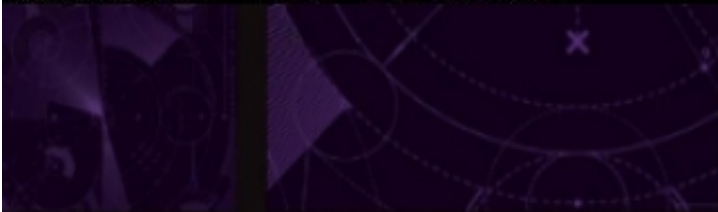




HANDBOOK OF MEASUREMENT

IN SCIENCE AND ENGINEERING

Volume 1



HANDBOOK OF MEASUREMENT

IN SCIENCE AND ENGINEERING

Volume 3

EDITED BY

MYER KUTZ



HANDBOOK OF MEASUREMENT

IN SCIENCE AND ENGINEERING

Volume 2

EDITED BY

MYER KUTZ

WILEY

Copyright © 2013 by John Wiley & Sons, Inc. All rights reserved.

Published by John Wiley & Sons, Inc., Hoboken, New Jersey.

Published simultaneously in Canada.

No part of this publication may be reproduced, stored in a retrieval system, or transmitted in any form or by any means, electronic, mechanical, photocopying, recording, scanning, or otherwise, except as permitted under Section 107 or 108 of the 1976 United States Copyright Act, without either the prior written permission of the Publisher, or authorization through payment of the appropriate per-copy fee to the Copyright Clearance Center, Inc., 222 Rosewood Drive, Danvers, MA 01923, (978) 750-8400, fax (978) 750-4470, or on the web at www.copyright.com. Requests to the Publisher for permission should be addressed to the Permissions Department, John Wiley & Sons, Inc., 111 River Street, Hoboken, NJ 07030, (201) 748-6011, fax (201) 748-6008, or online at <http://www.wiley.com/go/permission>.

Limit of Liability/Disclaimer of Warranty: While the publisher and author have used their best efforts in preparing this book, they make no representations or warranties with respect to the accuracy or completeness of the contents of this book and specifically disclaim any implied warranties of merchantability or fitness for a particular purpose. No warranty may be created or extended by sales representatives or written sales materials. The advice and strategies contained herein may not be suitable for your situation. You should consult with a professional where appropriate. Neither the publisher nor author shall be liable for any loss of profit or any other commercial damages, including but not limited to special, incidental, consequential, or other damages.

For general information on our other products and services or for technical support, please contact our Customer Care Department within the United States at (800) 762-2974, outside the United States at (317) 572-3993 or fax (317) 572-4002.

Wiley also publishes its books in a variety of electronic formats. Some content that appears in print may not be available in electronic formats. For more information about Wiley products, visit our web site at www.wiley.com.

Library of Congress Cataloging-in-Publication Data:

Handbook of Measurement in Science and Engineering / Myer Kutz, editor.
volumes cm

Includes bibliographical references and index.

ISBN 978-0-470-40477-5 (volume 1) – ISBN 978-1-118-38464-0 (volume 2) – ISBN 978-1-118-38463-3 (set) 1. Structural analysis (Engineering) 2. Dynamic testing. 3. Fault location (Engineering) 4. Strains and stresses—Measurement. I. Kutz, Myer.

TA645.H367 2012

620'.0044—dc23

2012011739

Printed in the United States of America

10 9 8 7 6 5 4 3 2 1

Table of Contents

VOLUME 1

PREFACE xxiii

CONTRIBUTORS xxvii

PART I CIVIL AND ENVIRONMENTAL ENGINEERING 1

1 New and Emerging Technologies in Structural Health Monitoring 3

Merit Enckell, Jacob Egede Andersen, Branko Glisic, and Johan Silfwerbrand

1.1 Introduction, 5

1.2 Background, 6

1.3 New and Emerging Technologies, 8

1.4 Fiber-Optic Technology, 16

1.5 Acoustic Emission, 24

1.6 Radar Technology, 27

1.7 Global Positioning System, 31

1.8 Corrosion Monitoring Systems, 33

1.9 Weigh-in-Motion (WIM) Systems, 35

1.10 Components of Structural Health Monitoring System, 37

1.11 Structural Health Monitoring System Design, 41

1.12 System Procurement and Installation, 44

1.13 Application of Structural Health Monitoring Systems, 47

1.14 Discussion, 67

1.15 Conclusion, 69

Acknowledgments, 70

References, 71

2 Applications of GIS in Engineering Measurements 79

Gary S. Spring

2.1 Introduction, 79

2.2 Background, 80

2.3 Basic Principles of GIS, 81

2.4 Measurement-Based GIS Applications, 96

2.5 Implementation Issues, 97

2.6 Conclusion, 100

References, 102

3 Traffic Congestion Management 105

Nagui M. Rouphail

3.1 Introduction and Background, 105

3.2 Scope of the Chapter, 106

3.3 Organization of the Chapter, 107

3.4 Fundamentals of Vehicle Emission Estimation, 107

3.5 Inventory of Traffic Congestion Management Methods, 112

3.6 Assessing Emission Impacts of Traffic Congestion Management, 119

3.7 Summary, 128

Acknowledgments, 129

References, 129

4 Seismic Testing of Highway Bridges	133
Eric V. Monzon, Ahmad M. Itani, and Gokhan Pekcan	
4.1 Introduction,	133
4.2 Similitude Requirements,	134
4.3 Specimen Fabrication,	141
4.4 Input Motion,	148
4.5 Instrumentation,	150
4.6 Data Acquisition and Processing,	155
4.7 Results,	157
References,	158
5 Measurements in Environmental Engineering	159
Daniel A. Vallero	
5.1 Introduction,	159
5.2 Environmental Sampling Approaches,	166
5.3 Laboratory Analysis,	169
5.4 Measurement Uncertainty,	183
5.5 Measurement Decision Making,	186
5.6 Environmental Indicators,	191
5.7 Extending Measurement Data Using Models,	199
5.8 Summary,	200
Nomenclature,	200
References,	202
6 Hydrology Measurements	205
Todd C. Rasmussen	
6.1 Introduction,	206
6.2 Precipitation,	209
6.3 Evapotranspiration,	212
6.4 Surface Flow,	216
6.5 Groundwater,	219
6.6 Soil Water,	223
6.7 Water Quality,	226
Suggested Readings,	231
7 Mobile Source Emissions Testing	233
Mohan Venigalla	
7.1 Testing for Regulatory Compliance,	234
References,	240

PART II MECHANICAL AND BIOMEDICAL ENGINEERING 241

8 Dimensions, Surfaces, and their Measurement 243

Mikell P. Groover

8.1 Dimensions, Tolerances, and Related Attributes, 244

8.2 Conventional Measuring Instruments and Gages, 245

8.3 Surfaces, 254

References, 256

9 Mass Properties Measurement 259

David Tellet

9.1 Introduction, 260

9.2 Mass and Weight, 262

9.3 Measurement Methodology, 264

9.4 Weight and Mass Measurement, 274

9.5 Center of Gravity Measurement, 275

9.6 MOI Measurement, 280

9.7 POI Measurement, 284

9.8 Measuring Large Vehicles, 287

9.9 Sources of Uncertainty, 292

References, 300

10 Force Measurement 301

Patrick Collins

10.1 Introduction, 302

10.2 Force Transducers, 303

10.3 Universal Testing Machines, 306

10.4 The Strain Gauge Sensor, 307

10.5 Resonant Element Transducers, 311

10.6 Surface Acoustic Wave Transducers, 314

10.7 Dynamometers, 317

10.8 Optical Force Transducers, 317

10.9 Magneto-Elastic Transducers, 320

10.10 Force Balance Transducers, 321

10.11 Force Transducer Characteristics, 321

10.12 Calibration, 323

10.13 Conclusion, 329

Glossary of Terms, 329

References, 340

11 Resistive Strain Measurement Devices 343

Mark Tuttle

11.1 Preliminary Discussion, 343

11.2 Resistance Metal Strain Gages, 349

11.3 Semiconductor Strain Gages, 363

11.4 Liquid Metal Strain Gages, 365

References, 366

12 Vibration Measurement 367

Sheryl M. Gracewski and Nigel D. Ramoutar

12.1 Introduction, 367

12.2 One-Degree-of-Freedom System Response, 369

12.3 Multi-Degree-of-Freedom Systems and the Frequency Response Function, 373

12.4 Vibration Measurement Equipment and Techniques, 388

12.5 Experimental Modal Analysis, 405

12.6 Applications of Vibration Measurement, 423

Nomenclature, 428

References, 431

13 Acoustical Measurements 433

Brian E. Anderson, Jonathan D. Blotter, Kent L. Gee, and Scott D. Sommerfeldt

13.1 Introduction, 434

13.2 Fundamental Measures, 436

13.3 Microphones, 445

13.4 Sound Pressure Level Measurements, 451

13.5 Measurement of Sound Isolation, 454

13.6 Room Acoustics Measurements, 457

13.7 Community and Environmental Noise, 463

13.8 Sound Intensity Measurements, 465

13.9 Sound Power Measurements, 472

13.10 Sound Exposure Measurements, 476

References, 479

14 Temperature Measurement 483

Peter R. N. Childs

Summary, 484

14.1 Introduction, 484

14.2 Selection, 487

14.3 Invasive Temperature Measurement, 489

14.4 Semi-Invasive Methods, 511

14.5 Noninvasive Methods, 514

14.6 Conclusions, 519

Nomenclature, 519

References, 521

15 Pressure and Velocity Measurements 527

Richard S. Figliola and Donald E. Beasley

15.1 Pressure Concepts, 528

15.2 Pressure Reference Instruments, 530

15.3 Pressure Transducers, 536

15.4 Pressure Transducer Calibration, 543

15.5 Pressure Measurements in Moving Fluids, 544

15.6 Modeling Pressure and Fluid Systems, 548

15.7 Design and Installation: Transmission Effects, 548

15.8 Fluid Velocity Measuring Systems, 552

Nomenclature, 563

References, 564

16 Luminescent Method for Pressure Measurement 567

Gamal E. Khalil, Jim W. Crafton, Sergey D. Fonov, Marvin Sellers, and Dana Dabiri

16.1 Introduction, 567

16.2 Principles of Pressure-Sensitive Paint, 569

16.3 Pressure-Sensitive Luminescent Dyes, 571

16.4 PSP Polymer and Binder, 572

16.5 Measurement Methods, 574

16.6 Pressure-Sensitive Paint Measurements, 588

Acknowledgments, 611

References, 612

17 Flow Measurement 615

Jesse Yoder

17.1 New-Technology and Traditional Technology Flowmeters, 616

17.2 Trends in Flow Measurement, 627

Further Readings, 628

18 Heat Flux Measurement 629

Thomas E. Diller

18.1 Introduction, 630

18.2 Important Issues, 631

18.3 Gages Based on Spatial Temperature Difference, 634

18.4 Gages Based on Temperature Change with Time, 643

18.5 Gages Based on Active Heating Methods, 648

18.6 Calibration and Errors, 653

References, 655

19 Heat Transfer Measurements for Nonboiling Two-Phase Flow 661

Afshin J. Ghajar and Clement C. Tang

19.1 Introduction, 661

19.2 Experimental Setup for Horizontal and Slightly Inclined Pipes, 662

19.3 Instruments for Measurement and Data Acquisition, 666

19.4 Heat Transfer Experiment Procedures, 667

19.5 Verifying the Functionality of the Experimental Setup, 670

19.6 Experimental Results of Two-Phase Flow, 673

19.7 Concluding Remarks, 682

Nomenclature, 683

References, 684

20 Solar Energy Measurements	687
Tariq Muneer and Yieng Wei Tham	
20.1 Introduction,	688
20.2 Measurement Equipment,	694
20.3 Equipment Error and Uncertainty,	703
20.4 Operational Errors,	704
20.5 Diffuse Radiation Data Measurement Errors,	704
20.6 Types of Sensors and their Accuracy,	711
20.7 Modern Developments,	711
20.8 Data Quality Assessment,	714
20.9 Statistical Evaluation of Models,	716
20.10 Outlier Analysis,	722
Acknowledgments,	722
References,	723
21 Wind Energy Measurements	727
Peter Gregg	
21.1 Introduction,	728
21.2 Concepts,	728
21.3 Measurements,	731
21.4 Evaluation,	739
References,	747
22 Human Movement Measurements	749
Rahman Davoodi	
22.1 Introduction,	749
22.2 Characterization of Human Movement,	750
22.3 Optical Motion Capture Systems,	751
22.4 Magnetic Motion Capture Systems,	754
22.5 Inertial Motion Capture Systems,	756
22.6 Discussion,	761
Acknowledgment,	762
References,	762
23 Flow Measurement	765
Arnold A. Fontaine, Keefe B. Manning, and Steven Deutsch	
23.1 Introduction,	765
23.2 Flow Measurement Applications,	768
References,	799

PART III INDUSTRIAL ENGINEERING 803

24 Statistical Quality Control 805

Magd E. Zohdi

24.1 Measurements and Quality Control, 805

24.2 Dimension and Tolerance, 805

24.3 Quality Control, 806

24.4 Interrelationship of Tolerances of Assembled Products, 812

24.5 Operation Characteristic (OC) Curve, 812

24.6 Control Charts for Attributes, 812

24.7 Acceptance Sampling, 815

24.8 Defense Department Acceptance Sampling by Variables, 817

Further Readings, 817

25 Evaluating and Selecting Technology-Based Projects 819

Hans J. Thamhain

25.1 Management Perspective, 819

25.2 Quantitative Approaches, 821

25.3 Qualitative Approaches, 826

25.4 Recommendations, 828

Variables and Abbreviations, 831

References, 831

26 Manufacturing Systems Evaluation 833

Walter W. Olson

26.1 Introduction, 833

26.2 Components of Environmentally Conscious Manufacturing, 834

26.3 Manufacturing Systems, 835

26.4 System Effects on ECM, 838

26.5 Assessment, 840

26.6 Summary, 844

References, 845

27 Measuring Performance of Chemical Process Equipment 847

Alan Cross

27.1 Introduction, 847

27.2 Direct Fired Heater Measurement and Process Control Instrumentation, 848

27.3 Crushing and Grinding Equipment Measurements, 851

References, 858

28 Industrial Energy Efficiency 859

B. Gopalakrishnan, D. P. Gupta, Y. Mardikar, and S. Chaudhari

28.1 Introduction, 860

28.2 Literature Review, 863

28.3 Data Analysis of Energy Efficiency Measures, 864

28.4 Energy Efficiency Measures in Major Energy Consuming Equipment, 872

28.5 Case Studies of Development of Energy-Efficiency Measures, 879

28.6 Conclusion, 881

Acknowledgments, 881

References, 881

29 Industrial Waste Auditing	885
C. Visvanathan	
29.1 Overview,	885
29.2 Waste-Minimization Programs,	886
29.3 Waste-Minimization Cycle,	888
29.4 Waste Auditing,	890
29.5 Conclusion,	909
Further Readings,	910
30 Organizational Performance Measurement	911
Jennifer A. Farris, Eileen M. Van Aken, and Geert Letens	
30.1 Introduction,	911
30.2 Summary,	940
References,	940
INDEX	

VOLUME 2

PART IV MATERIALS PROPERTIES AND TESTING 945

31 Viscosity Measurement	947
Ann M. Anderson, Bradford A. Bruno, and Lilla Safford Smith	
31.1 Viscosity Background,	947
31.2 Common Units of Viscosity,	949
31.3 Major Viscosity Measurement Methods,	959
31.4 ASTM Standards for Measuring Viscosity,	974
31.5 Questions to Ask When Selecting a Viscosity Measurement Technique,	976
References,	979
32 Tribology Measurements	981
Prasanta Sahoo	
32.1 Introduction,	982
32.2 Measurement of Surface Roughness,	983
32.3 Measurement of Friction,	988
32.4 Measurement of Wear,	992
32.5 Measurement of Test Environment,	994
32.6 Measurement of Material Characteristics,	998
32.7 Measurement of Lubricant Characteristics,	1001
32.8 Wear Particle Analysis,	1004
32.9 Industrial Measurements,	1005
32.10 Summary,	1006

33 Corrosion Monitoring 1007

Pierre R. Roberge

33.1 What is Corrosion Monitoring?, 1007

33.2 The Role of Corrosion Monitoring, 1008

33.3 Corrosion Monitoring System Considerations, 1010

References, 1116

34 Surface Properties Measurement 1121

Mrinalini Mulukutla and Sandip P. Harimkar

34.1 Introduction, 1121

34.2 Surface Properties, 1122

34.3 Microstructural Analysis, 1125

34.4 Compositional Analysis, 1128

34.5 Phase Analysis, 1130

34.6 Mechanical Testing, 1131

34.7 Corrosion Properties, 1141

34.8 Standards for Surface Engineering Measurement, 1145

References, 1147

35 Thermal Conductivity of Engineering Materials 1151

Juergen Blumm

35.1 Introduction, 1151

35.2 Stationary Methods for Measurement of the Thermal Conductivity, 1157

35.3 Transient Methods for the Measurement of the Thermal Conductivity, 1163

35.4 Test Results on Various Engineering Materials, 1173

References, 1188

36 Optical Methods for the Measurement of Thermal Conductivity 1189

Prabhakar R. Bandaru and Max S. Aubain

36.1 Thermal Boundary Resistance May Limit Accuracy in Contact-Based Thermal Conductivity (k) Measurements, 1189

36.2 Optical Measurements of k May Avoid Contact-Related Issues, 1192

36.3 Thermoreflectance (TR), 1196

36.4 Characteristics of Thermoreflectance from Si Thin Films—Modeling and Calibration, 1199

36.5 Experimental Procedures, 1202

36.6 Results and Discussion, 1204

36.7 Summary and Outlook, 1208

Acknowledgments, 1209

References, 1209

37 Selection of Metals for Structural Design 1213

Matthew J. Donachie

37.1 Introduction, 1214

37.2 Common Alloy Systems, 1215

37.3 What are Alloys and What Affects their Use?, 1215

37.4 What are the Properties of Alloys and How are Alloys Strengthened?, 1218

37.5 Manufacture of Alloy Articles, 1221

37.6 Alloy Information, 1221

37.7 Metals at Lower Temperatures, 1231

37.8 Metals at High Temperatures, 1233

37.9 Melting and Casting Practices, 1236

37.10 Forging, Forming, Powder Metallurgy, and Joining of Alloys, 1242

37.11 Surface Protection of Materials, 1245

37.12 Postservice Refurbishment and Repair, 1248

37.13 Alloy Selection: A Look at Possibilities, 1249

37.14 Level of Property Data, 1252

37.15 Thoughts on Alloy Systems, 1252

37.16 Selected Alloy Information Sources, 1259

Further Readings, 1261

38 Mechanical Properties of Polymers 1263

Daniel Liu, Jackie Rehkopf, and Maureen Reitman

38.1 Microstructure and Morphology of Polymers—Amorphous Versus Crystalline, 1264

38.2 General Stress–Strain Behavior, 1265

38.3 Viscoelasticity, 1271

38.4 Mechanical Models of Viscoelasticity, 1272

38.5 Time–Temperature Dependence, 1274

38.6 Deformation Mechanisms, 1274

38.7 Crazing, 1277

38.8 Fracture, 1279

38.9 Modifying Mechanical Properties, 1284

38.10 Load-Bearing Applications: Creep, Fatigue Resistance, and High Strain Rate Behavior, 1285

References, 1290

39 Electrical Properties of Polymers 1291

Evaristo Riande and Ricardo Diaz-Calleja

39.1 Introductory Remarks, 1291

39.2 Polarity and Permittivity, 1292

39.3 Measurements of Dielectric Permittivity, 1293

39.4 Polarization and Dipole Moments in Isotropic Systems, 1297

39.5 Thermostimulated Depolarization Currents, 1316

39.6 Conductivity in Polyelectrolytes and Polymer-Electrolytes as Separators for Low Temperature Fuel Cells and Electrical Batteries, 1318

39.7 Semiconductors and Electronic Conducting Polymers, 1324

39.8 Ferroelectricity, Pyroelectricity, and Piezoelectricity in Polymers, 1328

39.9 Nonlinear Polarization in Polymers, 1331

39.10 Elastomers for Actuators and Sensors, 1333

39.11 Electrical Breakdown in Polymers, 1336

References, 1338

40 Nondestructive Inspection 1343

Robert L. Crane and Jeremy S. Knopp

40.1 Introduction, 1344

40.2 Liquid Penetrants, 1347

40.3 Radiography, 1351

40.4 Ultrasonic Methods, 1361

40.5 Magnetic Particle Method, 1370

40.6 Thermal Methods, 1373

40.7 Eddy Current Methods, 1375

References, 1410

41 Testing of Metallic Materials 1413

Peter C. McKeighan

41.1 Mechanical Test Laboratory, 1414

41.2 Tensile and Compressive Property Testing, 1418

41.3 Creep and Stress Relaxation Testing, 1420

41.4 Hardness and Impact Testing, 1422

41.5 Fracture Toughness Testing, 1425

41.6 Fatigue Testing, 1429

41.7 Other Mechanical Testing, 1433

41.8 Environmental Considerations, 1434

Acknowledgments, 1436

References, 1436

42 Ceramics Testing 1437

Shawn K. McGuire and Michael G. Jenkins

42.1 Introduction, 1437

42.2 Mechanical Testing, 1438

42.3 Thermal Testing, 1451

42.4 Nondestructive Evaluation Testing, 1458

42.5 Electrical Testing, 1460

42.6 Summary, 1461

References, 1461

43 Plastics Testing 1463

Vishu Shah

43.1 Introduction, 1464

43.2 Mechanical Properties, 1464

43.3 Thermal Properties, 1481

43.4 Electrical Properties, 1484

43.5 Weathering Properties, 1488

43.6 Optical Properties, 1492

Further Readings, 1496

44 Testing and Instrumental Analysis for Plastics Processing:

Key Characterization Techniques 1499

Maria del Pilar Noriega

44.1 FTIR Spectroscopy, 1499

44.2 Chromatography (GC, GC-MSD, GC-FID, and HPLC), 1500

44.3 DSC and Thermogravimetry (TGA), 1510

44.4 Rheometry, 1518

References, 1527

45 Analytical Tools for Estimation of Particulate Composite

Material Properties 1529

Tarek I. Zohdi and Magd E. Zohdi

45.1 Introduction, 1529

45.2 Concepts in Statistical Quality Control, 1530

45.3 Effective Property Estimates, 1531

45.4 Summary, 1535

References, 1537

PART V INSTRUMENTATION 1539

46 Instrument Statics 1541

Jerry Lee Hall, Sriram Sundararajan, and Mahmood Naim

46.1 Terminology, 1541

46.2 Static Calibration, 1544

46.3 Statistics in the Measurement Process, 1547

References, 1570

47 Input and Output Characteristics 1573

Adam C. Bell

47.1 Introduction, 1574

47.2 Familiar Examples of Input–Output Interactions, 1575

47.3 Energy, Power, Impedance, 1578

47.4 Operating Point of Static Systems, 1586

47.5 Transforming the Operating Point, 1598

47.6 Measurement Systems, 1602

47.7 Distributed Systems in Brief, 1607

47.8 Concluding Remarks, 1609

References, 1610

48 Bridge Transducers 1611

Patrick L. Walter

48.1 Terminology, 1612

48.2 Flexural Devices in Measurement Systems, 1612

48.3 The Resistance Strain Gage, 1615

48.4 The Wheatstone Bridge, 1625

48.5 Resistance Bridge Balance Methods, 1634

48.6 Resistance Bridge Transducer Measurement System Calibration, 1636

48.7 Resistance Bridge Transducer Measurement System Considerations, 1646

48.8 AC Impedance Bridge Transducers, 1655

References, 1660

Further Readings, 1661

49 Signal Processing 1663

John Turnbull

49.1 Frequency-Domain Analysis of Linear Systems, 1663

49.2 Basic Analog Filters, 1666

49.3 Basic Digital Filter, 1672

49.4 Stability and Phase Analysis, 1680

49.5 Extracting Signal from Noise, 1682

References, 1683

50 Data Acquisition and Display Systems 1685

Philip C. Milliman

50.1 Introduction, 1686

50.2 Data Acquisition, 1687

50.3 Process Data Acquisition, 1688

50.4 Data Conditioning, 1691

50.5 Data Storage, 1699

50.6 Data Display and Reporting, 1704

50.7 Data Analysis, 1707

50.8 Data Communications, 1708

50.9 Other Data Acquisition and Display Topics, 1712

50.10 Summary, 1715

References, 1715

PART VI MEASUREMENT STANDARDS 1517

51 Mathematical and Physical Units, Standards, and Tables 1719

Jack H. Westbrook

51.1 Symbols and Abbreviations, 1720

Bibliography for Letter Symbols, 1731

Bibliography for Graphic Symbols, 1737

51.2 Mathematical Tables, 1742

51.3 Statistical Tables, 1765

51.4 Units and Standards, 1775

Bibliography for Units and Measurements, 1802

51.5 Tables of Conversion Factors, 1802

51.6 Standard Sizes, 1833

51.7 Standard Screws, 1886

52 Measurement Uncertainty 1911

David Clippinger

52.1 Introduction, 1911

52.2 Literature, 1914

52.3 Evaluation of Uncertainty, 1915

52.4 Discussion, 1924

Disclaimer, 1924

References, 1925

53 Measurements 1927

E. L. Hixson and E. A. Ripperger

53.1 Standards and Accuracy, 1927

53.2 Impedance Concepts, 1930

53.3 Error Analysis, 1935

References, 1942

INDEX I-1

CONTRIBUTORS

Jacob Egede Andersen, COWI A/S, Lyngby, Denmark

Ann M. Anderson, Union College, Schenectady, NY, USA

Brian E. Anderson, Brigham Young University, Provo, UT, USA

Max S. Aubain, University of California—San Diego, La Jolla, CA, USA

Prabhakar R. Bandaru, University of California—San Diego, La Jolla, CA, USA

Donald E. Beasley, Clemson University, Clemson, SC, USA

Adam C. Bell, Dartmouth, Nova Scotia, Canada

Jonathan D. Blotter, Brigham Young University, Provo, UT, USA

Juergen Blumm, NETZSCH-Geraetebau GmbH, Selb, Germany

Bradford A. Bruno, Union College, Schenectady, NY, USA

S. Chaudhari, West Virginia University, Morgantown, WV, USA

Peter R.N. Childs, Imperial College London, London, UK

David Clippinger, U.S. Coast Guard Academy, New London, CT, USA

Patrick Collins, MecMesin Ltd., Slinfold, West Sussex, UK

Jim W. Crafton, Innovative Scientific Solutions, Inc., Dayton, OH, USA

Robert L. Crane, Kettering, OH, USA

Alan Cross, Little Neck, NY, USA

Dana Dabiri, University of Washington, Seattle, WA, USA

Rahman Davoodi, University of Southern California, Los Angeles, CA, USA

María del Pilar Noriega, ICIPC, Medellin, Antioquia, Columbia

Steven Deutsch, Pennsylvania State University, University Park, PA, USA

Ricardo Díaz-Calleja, ITE (Universidad Politécnica de Valencia), Valencia, Spain

Thomas E. Diller, Virginia Tech, Blacksburg, VA, USA

Matthew J. Donachie, Winchester, NH, USA

Merit Enckell, Royal Institute of Technology (KTH), Stockholm, Sweden; COWI A/S, Lyngby, Denmark

Jennifer A. Farris, Texas Tech, Lubbock, TX, USA

Richard S. Figliola, Clemson University, Clemson, SC, USA

Sergey D. Fonov, Innovative Scientific Solutions, Inc., Dayton, OH, USA

Arnold A. Fontaine, Pennsylvania State University, University Park, PA, USA

Kent L. Gee, Brigham Young University, Provo, UT, USA

Afshin J. Ghajar, Oklahoma State University, Stillwater, OK, USA

Branko Glisic, Princeton University, Princeton, NJ, USA

B. Gopalakrishnan, West Virginia University, Morgantown, WV, USA

Sheryl M. Gracewski, University of Rochester, Rochester, NY, USA

Peter Gregg, GE, Schenectady, NY, USA

Mikell P. Groover, Lehigh University, Bethlehem, PA, USA

D.P. Gupta, West Virginia University, Morgantown, WV, USA

Jerry Lee Hall, Iowa State University, Ames, IA, USA

Sandip P. Harimkar, Oklahoma State University, Stillwater, OK, USA

E.L. Hixson, University of Texas, Austin, TX, USA

Ahmad M. Itani, University of Nevada, Reno, NV, USA

Michael G. Jenkins, University of Washington, Seattle, WA, USA

Gamal E. Khalil, University of Washington, Seattle, WA, USA

Jeremy S. Knopp, Wright Patterson Air Force Base, Dayton, OH, USA

Geert Letens, Royal Military Academy, Brussels, Belgium

Daniel Liu, Exponent, Bowie, MD, USA

Keefe B. Manning, Pennsylvania State University, University Park, PA, USA

Y. Mardikar, West Virginia University, Morgantown, WV, USA

Shawn K. McGuire, Stanford University, Palo Alto, CA, USA

Peter C. McKeighan, Warrenville, IL, USA

- Philip C. Milliman**, Weyerhaeuser Company, Federal Way, WA, USA
- Eric V. Monzon**, University of Nevada, Reno, NV, USA
- Mrinalini Mulukutla**, Oklahoma State University, Stillwater, OK, USA
- Tariq Muneer**, Edinburgh Napier University, Edinburgh, UK
- Mahmood Naim**, Union Carbide Corporation, Indianapolis, IN, USA
- Walter W. Olson**, University of Toledo, Toledo, OH, USA
- Gokhan Pekcan**, University of Nevada, Reno, NV, USA
- Nigel D. Ramoutar**, Gleason Works, Rochester, NY, USA
- Todd C. Rasmussen**, University of Georgia, Athens, GA, USA
- Jackie Rehkopf**, Exponent, Bowie, MD, USA
- Maureen Reitman**, Exponent, Bowie, MD, USA
- Evaristo Riande***, Instituto de Ciencia and Tecnología de Polímeros (Consejo Superior de Investigaciones Científicas), Madrid, Spain
- E.A. Ripperger**, University of Texas, Austin, TX, USA
- Pierre R. Roberge**, Royal Military College of Canada, Kingston, Ontario, Canada
- Nagui M. Roupail**, North Carolina State University, Raleigh, NC, USA
- Prasanta Sahoo**, Jadavpur University, Kolkata, India
- Marvin Sellers**, Aerospace Testing Alliance, Arnold Air Force Base, TN, USA
- Johan Silfwerbrand**, Royal Institute of Technology (KTH), Stockholm, Sweden
- Lilla Safford Smith**, Union College, Schenectady, NY, USA
- Vishu Shah**, Consultek Consulting Group, Brea, CA, USA
- Scott D. Sommerfeldt**, Brigham Young University, Provo, UT, USA
- Gary S. Spring**, Merrimack College, North Andover, MA, USA
- Sriram Sundararajan**, Iowa State University, Ames, IA, USA
- Clement C. Tang**, University of North Dakota, Grand Forks, ND, USA
- David Tellet**, Society of Allied Weight Engineers, Inc. (SAWE, Inc.)
- Yieng Wei Tham**, Edinburgh Napier University, Edinburgh, UK
- Hans J. Thamhain**, Bentley University, Waltham, MA, USA
- John Turnbull**, Case Western Reserve University, Cleveland, OH, USA
- Mark Tuttle**, University of Washington, Seattle, WA, USA
- Daniel A. Vallero**, Duke University, Chapel Hill, NC, USA

Eileen M. Van Aken, Virginia Tech, Blacksburg, VA, USA

Mohan Venigalla, George Mason University, Fairfax, VA, USA

C. Visvanathan, Asian Institute of Technology, Klongluang Pathumthani, Thailand

Patrick L. Walter, Texas Christian University, Fort Worth, TX, USA

Jack H. Westbrook, Ballston Spa, NY, USA

Jesse Yoder, Flow Research, Inc., Wakefield, MA, USA

Magd E. Zohdi, Louisiana State University, Baton Rouge, LA, USA

Tarek I. Zohdi, University of California, Berkeley, CA, USA

PREFACE

The idea for the *Handbook of Measurement in Science and Engineering* came from a Wiley book first published over 30 years ago. It was *Fundamentals of Temperature, Pressure and Flow Measurements*, written by a sole author, Robert P. Benedict, who also wrote Wiley books on gas dynamics and pipe flow. Bob was a pleasant, unassuming, and smart man. I was the Wiley editor for professional-level books in mechanical engineering when Bob was writing such books, so I knew him as a colleague. I recall meeting him in the Wiley offices at a time when he seemed to be having some medical problems, which he was reluctant to talk about. Recently, I discovered a book published in 1972 by a London firm, Pickering & Inglis, which specializes in religion. This book was *Journey Away From God*, an intriguing title. The author's name was Robert P. Benedict. I do not know whether the two Benedicts are in fact the same person, although Amazon seems to think so. (See the Robert P. Benedict page.) In any case, I do not recall Bob's mentioning the book when we had an occasion to talk.

The moral of this story, if there is one, is that the men and women who contributed the chapters in this handbook are real people, who have real-world concerns, in addition to the expertise required to write about technology. They have families, jobs, careers, and all manner of cares about the minutia of daily life to deal with. And that they have been able to find the time and energy to write these chapters is remarkable. I salute them.

I have spent a lot of time in my life writing and editing books. I wrote my first Wiley book somewhat earlier than Bob Benedict wrote his. When Wiley published *Temperature Control* in 1967, I was in my mid-twenties and was a practicing engineer, working on temperature control of the Apollo inertial guidance system at the MIT Instrumentation Lab, where I had done my bachelor's thesis. One of the coauthors of my book was to have been a Tufts Mechanical Engineering Professor by the name of John Sununu (yes, *that* John Sununu), but he and the other coauthor dropped out of the project before the contract was signed. So I wrote the short book myself.

Bob Benedict's measurement book, the third edition of which is still in print, surfaced several years ago, during a discussion I was having with one of my Wiley editors,

Bob Argentieri, about possible projects we could collaborate on. It turned out that no one had attempted to update Benedict's book. I have not been a practicing engineer for some time, so I was not in a position to do an update as a single author—or even with a collaborator or two. Most of my career life has been in scientific and technical publishing, however, and for over a decade I have conceived of, and edited, numerous handbooks for several publishers. (I also write fiction, but that is another story.) So, it was natural for me to think about using Benedict's book as the kernel of a much larger and broader reference work dealing with engineering measurements. The idea, formed during that discussion, that I might edit a contributed handbook on engineering measurements took hold, and with the affable and expert guidance of my other Wiley editor, George Telecki, the volume you are holding in your hands, or reading on an electronic device, came into being.

Like many such large reference works, this handbook went through several iterations before the final table of contents was set, although the general plan for arrangement of chapters has been the same throughout the project. The initial print version of the handbook is divided into two volumes. The chapters are arranged essentially by engineering discipline. The first volume contains 30 chapters related to five engineering disciplines, which are divided into three parts:

- Part I, Civil and Environmental Engineering, which contains seven chapters, all but one of them dealing with measurement and testing techniques for structural health monitoring, GIS and computer mapping, highway bridges, environmental engineering, hydrology, and mobile source emissions (the exception being the chapter on traffic congestion management, which describes the deployment of certain measurements);
- Part II, Mechanical and Biomedical Engineering, which contains 16 chapters, all of them dealing with techniques for measuring dimensions, surfaces, mass properties, force, resistive strain, vibration, acoustics, temperature, pressure, velocity, flow, heat flux, heat transfer for non-boiling two-phase flow, solar energy, wind energy, human movement, and physiological flow;
- Part III, Industrial Engineering, which contains seven chapters dealing with statistical quality control, evaluating and selecting technology-based projects, manufacturing systems evaluation, measuring performance of chemical process equipment, industrial energy efficiency, industrial waste auditing, and organizational performance measurement.

The second volume contains 23 chapters divided into three parts:

- Part IV, Materials Properties and Testing, which contains 15 chapters dealing with measurement of viscosity, tribology, corrosion, surface properties, and thermal conductivity of engineering materials; properties of metals, alloys, polymers, and particulate composite materials; nondestructive inspection; and testing of metallic materials, ceramics, plastics, and plastics processing;
- Part V, Instrumentation, which contains five chapters covering electronic equipment used for measurements;
- Part VI, Measurement Standards, which contains three chapters covering units and standards, measurement uncertainty and error analysis.

Major reference works, like this handbook, are generally incomplete when they are first published. Editors cannot wait for tardy contributors, some contributors simply cannot manage to deliver their chapters no matter how much time they are given, and contributors cannot be secured for all the chapters an editor has in mind for a reference work. Among the topics that were either contracted for but were not delivered or for which contributors were not found are surveying engineering, engineering seismology, construction materials properties, turbulence, water quality, wastewater engineering, trace gases in the atmosphere, experimental methods, experimental design, shape and deformation, thermal systems, energy audits, electrical properties of materials, rheology, software engineering, biomedical electronics, physiology, dielectric properties of tissues, productivity, remote sensing, and data analysis.

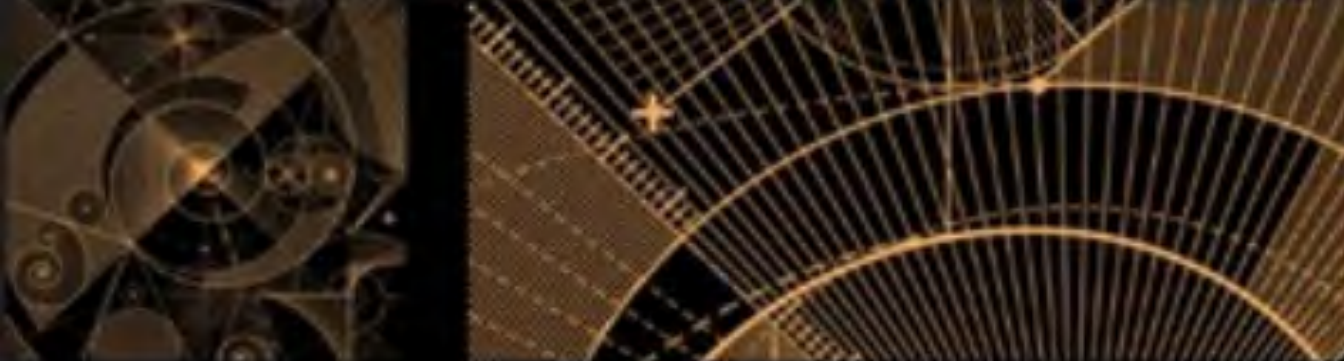
Of course, such a list, when combined with the chapters being published, does not exhaust the list of possible topics for a measurements handbook. Be that as it may, the usual practice has been to attempt to include additional topics, together with updates of existing chapters, in new editions of a reference work, which tend to appear in 5-to-10-year, or longer, intervals. I have done this successfully in the *Mechanical Engineers' Handbook* that I edit for Wiley. That work is now, in its forthcoming fourth edition, a four-volume handbook, in contrast to the single-volume first and second editions.

In the case of this measurements handbook, however, George Telecki has proposed that the online version be dynamic, with 20 or so articles added annually. Furthermore, coverage will be expanded beyond engineering disciplines to include chemistry, life sciences, and physics, thereby justifying the handbook's title, *Handbook of Measurement in Science and Engineering*. In addition, existing chapters will be updated as the need arises. I have campaigned for years to get my publishers to adopt this scheme, and I am gratified that Wiley intends to pursue it. I will attempt to get the others to follow suit.

Thanks to Kari Capone for shepherding the manuscript toward production and to the stalwarts Kristen Parrish and Shirley Thomas, for bringing the handbook home. Thanks, also, to my wife Arlene, who helps me with everything else.

MYER KUTZ

Delmar, NY
June, 2012



HANDBOOK OF MEASUREMENT

IN SCIENCE AND ENGINEERING

Volume 1



EDITED BY

MYER KUTZ



 WILEY

(0.1.2.3)

PART I

CIVIL AND ENVIRONMENTAL ENGINEERING

1

NEW AND EMERGING TECHNOLOGIES IN STRUCTURAL HEALTH MONITORING

MERIT ENCKELL, JACOB EGEDE ANDERSEN, BRANKO GLISIC, AND
JOHAN SILFWERBRAND

- 1.1 Introduction
- 1.2 Background
- 1.3 New and emerging technologies
 - 1.3.1 General
 - 1.3.2 Fiber-optic sensors (FOS)
 - 1.3.3 The global positioning system (GPS)
 - 1.3.4 Microelectromechanical Systems
 - 1.3.5 Corrosion monitoring
 - 1.3.6 B-WIM, WIM
 - 1.3.7 Nondestructive testing (NDT)
 - 1.3.8 Interferometric radar
 - 1.3.9 Photogrammetry
 - 1.3.10 Smart technical textiles
 - 1.3.11 Specific issues around usage of new technologies
 - 1.3.12 Chosen technologies and motivation
- 1.4 Fiber-optic technology
 - 1.4.1 General
 - 1.4.2 Sensors based on Sagnac, Michelson, and Mach–Zehnder interferometers
 - 1.4.3 Sensor based on the Fiber Bragg Gratings
 - 1.4.4 Sensors based on Fabry–Perot interferometry
 - 1.4.5 Best performances of discrete FOS
 - 1.4.6 Distributed sensors
- 1.5 Acoustic emission
 - 1.5.1 Theory of acoustic emission

- 1.5.2 Sources of acoustic emission
 - 1.5.3 The development of acoustic emission in industry and civil engineering
 - 1.5.4 Acoustic emission systems
 - 1.5.5 Codes, standards, and recommended practice in acoustic emission
 - 1.6 Radar technology
 - 1.6.1 General
 - 1.6.2 Ground-penetrating radar
 - 1.6.3 Interferometric radar
 - 1.7 Global Positioning System
 - 1.8 Corrosion monitoring systems
 - 1.9 Weigh-in-motion (WIM) systems
 - 1.9.1 Weigh-in-motion
 - 1.9.2 Railway weigh-in-Motion
 - 1.9.3 Bridge weigh-in-Motion
 - 1.10 Components of structural health monitoring system
 - 1.10.1 Sensory system
 - 1.10.2 Data acquisition system
 - 1.10.3 Data processing and control system
 - 1.10.4 User interface
 - 1.10.5 Maintenance tools
 - 1.11 Structural health monitoring system design
 - 1.11.1 Structural analysis for new structure
 - 1.11.2 Structural analysis for existing structure
 - 1.11.3 Sensor selection
 - 1.11.4 Data acquisition issues
 - 1.11.5 Responsibilities and installation planning
 - 1.12 System procurement and installation
 - 1.12.1 System procurement
 - 1.12.2 Commissioning
 - 1.12.3 Installation
 - 1.12.4 Lifetime support
 - 1.12.5 System efficiency and redundancy
 - 1.12.6 Dismantling environmental issues
 - 1.13 Application of structural health monitoring systems
 - 1.13.1 High-rise building, Singapore—2001
 - 1.13.2 The New Årsta Railway Bridge, Sweden—2005
 - 1.13.3 Stonecutters Bridge, Hong Kong—2010
 - 1.13.4 Severn River Crossing, UK—2010
 - 1.13.5 A4 Hammersmith Flyover, UK—2010
 - 1.13.6 Streicker Bridge, United States—2010
 - 1.13.7 Messina Bridge, Italy—2018
 - 1.14 Discussion
 - 1.14.1 Development of new and emerging technologies
 - 1.14.2 Obstacles
 - 1.14.3 Need for education and collaboration
 - 1.14.4 Future use and development
 - 1.15 Conclusion
- Acknowledgments
- References

1.1 INTRODUCTION

Structural Health Monitoring (SHM) and modern sensory technology together with advanced data acquisition are currently available for various applications. The area of monitoring is also very wide and incorporates several disciplines. Civil engineers working with monitoring have to cooperate very closely with various kinds of specialists to assure that the chosen monitoring system provides the information that they are looking for.

Organized SHM became a well-known concept during the last decades. Health Monitoring according to Aktan et al. (2000), may be defined as

the measurement of the operating and loading environment and the critical responses of a structure to track and evaluate the symptoms of operational incidents, anomalies, and/or deterioration or damage indicators that may affect operation, serviceability, or safety reliability

SHM helps to control and verify structural behavior: the condition or changes in the condition of a structure. SHM gives more improved and precise information than visual inspection about the real condition of the structure at real time. Decision making concerning the maintenance, economy, and the safety of the structures is easier with an appropriate Structural Health Monitoring System (SHMS).

Factors such as shortened construction periods, increased traffic loads, new high speed trains causing new dynamic and fatigue problems, new materials, new construction solutions, slender constructions, limited economy, and need for timesaving demanded for better control as well as verification and benefited for SHM.

Old deteriorated structures, especially, the ones that do suffer about fatigue effects may have malfunction and could collapse. But also structures that are not at the end of their lifetime do fail. Some serious collapses have taken place in recent years, for example, Sport Arena Bad Reichenhall in south Germany and Arena in Katowice, south Poland in 2006; I-35W Mississippi River bridge in Minnesota, United States in 2007. Many people were killed and injured caused by these mentioned collapses. The newly built bridges called Gröndal Bridge and Alvik Bridge in Stockholm revealed extensive cracking in the webs of their concrete hollow box girder sections just after a few years of operation (Sundquist and James, 2004), and two tension rods and a crossbeam from a recently installed repair collapsed in Oakland Bay Bridge, San Francisco in 2009, causing the bridge to be closed temporarily. Structures also do collapse during the construction period, and workers may be killed or injured as recently for the cable-stayed bridge across the Chambal River in India. SHM can start with a sensor installation and monitoring already during the construction period and therefore provide information throughout the whole life span of the structure: construction, testing, operation and also demolition. It may capture the behavior that visual inspection does not accomplish and therefore there is possibility that it may save human life.

The rapid development of technology in the fields of sensors, data acquisition and communication, signal analysis and data processing provide SHM with great profit. Buildings, bridges, wind farms, nuclear power plants, geotechnical structures, historical buildings and monuments, dams, offshore platforms, pipelines, ocean structures, airplanes, wind plants, turbine blades, and so on, may be objects for monitoring activities. The monitoring can be periodic or continuous, short term or long term, local or global, and the monitoring system can consist of a few sensors up to hundreds or even thousands of them depending on the demands of the monitoring object.

Structural Health Monitoring System for a structure consists of sensors, data acquisition systems, data transfer and storage systems, data management that normally includes data analysis as well as presentation, and data interpretation. The number of sensors used in monitoring is endless. Different applications with various techniques such as electrical, optical, acoustical, and geodetical are available. Various parameters such as strain, displacement, inclination, stress, pressure, humidity, temperature, different chemical quantities, and environmental parameters such as wind speed and direction can be monitored.

Conventional sensors used for civil engineering such as strain gauges, traditional accelerometers, inclinometers, load cells, vibrating wires, linear variable differential transformers (LVDT) are able to measure many parameters and have a long experience in use. On the other hand, the evolution of emerging technologies together with computer-based data acquisition, advanced signal and data communication have made the evaluation of new techniques and sensors for civil engineering purposes possible.

Fiber-optic sensors, microelectromechanical systems (MEMS), optical distance measurement techniques, acoustic emission, and different type of lasers and radars have been under great development in recent years and are now available on the market. They are characterized by high accuracy, straightforward usage, and data-collecting concept. These techniques often allow very delicate measuring in harsh conditions and in various applications. The automatic collection of the data saves time, and it has advantages with respect to manual measurements. The reliability and durability of the sensors become significant when choosing the appropriate instrumentation. These new high-tech sensors also allow for not only high accuracy but also high precision, high and constant sensitivity, stability over time, no drift, and they are often temperature compensated.

The market with fiber-optic sensors and their applications is massive. There are several different techniques and various kinds of sensors that also can be modified for unique monitoring needs for a particular structure. Fiber-optic sensors allow for measurements that have been unpractical or too costly with the traditional sensor technology. Hundreds measuring points along the same fiber, as well as distributed sensing, versatility, insensitivity for electromagnetic fields, operability under extreme climate conditions and also the fact that there is no need for protection against lightning are some of the advantages over the electrical-based counterparts (Ross and Matthews, 1995).

Numerous different monitoring projects have taken place in recent years, and SHM has become a standard when designing large or complicated structures all over the world. It is now an essential implement in managing structures for civil engineers as well as owners. The modern sensory technology is also more commonly used as well as accepted by the civil engineering society.

As the subject is large and the number of emerging technologies and SHM applications is numerous, the chapter principally brings up and discusses the subjects mainly from the civil engineering point of view. The most common techniques that are suitable for civil engineering applications are presented and discussed and examples are given.

1.2 BACKGROUND

Testing and measuring of certain desired parameters have taken place in the field of civil engineering in the latest century (Mufti, 2001). Steel strains, rock stresses, concrete curing temperature, shrinkage and stresses, pressure of the concrete in formworks, vibrations, and many other phenomena that engineers felt uncertain about have been measured and recorded.

Lack of knowledge or experience was the driving force. The earliest reference to dynamic testing in the late nineteenth century can be seen in Salawu and Williams (1995). But as the Tacoma Narrows Bridge collapsed in 1940, it forced the engineers to face the problem with long-span bridge aerodynamics (Miyata et al., 2002; Miyata, 2003). As a result, the dynamic measuring techniques developed and increased significantly in the following decades. These activities were small scale and not really organized structural monitoring. The technology was not yet well developed in term of automation and data handling. The amount of data was held in small portions in order to be handled and used in a decent way.

Damage identification in aerospace and mechanical engineering started the organized SHM but were followed by civil engineering society. Organized SHM activities were acknowledged in the last decades. The subject also emerged various engineering disciplines as the new sensor technology and information technology entered the field. At present, many civil engineering structures are monitored continuously and true real-time information of these structures is provided.

Furthermore, a lot of discussion is going on between scientists and other related disciplinary in order to create standards and international guidelines. The concept “Health Monitoring” was defined by Aktan et al. (2000). They also published the report “Development of a Model health Monitoring Guide for major Bridges” (Aktan et al., 2001). ISIS Canada Research Network was established in 1995. ISIS published a report “Guidelines for Structural health Monitoring” in 2001 (Mufti, 2001) and several others design manuals and reports are published up today. International Society for Structural Health Monitoring of Intelligent Infrastructure (ISHMII), a nonprofit organization was founded in 2003. The aim of the Society is to advance the understanding and the application of SHM in the civil engineering infrastructure, in the service of the engineering profession and society. ISHMII also publish a paper called *The Journal of Civil Structural Health Monitoring (JCSHM)*. JCSHM publishes articles to advance the understanding and the application of health monitoring methodologies for the condition assessment and management of the civil infrastructure systems.

The European Union project, sustainable bridges also started to work with functional requirements for railway bridges in 2003. The goal was to achieve increased capacities required to meet future demands for increased traffic levels and heavier axle loads. The activities would also deal with efficient condition monitoring systems for railway bridges, and numerous reports were published and can be seen in www.sustainablebridges.net/.¹

Farrar and Worden (2007) gives an introduction to SHM. A very comprehensible and extensive paper about SHM of civil infrastructure that illustrates the different topics of SHM of dams, bridges, offshore, buildings, towers, nuclear installations, tunnels, and excavations can be seen in Brownjohn (2007). Problems, challenges, and limitation of development of SHM are brought up in these papers, the subject that is very important but still left out completely by many other authors. Some standards are also already created and can be seen in BSI (2004).

Recently, the SmartEN research project for research into smart use of wireless sensor networks and the integration of SHM monitoring results with Maintenance Management Systems for the optimization of maintenance have been formed funded by the European Commission seventh framework for research. These developments open up a completely novel area of multidisciplinary research toward the “smart” management of sustainable

¹Accessed January 27, 2010.

environment. Even though there are top research institutions working in the field of wireless sensors and others in the civil infrastructure reliability and management (Stochastic Optimization Methods for Infrastructure Management with Incomplete Monitoring Data Le Thanh Nam, 2009, Kyoto University), most of the activity is fragmented and there is no significant activity in performing multidisciplinary structured research for developing integrated smart and dynamic systems for effective management of the built and natural environment. The aim of SmartEN is to fill this gap and push innovation through the development of an ITN network that will focus on the development and effective integration of emerging technologies targeting key application areas of current interest to the European Commission and internationally.

Major construction projects such as large bridges, dams, high-rise buildings, nuclear power plants, offshore structures, tunnels, and harbor structures demand a lot of investment and resources. Their malfunction or collapse may cause severe damage to the society, its inhabitants or to the environment. Therefore, need to control these structures have emerged needs for monitoring that in turn has been adapted to smaller scale projects.

1.3 NEW AND EMERGING TECHNOLOGIES

1.3.1 General

Established technologies are well known and have a proper long-term experience. Emerging technologies, on the other hand, are science-based innovations that have the potential to create a new industry or transform an existing one (Day et al., 2000). Emerging technologies demand for new kind of thinking in order to prevail and copy with them. They are also characterized with certain ambiguity and complexity as they are in accelerating change.

Authors of this chapter have been working with established as well as new and emerging technologies: both in theory and practice. Several both small-scale and large-scale installations, testing and measuring campaigns are completed and Structural Health Monitoring Systems are designed, mostly for international major bridges but also for other applications such as high-rise buildings, tunnels, machinery, and heritage structures. The authors' intention is to provide a comprehensive presentation about the subjects as well as about related tasks. This will help the reader to understand the new challenges in order to accomplish sustainable SHM with new and emerging technologies.

Due to tremendous change in information technologies and sensor development, it is possible today to measure nearly any asked parameter and also perform automatic data processing and analysis in real time and with remote access. Sophisticated systems are present: over icing, corrosion, vibration, deflection, chemical concentration, humidity, strain, stresses and combined as well as distributed parameters can be measured. Some new technologies are better established like fiber-optic sensors but there are still a lot of challenges as research is ongoing and new sensors enter the field constantly.

A lot of work needs also be done in order to guarantee severity and long-term function of these innovative systems; one of the main tasks is to provide for defect localization as well as prediction of the future condition of the structure by automatic data analysis. New and emerging technologies have many advantages compared with old technologies such as adaptability, reliability, possibility for sophisticated measurements in real-time, and in harsh conditions. There are also disadvantages as some uncertainty is present. People working with new and emerging technologies need to be open for new ideas, ways of

thinking and able to have a idea about the future development in order to find flexible, adaptable solutions that will meet the requirement not only now but also in the future. The following sections present some new and emerging technologies and areas that the authors' find interesting and relevant for civil engineering structures. A few technologies are then presented in more detail.

1.3.2 Fiber-Optic Sensors (FOS)

The development that made possible to optical communication was first, the invention of the laser in 1960; and second, the invention of optical fiber. Anyhow, the first optic fibers had a lot of losses and it took some years before the discovery of low-loss silica-glass fiber led to the technology of fiber-optic communications. Telecommunication systems made fiber optics familiar to everybody. The use of fiber-optic applications in different kinds of engineering fields made also a huge expansion in the last decades, especially in communications. Fiber-optic sensory technology provides accurate, wide bandwidth measurements of various parameters in diverse harsh environments where the traditional techniques might fail. Many sensors do not need calibration, have no drift, allow either parallel or serial multiplexing, perform static and/or dynamic measurements and are insensitive to electromagnetic interference.

1.3.3 The Global Positioning System (GPS)

U.S. Department of Defense started the project for Global Positioning System in early 1970s to overcome the limitations of previous navigation systems. GPS system was created and realized, and it is a space-based global navigation satellite system that provides reliable location and time information in all weather and at all times and anywhere on or near the Earth. In order to work properly, GPS needs an unobstructed line of sight to four or more GPS satellites. The system that was developed as a military system became fully operational in 1993, and it is maintained by the U.S. Government (El-Rabbany, 2002). Nowadays, it is also available for civil use and civil engineering needs. Dynamic deformation monitoring of structures, such as long bridges, towers, and tall buildings, is now possible using GPS technology. GPS can also monitor a wide variety of other structures and geologic features: dams, landslides, platforms, slopes, pipelines, volcanoes, retaining walls, constructions zones and railways, just to mention some.

1.3.4 Microelectromechanical Systems (MEMS)

Microelectromechanical Systems is the technology of the very small devices or systems that combine electrical and mechanical components (EMPA, 2004; Leondes, 2006). These devices usually range in size from a micrometer to a millimeter. They are fabricated using modified silicon fabrication technology, molding and plating, wet etching and dry etching, electro discharge machining and other technologies capable of manufacturing very small devices. A particular system can accommodate from a few to millions devices. These systems can sense, control and activate mechanical processes on the micro scale. On the macro scale they are able to function individually or in arrays.

Accelerometers, inertial sensors, optical scanners, fluid pumps, chemical, flow, humidity, temperature, and pressure sensors are a few application examples. Accelerometers based on MEMS technologies are fairly new but perhaps the simplest

MEMS device possible, consisting of little more than a suspended cantilever beam or seismic mass with some type of deflection sensing and circuitry. MEMS Accelerometers are available in single axis, dual axis, and three axis models and in wide variety of ranges. Use of these accelerometers in the automotive industry has pushed their cost down radically and they are now available on the market for civil engineering monitoring purposes. A SHM project with MEMS Accelerometers can be seen in Wiberg (2006).

MEMS industry is still a young industry. MEMS will certainly invade more and more area as new products are invented, and frequency response and sense range are getting wider. Sensors based on MEMS technology are also getting more reliable, and their sensitivity is better and even the price is dropping down.

1.3.5 Corrosion Monitoring

Reinforced concrete (RC) structures and steel structures in aggressive environments are exposed to chemical attacks. For that reason, SHM techniques for evaluating the condition of these structures are essential. Maintenance and repair may be very expensive: if errors can be identified at an early stage of their occurrence, a lot of capital can be saved. Corrosion monitoring has many applications in the construction and maintenance of civil structures. High-rise buildings, bridges, dams, spillways, flood control channels, tunnels, piers, pylons, and harbor constructions are continuously monitored.

Parameters to be measured in RC structures are chloride concentration, resistivity, and temperature. The market provides with new innovative devices and methods, and a lot of research is ongoing as corrosion cost a lot of money for governments. The movement of aggressive substances in concrete occurs due to differentials in humidity, ionic concentration, and pressure and temperature within the microstructure of concrete cover (Srinivasan et al., 2009). The moisture content and temperature of the concrete can be monitored during the curing process to ensure maximum strength of the concrete (Dunn et al., 2010). Once construction is complete, the instrumentation can be used to conduct long-term monitoring of corrosion conditions over time.

1.3.6 B-WIM, WIM

Weigh-in-motion or weighing in motion (WIM) devices capture detailed data for each individual vehicle as vehicles drive over a measurement site. The following parameters can be measured dynamic weights of all axles, gross vehicle weights, axle spacing, vehicles distance and speed, vehicle classification according to various schemes, and statistic representations for all types of traffic parameters.

Modern WIM systems are efficient as they are capable of measuring at normal traffic speeds. Bridge weigh-in-motion (B-WIM) is the process by which axle and gross vehicle weights of trucks traveling at highway speeds can be determined from instrumented bridges. B-WIM systems are performed by attaching strain transducers to the soffit of a bridge and placing sensors for detecting axles in order to provide information on vehicle velocity, axle spacing and position of each vehicle. This is called nothing-on-the-road (NOR) or free-of-axle detector (FAD). B-WIM system and a wide range of field trials have been completed in recent years. These systems are becoming increasingly accurate, and they are exceptionally durable as no contact with tires is required (Obrien et al., 2008).

1.3.7 Nondestructive Testing (NDT)

Nondestructive testing is a commonly used tool in several engineering fields as well as in medicine and arts. McCann and Forde (2001) give an overview to NDT methods as applied to the civil engineering. The basic principles of NDT methods are described, and the main NDT methods used in engineering investigations are discussed. In addition, brief case histories from the literature are illustrated.

NDT methods can be divided into following main categories: ultrasonic, magnetic-particle, liquid penetrant, radiographic, remote visual inspection (RVI), radar-based non-contact applications, eddy-current testing, and low-coherence Interferometry. Methods such as ultrasound, radiography, impact echo, rebound hammer, photogrammetry, and crack detection with help of very fine ferromagnetic particles are also widely adopted by industry and engineers. Steel industry, for instance, use these methods for testing on welding, homogeneity of material, and so on, and these testing methods are also described in the handbooks.

1.3.7.1 Noncontact 3D Laser Scanning Noncontact three-dimensional laser scanners are a tool for capturing geometrical information of the objects (Feng, 2001). This technique that was initially developed for car manufacturing analyzes objects or environment to collect data on its shape and possibly its appearance (i.e., color). It is widely used in industrial applications as well as in rock and bridge engineering. Use of three-dimensional lasers has increased in the untraditional fields in recent decades, and modern three-dimensional laser scanners provide improved resolution, high-measurement accuracy, high scanning speed as well as lower cost. They can also accommodate integrated camera and laser plummet. Adapted software converts the stored data to useful information such as digital, three-dimensional models. Several companies worldwide provide this technique in various performances. These existing systems in the market are based on three scanning principles, for example, triangulation, pulse-based, and phase-based techniques.

A high speed phase-based three-dimensional laser scanner consists of two major components: the single point laser measuring system and the mechanical beam reflection system. A scan takes just few minutes as the system is developed for high-speed, high-performance, and eye-safe scanning tasks. Sampling rate up to hundreds of thousands of points per second can be achieved and the system works both in indoor and outdoor environments. Scanning ranges are from around 0.1 m to about 50 m in average lasers and up to few 100 m for the latest technology. The achieved measurement accuracy is in millimeter range and can also be performed in darkness. An interesting study of development of a prototype system that combines the noncontact measurement technologies of photogrammetric imaging and three-dimensional laser scanning to create dimensionally accurate and pictorially correct three-dimensional models and orthoimages of a rock slope can be seen in Kwong et al. (2007).

Optical technologies encounter many difficulties with shiny, mirroring, or transparent objects. The reflectivity of an object is based on the object's color or reflecting power of a surface. Transparent objects such as glass will only refract the light and give false three-dimensional information. Shiny objects can be scanned by covering them with a thin layer of white powder. A white surface reflects a lot of light and more light photons will reflect back to the scanner.

1.3.7.2 Infrared Thermography Infrared thermography or thermal imaging is example of infrared imaging science. Thermal images are visual displays of the amount of infrared energy emitted, transmitted, and reflected by an object. Thermography allows fast scanning of objects and produces immediate images in real time. It is a noncontact and remote measurement that can be used on both still and moving objects.

Thermography is used in many different areas as construction industry, condition monitoring and predictive maintenance industry. It is a safe and convenient method that allows large areas of structures to be surveyed quickly, remotely, and cost effectively. Improved camera and software technology have supported the use in many various applications within the construction industry. Detection of hidden structures, deterioration, moisture, and heat losses can also be performed.

A thermal imaging camera detects and measures invisible infrared energy being emitted from an object. Data are processed and images of that radiation, called thermograms with variations in temperature can be produced and plotted. It is very easy to produce these thermograms, but the interpretation of the information gathered takes both education and experience in order to perform proper analysis.

This noncontact method technique is very interesting and promising. Improved manufacturing efficiencies and product quality emerge new applications for Infrared Thermography. More information can be seen in the website of the Institute of Infrared Thermography (<http://www.infraredinstitute.co.uk/index.html>).²

1.3.7.3 Acoustic Emission A structure starts to deform elastically when it is applied to a load; either by internal pressure or by external mechanical loading. Thereby, the stress distribution and storage of elastic strain energy in the structure changes. Acoustic emission (AE) (Jaffrey, 1982) is a naturally occurring phenomenon that takes place and generates elastic waves with these before mentioned loading conditions that relate to rapid release of energy. It detects ultrahigh frequency sound that stressed materials release. Acoustic emission monitoring is classified as a passive nondestructive testing method and AE tests can be used to evaluate the structural integrity of a component or a structure, structural damage diagnosis, lifetime assessment and SHM.

In AE monitoring, a transient elastic wave generated within a material by rapid release of energy is electronically monitored. Defects are detected and located in real time while the phenomena are taking place and instant action can be taken in order to save resources and provide safety.

A lot of development took place in the past decades. This technique is reliable and there is genuine long-term practice (Miller and McIntire, 1987) to monitor material flaws, corrosion, leakage in tanks, pressure vessels, piping systems, and steam generators. Anyhow, AE monitoring have entered civil engineering field; and bridges, offshore structures, heritage structures, dams, and many other structures are monitored continuously. Phenomenon that is studied nowadays with AE monitoring is corrosion, occurrence, and extension of fatigue cracks, fiber breakages in composite materials or fiber breakages in bridge main cables, stay cables or pre-stressed cables as well as cracking in concrete or reinforce concrete members. This is also a perfect method to verify post-tensioning in a structure.

1.3.7.4 Ground-Penetrating Radar GPR is an accurate geophysical method. It uses electromagnetic waves to map the spatial extent of near-surface objects, interfaces, or

²Accessed March 1, 2011.

changes in soil media and produce images of those features. Radar waves are propagated in distinct pulses from a surface antenna, reflected off buried objects, features or bedding contacts in the ground, and detected back at the source by a receiving antenna. As radar pulses are being transmitted through various materials on their way to the buried target feature, their velocity changes, depending on the physical and chemical properties of the material through which they are traveling. When the travel times of the energy pulses are measured and velocity through the ground is known, distance (or depth in the ground) can be accurately measured. A three-dimensional data set is then produced. In the GPR method, radar antennas are moved along the ground in transects, and two-dimensional profiles of a large number of periodic reflections are created. There are now a number of commercially available equipments for civil engineering purposes (Conyers, 2002).

1.3.7.5 Remote Sensing Remote sensing is technology of obtaining reliable information on a given object or area either wireless, or not in physical or intimate contact with the object. Any form of noncontact observation can be regarded as remote sensing. Microwave Interferometry and Photogrammetry are good examples of remote sensing and presented latter.

1.3.8 Interferometric Radar

Interferometric radar is a pioneering revolutionary technology in the domain of geodetic measurements that is now spreading out to other areas such as civil engineering. The measurement device is coherent radar generating, transmitting, and receiving the electromagnetic signals to be processed in order to provide movement and deformation measurements (Gentile, 2010). Both static and dynamic measurements of structures can be performed. This noncontact method to measure objects distances up to kilometers is very convenient for many applications such as stay cables and main cables in bridges. No instrumentation is needed, traffic can continue and the method saves time, money, and resources.

1.3.9 Photogrammetry

Geometric properties of object can be determined from photographic images; this practice is called photogrammetry (Mikhail et al., 2001). The American Society for Photogrammetry and Remote Sensing (ASPRS) founded in 1934 is a scientific association. Their mission is to advance knowledge and improve understanding of mapping sciences to promote the responsible applications of photogrammetry, remote sensing, geographic information systems (GIS), and supporting technologies. They do define photogrammetry as follows (<http://www.asprs.org/>)

Photogrammetry is the art, science, and technology of obtaining reliable information about physical objects and the environment, through processes of recording, measuring, and interpreting images and patterns of electromagnetic radiant energy and other phenomena.

Photogrammetry is used in different fields: topographic mapping, architecture, engineering, manufacturing, quality control, archeology, meteorology, and geology; and it enables producing plans of large or complex sites very effectively. It can be divided to aerial photogrammetry, close-range photogrammetry or Stereophotogrammetry.

In aerial photogrammetry, a camera is installed on an aircraft and multiple overlapping photos of the ground are taken in order to create two-dimensional or three-dimensional models from aerial photographs. Close-range photogrammetry cameras are used to model buildings, engineering structures, vehicles, forensic, and accident scenes. The camera is normally set close to the subject and is typically hand held or on a tripod and the produced output is three-dimensional model or a drawing. In Stereophotogrammetry three-dimensional coordinates of points on an object are estimated. Two or more photographic images taken from different positions are required. Common points are identified on each image and a line of sight can be constructed from the camera location to these points on the object. Triangulation is used to establish three-dimensional locations of the points (Ackermann, 1984). A lot of information such as publications, and so on, can also be found on the web of the International Society of Photogrammetry and Remote Sensing (ISPRS) (<http://www.isprs.org/>).³

Monitoring of crack origin and evolution with photogrammetry can be seen in Benning et al. (2004).

1.3.10 Smart Technical Textiles

Technical textiles are frequently used in civil engineering and geotechnical applications for reinforcing, repairing, or retrofitting purposes (Veldhuijzen Van Zanten, 1986). Typical applications in civil engineering domain include repair of damaged parts of structures (e.g., cracking of the bridge deck), retrofitting of seismically weak structure (e.g., old masonry structure). Typical geotechnical applications include reinforcing bearing capacity of soils beneath foundations (e.g., dams, dikes, tunnels) and stabilization of landmasses prone to subsidence or sliding (e.g., to prevent land sliding, or creation of sinkholes).

Once the technical textile is applied to the host structure, there is a need to evaluate both its performance and the performance of the host structure in long term, in order to assess effectiveness of the technical textile application and the improvement made to the host structure. It is important to detect any deterioration in the performances, since it may lead to failure of the structure. Two parameters that are of particular interest for assessment the condition of civil structures are strain and temperature. In geotechnical applications, besides these two parameters, monitoring water pressure and leakage is important for assessment of condition of reinforced soils and foundations.

By embedding a monitoring system in the technical textile, the latter is transformed into an innovative intelligent multifunctional material—smart textile—that simultaneously provides both reinforcing and monitoring capabilities. An integration of sensors into the technical textile is functionally beneficial for both parties. The sensors provide with assessment of performances of the technical textile, whereas the latter provides for protection and an easy and practically inexpensive installation of the sensors, because they were integrated in the technical textile during the production. Thus, research is being performed in order to create a multifunctional technical textile with monitoring capabilities (e.g., Messervey et al., 2010).

Various sensing technologies can be embedded in the technical textile; however, the fiber-optic sensors seem to yield the most promising results. Sensors based fiber Bragg gratings (FBG) are particularly suitable for embedding in layers of technical textiles used to repair and retrofit civil structures such as bridges and buildings

³ Accessed March 1, 2011.

(Messervey et al., 2010), owing to their small size and both serial and parallel multiplexing capability. Distributed sensing technologies are particularly suitable for embedding in technical textiles used in geotechnical applications, since they provide with coverage of long lengths (up to several kilometers), which are characteristic for geotechnical structures (Belli et al., 2009).

Besides the developments in domain of traditional, glass fiber-based sensors, research is ongoing in domain of plastic optical fiber based sensors too (Liehr et al., 2009). Although the plastic fiber features significantly higher losses, and consequently can be applied to cover limited length (few hundreds of meters), they are highly deformable (few tens of percents of the original length), which makes them particularly suitable for monitoring large deformations such as found in geotechnical applications.

The first smart technical textile based on Brillouin scattering in glass optical fibers appeared on the market and they are commercially available. The research is ongoing, and various new products are expected emerge in the next decade.

1.3.11 Specific Issues Around Usage of New Technologies

Structures such as wind turbines are increasing, they are often located in remote areas with difficult or no access at all. They do require new kind of solutions in order to be monitored. Lot of research is ongoing in the field and solutions can be adapted from other areas like from offshore structures that also have no easy accessibility.

Other important structures and structural components difficult to access are the insides of main cables and stay cables in bridges, heritage structures in poor condition, residential areas of high-rise buildings, surfaces of tunnel structures, and so on. Sensors can be required implemented by manufacturers when producing components, as installation is difficult or impossible afterwards.

New technologies are merging the field and techniques from other engineering fields as well as monitoring methods are adapted to civil engineering purposes. These applications might though lack long-term experience; therefore, it is good to perform proper investigations when choosing the technology so that capital is not invested in nonproven technology. Small-scale test in the field is sometimes appropriate to prove, if the technique is feasible in the given circumstances. It is also very important that the installation is done in a proper way; otherwise, there is a risk to jeopardize the function of the whole system.

1.3.12 Chosen Technologies and Motivation

Many large construction projects were instrumented with thousands of sensors in the last decades. New technologies entered the field and a lot of heuristic knowledge was gained when working with new technologies, both in theory and in practice.

FOS appears to have a bright future, and a lot of research has been going on presently. Corrosion is present in reinforced concrete and steel structures; therefore, better methods for corrosion monitoring need to be found in order to provide long lasting structures and to save capital. New NDE techniques offer for straightforward monitoring and testing of structures. True information about old structures can be recorded in order to avoid uncertainties and malfunctions.

First, the chosen technologies and areas of interests are FOS, acoustic emission, radar technology, WIM applications, and GPS, as they have shown their adaptability to civil engineering structures in recent decades.

And second, corrosion monitoring as this issue is present in nearly every structure and huge amounts of capital can be saved if corrosion problems can be solved. New developments and the chosen technologies that have potential for SHM of civil structures are discussed below.

This chapter does not discuss wireless sensing technologies as wireless structural monitoring systems are still in their infancy. A detailed overview in designing wireless sensing units for the health monitoring of civil structures can be seen in Lynch (2007). The chapter has reviewed the design of a number of different wireless sensing unit prototypes that have been designed explicitly for structural monitoring applications. The firmware required to operate the units and to locally interrogate structural response data and a case study is presented, and a novel wireless active sensing unit design is proposed.

1.4 FIBER-OPTIC TECHNOLOGY

1.4.1 General

An optical fiber is a thin, transparent fiber, usually made of fused silica for transmitting light over large distances with very little loss. The diameter of the optic fiber is similar to that of a human hair and the core of it serves to guide the light along the length of the optical fiber. There are both single mode and multimode fibers: the core of the single mode fiber is very small, 5–10 μm , whereas core of the multimode fiber is around 50 μm . The core is surrounded by cladding with slightly lower index of refraction than the core. The purpose of the cladding is to minimize the losses and also physically to support the core region as the light propagates in the fiber. Optical fibers operate over a range of wavelengths; 1310 and 1550 nm is common for single mode fibers with minimal losses and 850 and 1300 nm for multi mode fibers.

Development and research is ongoing; existing technology is improved and new products appear on the market continuously. Clear and detailed descriptions over the different fiber-optic technologies can be seen in Measures (2001). The book is highly relevant still today and a perfect introduction into fiber-optic sensors, interrogators, and related aspects.

Several companies working with fiber-optic sensors were founded in the 1990s in collaboration with universities. Several doctoral theses, as well as books were published about various kinds of fiber-optic systems and SHM related topics. Fiber-optic sensors, smart materials, and structural technology were developed, and examples can be seen in Udd (1991, 1995), Inaudi (1997), Vurpillot (1999), Glisic (2000), Clark et al. (2001), Glisic and Inaudi (2007), and Imai et al. (2009).

The high performances of the fiber-optic sensors (FOS) are intrinsically linked to the optical fiber itself. The optical fiber can be used for both sensing and signal transmission purposes. The silica of which the optical fiber is composed is an inert material, which is resistant to most chemicals in wide range of temperatures and is, therefore, suitable for applications in harsh chemical environments (Udd, 2006). Various packaging especially designed for field applications made fiber-optic sensors robust and safe to use even in very demanding environments (Udd, 2006; Glisic and Inaudi, 2007).

The light used for sensing purposes in the core of the optical fiber does not interact with any surrounding electromagnetic (EM) field. Consequently, the fiber-optic sensors are intrinsically immune to any EM interference (EMI), which contributes significantly to their long-term stability and reliability. The ability to measure over distances of several

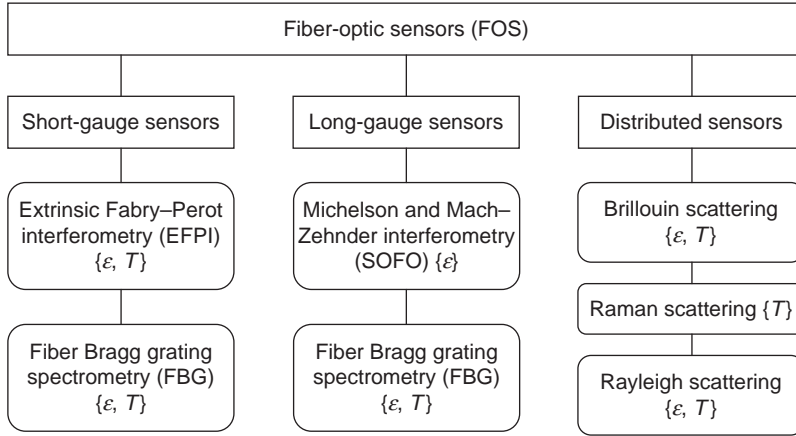


FIGURE 1.1 Division of FOS based on gauge length and functional principle.

tens of kilometers without the need for any electrically active component is an important feature when monitoring large and remote structures, such as landmark bridges, dams, tunnels, and pipelines (Udd, 2006).

Fiber-optic sensors cover large spectrum of parameters that can be monitored (e.g., strain, inclination, temperature, humidity); thus, multiple parameters can be combined on the same network (e.g., Del Grosso et al., 2005). Compared with conventional electrical sensors, fiber-optic sensors offer two new and unique sensing tools: long-gauge strain sensors and truly distributed strain and/or temperature sensors. The former can be combined in topologies that allow for global structural monitoring while latter allows for one-dimensional strain field and integrity monitoring.

As the area of fiber-optic sensors is really wide, it is not possible to present every single technique. Sensors that are generally used in civil engineering applications and their division based on gauge length and the functional principle are presented in Figure 1.1.

Recent significant developments in the optical telecommunications market helped reduce the cost of the FOS, which is still higher compared with conventional sensors, but, however, affordable and justified by superior long-term performance of the FOS. In this section, several FOS techniques are presented, along with their typical applications and summaries of the best performances. Many sensors like the ones based on fluorescence are mostly used for medical and chemical applications and are beyond the scope of this chapter.

1.4.2 Sensors Based on Sagnac, Michelson, and Mach–Zehnder Interferometers

Sagnac interferometer, also called as a fiber-optic gyroscope (FOG) is a gyroscope that uses the interference of light to detect mechanical rotation. This is very developed sensor and principally used for rotation rate measurements as well as in aerospace engineering. Two counter propagating light beams travel along the fiber inside an interferometer in opposite directions. As the path is closed and same for both beams, the beam traveling against the rotation experiences a slightly shorter path than the other beam. This is called Sagnac effect. This sensor can also be used in dynamic measurements but is not well known in commercial civil engineering applications.

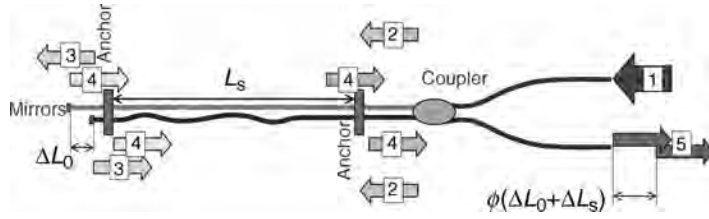


FIGURE 1.2 Schematic representation of standard SOFO sensor.

Both Michelson and Mach–Zehnder interferometers are easy to understand and manufacture. They are pretty similar in appearance: two fiber beams are inserted with light and then a relative phase shift between these beams is measured in order to measure eventual length difference. In the Michelson interferometer, the end of the fiber paths includes chemical mirrors that reflect the light back to the photo detector. Mach–Zehnder interferometer has two outputs instead of the mirrors and the light passes the fiber paths and is then measured by a photo detector. Michelson interferometer-based decoder allows for very stable, long-term static measurements, whereas Mach–Zehnder interferometer allows for very sensitive short-term dynamic measurements. A detailed description of Michelson Interferometer that is commercially available is given in the following subchapter.

1.4.2.1 SOFO Michelson interferometer that is proven in large number of projects (Glisic et al., 2010a) is called SOFO (French acronym for Structural Monitoring using FOS, Inaudi, 1997). The standard SOFO sensor comprised two zones: the active zone that measures the deformations, and the passive zone that serves as the carrier of information between the active zone and the reading unit. The sensor is schematically represented in Figure 1.2.

The active zone is limited by two anchor pieces and consists of two optical fibers placed in a protection tube. The anchor pieces have a double role: to attach the sensor to the monitored structure and to transmit the deformation from the structure to the sensing fiber. The measurement fiber is pretensioned between the anchor pieces in order to measure the shortening of the structure as well as its elongation. The reference fiber is independent of both the measurement fiber and the deformation of the structure, and its purpose is to compensate for temperature changes. Both fibers have mirrors silvered at their extremities.

The passive zone transmits the information from the active zone to the reading unit. It comprised one single-mode fiber, connector, and coupler, which are all protected by a plastic tube. The coupler is placed in the passive zone of the sensor, close to the anchor piece in order to increase the precision and to facilitate the manipulation during the measurement.

The light inserted in the passive zone is split into two fibers of active zone, travels to the extremities and reflects back off the mirrors. Since two fibers have in general different length, a shift in phase is created between the two reflected lights, and this shift in phase is proportional to the length difference in the optical fibers. The deformation of monitored structure will be transferred only to measurements fiber while the reference fiber remains unchanged. As a consequence, the shift in phase will change. The shift in phase is converted to length difference in the reading unit, which contains decoder in form of Michelson interferometer.

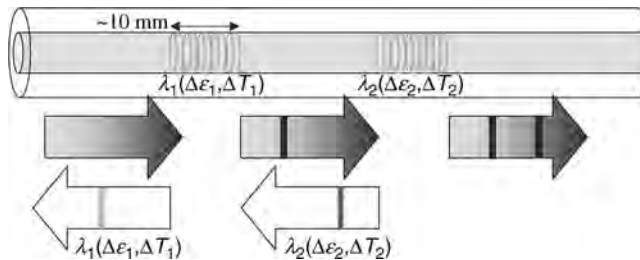
TABLE 1.1 The Best Performances of Discrete FOS

	Interferometric (SOFO)	Fiber Bragg Gratings (FBG)	Fabry–Perot (EFBI)
Gage length	250 mm–20 m	10 mm–2 m	51–70 mm
Multiplexing	Parallel	In-line and parallel	Parallel
Maximum number of sensors in the network	Static: unlimited; dynamic: 8	16 channels 5–10 sensors/channels	32
Stability	Static: long term	Long term	Long term
Resolution	Static: 2 μm ; dynamic: 10 nm	0.2 me	$\pm 0.01\%$ full scale
Repeatability (precision)	Static: <0.2% of measured value	<1 $\mu\epsilon$	N/A
Dynamic range	Static: –5,000 to +10,000 $\mu\epsilon$; dynamic: ± 5 mm	–5,000 to +7,500 $\mu\epsilon$ packaging dependent	$\pm 3,000$ $\mu\epsilon$
Temperature sensitivity	Self-compensated	Compensation needed	Insensitive or packaging dependent
Maximum measure frequency	Static: 0.1 Hz; dynamic: 10 kHz	0.5 MHz	20 Hz

The SOFO sensor is true long-gauge sensor, because the light “integrates” the strain along its gauge length, which is typically limited between 20 cm and 10 m. The best performances of the SOFO system are given in Table 1.1, and applications can be seen in Enckell and Larsson (2005) and Enckell (2006).

1.4.3 Sensor Based on the Fiber Bragg Gratings

Fiber Bragg gratings (FBG) consist of periodical changes created in fiber core by appropriate exposing to ultraviolet light (Dadpay et al., 2008). If the light containing certain range of wavelengths is inserted into the fiber with an FBG, the latter will reflect back one specific wavelength and let pass through all the other wavelengths, as shown in Figure 1.3. The specific wavelength that is reflected depends on optical properties of the FBG that were preimposed by manufacturing process. The optical properties of the FBG depend linearly on strain and temperature. Increase in one or both of these parameters will

**FIGURE 1.3** Principle of FBG sensors.

practically change the wavelength that is reflected back to the light source. By determining the difference from initial wavelength it is possible to determine strain or temperature in the FBG. Since the change in wavelength depends on both strain and temperature simultaneously, strain sensors must be compensated for temperature using an appropriate procedure. The typical length of an FBG is several millimeters, thus, depending on packaging, they can be used either as short-gauge sensors (fiber-optic equivalent for classical strain-gauges) or as long-gauge sensors (e.g., if the fiber with FBG is pretensioned between two anchoring points). There is a large variety of both types of the sensors available on the market.

The source and decoder for FBG-based sensors are usually combined in the same device. The source is able to generate desired range of wavelength, while decoder measures intensity of reflected light and determines reflected wavelength using tunable laser with wavelength filter (e.g., Fabry–Perot cavity) or spectrometer. Both static and dynamic measurements are possible, depending on type of the reading unit.

An important advantage of FBG-based sensors is that several gratings with different specific wavelengths can be placed along a single fiber, allowing an easy multiplexing. Thus, several sensors can be read from a single channel on the reading unit. The total number of sensors that can be placed along single line depends on range of strain and temperature changes in the monitored structure. For typical civil engineering applications, involving steel or concrete structure, this number varies between 5 and 10. The best performances of the FBG sensors are given in Table 1.1.

1.4.4 Sensors Based on Fabry–Perot Interferometry

The previously presented sensors, SOFO and FBG, are called intrinsic sensors, because the optical fiber is used as the sensing body. Fabry–Perot interferometric sensors can be either intrinsic or extrinsic, but latter was proven to be more effective and reached market maturity, and that is why they are presented in this section. Schematic principle of functioning of an extrinsic Fabry–Perot interferometric (EFPI) sensor is shown in Figure 1.4.

The sensor consists of lead fiber and target fiber, both cleaved at 90° and both with partially reflective surfaces. There is a few millimeters air gap between the cleaved surfaces. The broadband light is sent from the source (multimode optical fiber is used as opposed to single-mode used in the case of SOFO and FBG sensors) through lead fiber. When the light reaches the cleaved end of the lead fiber a part of it is reflected back to the reading unit, and the other part passes through the air gap, reflects off the surface of the target fiber, and re-enters in the lead fiber. Two lights are then combined in the lead fiber into an optical signal (constructive or de-constructive interference of several wavelengths)

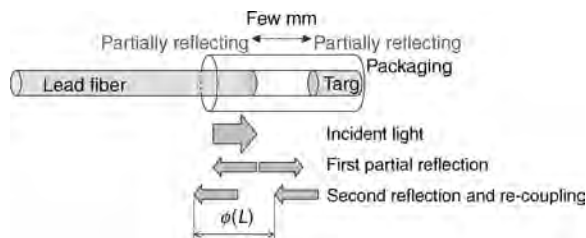


FIGURE 1.4 Schematic principle of EFPI sensors.

that contains information about the size of the air gap (Krohn, 2000). This information is finally decoded in the reading unit.

The lead fiber is connected to target fiber using a packaging with purpose is to couple the sensor to the structure. Any deformation of the structure will result in the change of the size of the air gap that is sensed by optical signal and determined in the reading unit.

The great advantage of the EPFI principle is its insensitivity to temperature changes. However, depending on the gauge length of the sensor and the position of the anchoring points between the packaging and the optical fibers, sensitivity to temperature is often present due to thermal expansion of the mechanical components of the sensor, and in general the sensors should be compensated for temperature.

Besides the strain, the EFPI sensors can be packaged so that they can sense temperature and pressure. They can be made extremely small (practically the size of optical fiber is the lower size limit) in and they found many applications in biomedical domain.

In civil applications, the EFPI sensors face some limitations regarding multiplexing and remote position of the reading unit; however, the complete system for smaller projects or for manual measurements is less expensive than SOFO and FBG systems. The best performances of EFPI sensors are shown in Table 1.1.

1.4.5 Best Performances of Discrete FOS

The best performances of discrete FOS are for comparison given in Table 1.1. These performances were extracted from commercial data sheets and web page of various companies.

1.4.6 Distributed Sensors

Distributed sensor (or sensing cable) can be represented by a single cable, which is sensitive at every point along its length. Hence, one distributed sensor can replace large number of discrete sensors. A distributed sensor requires single connection cable to transmit the information to the reading unit, instead of a large number of connecting cables required in case of wired discrete sensors. Finally, distributed sensors are less difficult and more economic to install and operate. An illustrative comparison between pipelines equipped with distributed and discrete sensors is shown in Figure 1.5 (this schematic drawing does not refer to real case, e.g., redundancy is not included).

Although a distributed sensor is sensitive to measured parameters (strain and/or temperature) at every point of its length, it delivers measurements at discrete points that are spaced by a constant value, called the sampling interval, and the measured parameter is actually an average strain measured over a certain length, called the spatial resolution (Lanticq et al., 2009). The spatial resolution can be considered as the gauge length over which the measurement is made. Sampling interval and spatial resolution are two parameters characteristic for the distributed systems. They can be set by the user depending on application. However, they are correlated with resolution and precision of measurement, and the time of measurement, and the trade-offs must be made. For example, a measurement that is very precise, and sensitive (small spatial resolution), are slow, while a fast measurement affects one or both other parameters.

There are three main principles for distributed sensing in the domain of FOS: Rayleigh scattering (e.g., Posey et al., 2000), Brillouin scattering (e.g., Karashima et al., 1990), and Raman scattering (e.g., Kikuchi et al., 1988). Each technique is based on the relation

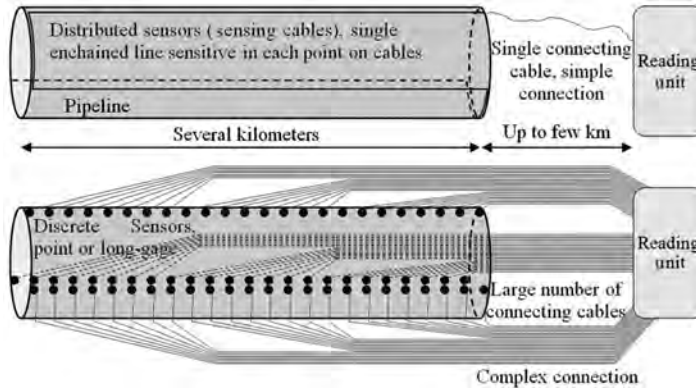


FIGURE 1.5 Distributed versus discrete monitoring; schematic comparison (does not refer to real case).

between the measured parameters, that is, strain and/or temperature, and encoding parameter, that is, changes in optical properties of the scattered light. This is schematically presented in Figure 1.6.

Rayleigh scattering can be used for both strain and temperature monitoring. It is based on the shifts in the local Rayleigh backscatter pattern, which is dependent on the strain and the temperature. Thus, the strain measurements must be compensated for temperature. The main characteristics of this system are high resolution of measured parameters and short spatial resolution, but the maximal length of sensor is limited to 70 m (Lanticq et al., 2009). Thus, this system is suitable for monitoring of localized strain changes over relatively short distances. Best performances achievable in strain monitoring using the Rayleigh scattering are given in Table 1.2.

Brillouin scattering can also be used for both strain and temperature monitoring. It is based on the change in frequency of Brillouin scattered light, which is dependent on the strain and the temperature. Thus, as in case of the Rayleigh scattering, the strain measurements must be compensated for temperature, and temperature measurements are to be performed with sensors containing a loose (strain-free) optical fiber. Both spontaneous (Wait and Hartog, 2001) and stimulated (Nikles et al., 1994, 1997) Brillouin scattering can be used for sensing purposes. Monitoring system

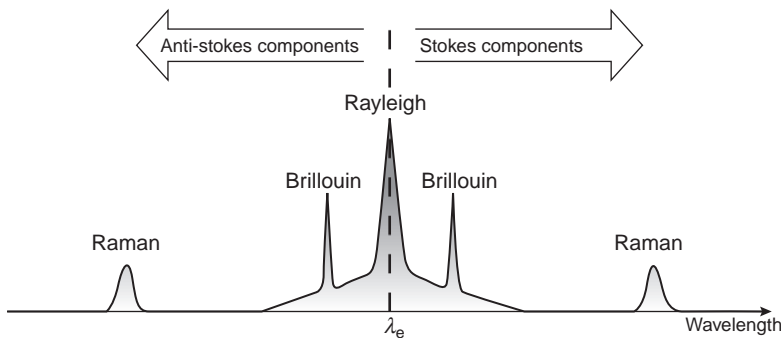


FIGURE 1.6 Scattered light properties as encoding parameter for strain and/or temperature measurements (Glišić and Inaudi, 2007, courtesy of SMARTEC SA).

TABLE 1.2 Comparison of Best Performances Achievable in Strain Monitoring Using Distributed Systems

	Brillouin, Stimulated	Brillouin, spontaneous	Rayleigh
Spatial resolution	0.5–5 m	1 m	10 mm
Sampling interval	100 mm	50 mm	10 mm
Maximum no. of sensing cables in network	16	N/A (1?)	N/A (1?)
Stability	N/A	N/A	N/A
Resolution	2 $\mu\epsilon$	30 $\mu\epsilon$ (“accuracy,” 2 RMS)	1 $\mu\epsilon$
Repeatability	N/A	<0.02%	N/A
Instrument range	$\pm 30,000 \mu\epsilon$	10,000 $\mu\epsilon$	$\pm 7,000 \mu\epsilon$
Maximum length of sensor	250 m to 5 km (up to 20 db)	N/A	70 m
Temperature sensitivity	Compensation needed	Compensation needed	Compensation needed
Measurement speed	10 s–15 min	4–25 min	4 s

based on stimulated Brillouin scattering is less sensitive to cumulated optical losses that may be generated in sensing cable due to manufacturing and installation, and allows for monitoring of exceptionally large lengths (Thevenaz et al., 1998), for example, in the case of strain monitoring, a single reading unit with two channels can operate measurement over lengths of 10 km, while in the case of temperature monitoring, the lengths of 50 km can be reached. Remote modules can be used to triple the monitoring lengths. The measurement specifications of Brillouin-based measurements are not as good as these of Rayleigh-based measurements; however, the great advantage of Brillouin-based systems is significantly longer length of the sensor (several kilometers). Thus, the Brillouin-based systems are better suited for monitoring global strain changes over large distances. Applications and different methods of Brillouin monitoring can be seen in Nikles et al. (2004) and Inaudi and Glisic (2005, 2006). The best performances in strain monitoring achievable with Brillouin-based systems are given in Table 1.2.

Raman scattering is the result of a nonlinear interaction of the light traveling in the silica fiber core and can only be used for temperature monitoring. It is based on the change in amplitude of Raman scattered light, which is dependent on temperature only. The insensitiveness of this parameter to strain is actually an advantage compared with Rayleigh and Brillouin-based temperature monitoring, since no particular packaging of the sensor must be made to make sensing fiber strain-free. Typical spatial resolution of Raman systems is 1 m, and typical resolution is better than 1°C. Since the leakage of pipelines, dykes, dams, and so on, often changes thermal properties of surrounding soil, besides the temperature monitoring; the Raman-based systems are used for leakage monitoring in large structures.

There is a sever tradition of monitoring dam constructions with Raman systems and some examples can be seen in Aufleger (1998).

1.4.6.1 Crack Detection with Distributed Techniques Crack-related parameters: detection, identification, localization, and quantification, can be measured with

distributed fiber-optics techniques. Brillouin-based technology for crack detection and crack width estimation is tested and discussed by Bao et al. (2005), Imai et al. (2009), and Nöther et al. (2009). Shi et al. (2005) and Zhang and Wu (2008) do report about distributed sensing system based on Brillouin scattering for monitoring geotechnical engineering structures and large Earth structures.

Enckell et al. (2011) do report about evaluation of a large-scale bridge strain, temperature and crack monitoring with distributed fiber-optic sensors (Ravet et al., 2009). The main goals of the system are to detect and localize new cracks, measure high strain, and unusual strain behavior. Also new methods and procedures in installing, testing, modifying, and improving a SHMS were developed, tested and proven, both in laboratory and on-site. The system sends warning messages to the bridge authorities with well-specified scenarios (Myrvoll et al., 2009; Enckell et al., 2011). The system is installed over five 1-km long girders of the Gota Bridge, Gothenburg, Sweden (Glisic et al., 2007, 2009) and monitoring is ongoing and intended to the last about 10 more years, to the end of the bridge's lifetime. This is the first application of distributed systems on the bridges at such a large scale.

1.5 ACOUSTIC EMISSION

1.5.1 Theory of Acoustic Emission

When a material deforms and cracks, the subsequent atomic rearrangement produces transient elastic waves, which is known as acoustic emission (AE). Degrading mechanisms within materials, such as cracking, corrosion, and yielding, which produce AE signals, are called sources. AE signals contain information about the nature and the severity of the changes occurring within the material. Signals propagate through the material away from the source. AE technology is a passive technique that detects these signals using surface mounted transducers (sensors). Active defects are point sources, emitting nondirectional AE, which radiate out in a spherical wave front and detection is permitted by sensors placed anywhere on the structure. AE has the ability to locate the position of active defects, using two or more sensors. Location is achieved by identifying the time of arrival of an AE signal at different transducers. AE has progressed greatly over the past 30 years, with the increase in computing power allowing increased accuracy through data filtering and processing, which has seen the establishment of many commercial applications. Advantages of acoustic emission are high sensitivity; early and rapid detection of defects such as corrosion, flaws, wire break, and cracks; real-time monitoring of structures in-service; cost and time reduction; defect localization; global and local monitoring; monitoring of nonaccessible zones; defective area location; repeatable; and the evaluation is comprehensive and source detection; and it is not limited by defects geometry. Disadvantages are complicated signal analysis that requires highly trained and experience personal, low signal to noise separation and difficulty to separate external noise in some applications.

1.5.2 Sources of Acoustic Emission

Sources that emit detectable AE are numerous; in fact most materials emit AE when they are put under load. Common sources include the fracture of ice cubes in a drink or the

crunch of an apple being eaten. Acoustic emission can be generated by many mechanisms, which include

- Fracture of crystallites/inclusions
- Crack nucleation and propagation
- Micro-cracking including fracture of inclusions
- Phase transformations in solids (e.g., martensitic)
- Boiling and electrical discharge (electrical transformers)
- Matrix cracking (fiber-reinforced plastics)
- Delamination (concrete and fiber-reinforced plastics)
- The creation, annihilation of point, line or surface defects

Other forms of acoustic emission mechanisms commonly encountered include

- Mechanical sources, for example, pumps, actuators
- Joints and supports of structures, for example bridge expansion joints and bearings

Practitioners of AE try to differentiate between real damage sources and extraneous noise. Fundamental to the identification of signals of an individual application is the development of a database. A database contains examples of AE data from tests where the origin of the signals was confirmed either through visual inspection or other nondestructive testing. Using source confirmed AE examples, AE data taken in relatively blind conditions can be attributed to a certain type of source and be used to measure the severity/significance of damage.

1.5.3 The Development of Acoustic Emission in Industry and Civil Engineering

The technology was born in the early 1960s when it was recognized that growing cracks and discontinuities in fiber reinforced plastic tanks and pressure vessels could be detected by monitoring their acoustic emission signals. Measuring bridge fatigue crack growth in steel with acoustic emission can be seen as early on as 1972 (Pollock and Smith, 1972), with further fatigue work documented in later in the decade (Sinclair et al., 1977; Lindley et al., 1978). Great advancements were made from the 1990s by Cardiff Universities research into steel bridge fatigue using modern fast DSP AE systems coupled with neural network software (Carter and Davies, 1996; Watson et al., 2001; Pullin et al., 2007). AE studies of concrete structures were predominantly led by research and application in Japan, in particular by professor Ohtsu (1987) and professor Shigish and Dr. Yuyama (Ohtsu and Yuyama, 2000). This Japanese looked at the application of AE on a variety of common concrete construction elements; beams, columns, foundations and slabs. In situ tests were carried out on concrete docks, pylon foundations, bullet train bridges, tunnels, and an arch dam. The ability of acoustic emission to detect wire break has long been recognized (Holford, 1996) and is now commonly employed on both suspension bridges and post-tension concrete ([http://www.ndtnews.org.2008](http://www.ndtnews.org.2008;);⁴ http://www.mistrasgroup.com/news/itn/print/09012010_BDE.pdf).⁵

⁴Accessed February 25, 2011.

⁵Accessed January 27, 2010.

1.5.4 Acoustic Emission Systems

A typical AE system comprises a high speed digital signal processor (DSP), AE processing boards with individual processing channels for each sensor (i.e., a nonmultiplexing system) and the ability to program the settings for signal thresholds and frequency range to enable the AE signal to be filtered. It should also have software for source location in one, two, and three dimensions, feature extraction capability to allow characterization of the signals, and stable software for long-term monitoring. The acquisition of raw RF waveforms is a distinct advantage for source analysis. An AE signal is recorded when the signal reaching the sensor exceeds a threshold value, set by the operator in order to avoid recording unwanted noise (e.g., 30–45 dB is generally satisfactory for most bridges). AE sensors are piezoelectric crystals that convert movement (a variation of pressure) into an electrical voltage. The sensors must all have an identical response and they should be calibrated annually. They are normally held in place using metallic clamps for steel structures or bonded to concrete. These are connected to the AE system using coaxial cables with shielding to prevent electromagnetic interference. There are two types of AE sensor: resonant and broadband. Resonant sensors are more sensitive to sources at particular frequencies—thus, sensor spacing can be maximized—and they help minimize background noise by filtering out unwanted frequencies. Sensors with a low-resonant frequency are more susceptible to noise but high-frequency signals are readily attenuated which limits their use. A resonant frequency of 30–100 kHz is typical for concrete applications; whereas 100 and 200 kHz is used for metallic structures. Higher frequency sensors can be used in high noise environments but only for local monitoring due to the higher attenuation at these frequencies.

1.5.5 Codes, Standards, and Recommended Practice in Acoustic Emission

ISO standards exist for calibration of acoustic emission sensors

- ISO 12713 : 1998 (ISO, 1998)
- ISO 12714 : 1999 (ISO, 1999)

The American Society for Testing and Materials (ASTM) has several standards concerning acoustic emission technology and monitoring⁶

- ASTM E1106-07 (ASTM E1106-07, 2007)
- ASTM E1781-08 (ASTM E1781-08, 2008)

Example of European and British standards for AE can be seen in

- BS EN 1330-9:2009 Nondestructive testing terminology part 9: terms used in acoustic emission testing (British Standard, 2009)
- BS EN 13477-2:2001 Nondestructive testing—acoustic emission—equipment characterization (British Standard, 2001)
- BS EN 13554:2002 Nondestructive testing—acoustic emission—general principles (British Standard, 2002).

⁶Accessed January 27, 2010.

Procedures and recommended practices for steel and concrete structures monitoring by acoustic emission have been developed and can be seen in

- BA86/06 Advice notes on the nondestructive testing of highway structures—Section 3.3—acoustic emission by the UK highways agency defines a wide range of uses of AE and guidance for use on concrete and steel structures. Specific case studies are defined (Highways Agency, UK, 2006).
- Recommended practice for in situ monitoring of concrete structures by acoustic emission, by Ohtsu and Yuyama (2000) reports about AE monitoring for concrete structures.
- A proposed standard for evaluating structural integrity of reinforced concrete beams by acoustic emission by Yuyama et al. (1999) for evaluating structural integrity.
- Acoustic emission for bridge inspection, report no. FHWA-RD-94—prepared for FHWA and U.S. Department of Transportation, June 1995. This defines the use of AE for local area monitoring of steel defects (Physical Acoustics Corporation, 1995).
- Developments for the nondestructive evaluation of highway bridges in the United States (Washer, 1998).
- Development of an advanced structural monitoring system for fracture critical steel bridges (<http://www.lrrb.org/pdf/201039.pdf>).⁷

Golaski from Kielce University of Technology has also published an overview (Golaski et al., 2002) concerning AE monitoring of reinforced concrete bridges. AE studies of other concrete structures can be seen in Ohtsu (1987), Matsuyama et al. (1993), Colombo et al. (2002), and Watson et al. (2002). Publications from American universities around AE testing and monitoring can be seen in Fowler et al. (1998) and Hamstad and McColskey (1999). A study of stress corrosion cracking of high-strength steel used in prestressed concrete structures by AE can be seen in Ramadan et al. (2008).

1.6 RADAR TECHNOLOGY

1.6.1 General

RADAR is an acronym coined in the 1934 for radio detection and ranging (Buderi, 1996). The first ground-penetrating radar survey was performed in Austria in 1929 to sound the depth of a glacier (Stern, 1929, 1930). The technology was largely forgotten until the late 1950s when U.S. Air Force started investigations into the ability of radar to see into the subsurface. A system much like Stern's original glacier sounder was proposed, built, and sent on Apollo 17 to the Moon (Simmons et al., 1972) in 1972 to study the electrical and geological properties of the crust.

1.6.2 Ground-Penetrating Radar

Ground-penetrating radar (GPR) is one of the most inclusive archaeological geophysical methods because it collects large amounts of reflection data and produces massive three-dimensional databases. It can also locate objects and map features without any risk of

⁷Accessed January 27, 2010.

damaging them (Conyers, 2002). GPR is now available for other areas, and it is an effective nondestructive tool for structural analysis. It offers a set of unique solutions featuring dedicated high-frequency antenna configurations and three-dimensional software processing solutions. GPR can be used in various applications: utility detection and mapping, civil engineering, geology, geophysics, geotechnics and environment, archeological and cultural heritage, forensic, and security. Ground-penetrating radars for civil engineering nondestructive surveys of structures and buildings provide valuable information on: concrete thickness and inner rebar structure, presence of wall cavities, presence of water, and reconstruction of wall internal structure.

The principal operation for GPR uses electromagnetic wave propagation and scattering to image, locate and quantitatively identify changes in electrical and to some extent magnetic properties in the object that is under survey. A surface antenna or several antennas produces radar waves that are propagated in distinct pulses in the microwave band of the radio spectrum. These waves are reflected off buried objects and detected back at the source by receiving antenna/antennas. As radar pulses are transmitted through various materials on their way to the buried targets, their velocity changes depending on the physical and chemical properties of the material through which they travel (Conyers, 2004; Conyers and Goodman, 1997). Velocity of energy pulses through the ground is known, and their travel times are measured so can their distance be accurately measured to produce a three-dimensional data set (Conyers and Lucius, 1996). When a radar pulse traverses a material with a different composition or presence of water, the velocity changes and a portion of the radar energy are reflected back to the surface and recorded at the receiving antenna/antennas. Data can be plotted in many different formats, depending what is suitable for the project; it may be plotted as profiles, as plan view maps isolating specific depths, or as three-dimensional models.

Modern GPR also uses high- and low-frequency antennas (dual frequency antennas) that allow for better detection performance; both shallow and deep targets can be detected with a single scan (Simi et al., 2010). Possible errors are reduced and the method saves both time and money. The penetration depth depends on the soil conditions, and these devices can be attached to vehicles, trolleys, or other devices that might be practical in a specific project.

One of the most actual and crucial need in highway engineering is the maintenance and rehabilitation of existing road. There are specialized GPRs for road, railway, and tunnel surveys. Another special application for civil engineering is bridge deck surveying. Deck surveying detects early phase deterioration and current damage to concrete bridge decks. This method can also be used with old structures where drawings are missing in order to locate reinforcement layers and/or prestressing cables, their position and current status. Load capacity of these bridges can, therefore, be calculated.

An example of a novel GPR system for high resolution inspection of walls and structures can be seen in Sarri et al. (2002). This chapter gives an overview, describes the theory, reports measurements and calculations and highlights the field test results. These field test results also confirmed excellent performance of the system.

Use of GPRs in civil engineering structures such as bridge decks, walls, heritage structures, roads, railway beds, and many other applications is innovative and offers real-time information about the real condition of the structure. These are very promising techniques and their use can save a lot of money and time as well as provide safe structures without hidden malfunctions. Two different setups can be seen in Figure 1.7.



FIGURE 1.7 (a) Setups for bridge pavement measurements. (b) Setup for soil measurement with several units connected together (Courtesy of IDS).

1.6.3 Interferometric Radar

Instruments based on radar Interferometry for the measurement of displacements and vibrations were developed in last decades (Farrar et al., 1999; Tarchi et al., 2003) and are now commercially available. For example, landslides, mine slopes, volcanoes, glaciers, dams, bridges, towers, and buildings can be monitored with submillimeter accuracy.

Ground-based radar interferometry is a pioneering revolutionary technology in the domain of geodetic measurements that is now spreading out to other areas such as civil engineering. The measurement device is coherent radar generating, transmitting, and receiving the electromagnetic signals to be processed in order to provide movement and deformation measurements (Gentile, 2010). This new technology allows also dynamic measurements of structural vibrations such as resonance frequencies and vibration modes up to 100 Hz. The method is based on the detection of amplitude and phase of an electromagnetic wave transmitted by the instrument and reflected back from the object to be measured. Each submillimetric displacement gives rise to a phase difference of the reflected wave that is detectable by the radar. More comprehensive introduction to theory of the radar Interferometry can be seen in Rödelsperger et al. (2010).

The instrument is installed on a tripod that allows the movement in the desired area or direction. The complete equipment consists of a radar instrument, a control PC with adequate software in order to store, process, and view the data. The high-resolution capacity over distance provided by the radar produces a displacement map showing the displacement of many points across the entire target. The radar samples that target about every 0.5 m, which correspond as if the target was applied with large amount of sensors. Normally, every structure has reflective points in itself and does not need any additional equipment. Anyhow, there is a possibility to apply passive radar reflectors to the points of interest if needed.

These systems are excellent early warning system. The great benefit with radar interferometry radar compared with current technology is remote operation. No contact whatsoever with the target to be monitored and no complicated or time consuming installation is required. The system is able to supply continuous deformation maps with an extraordinary measurement speed as well as accuracy. Simultaneous displacement of the entire scenario can be remotely measured to distances up to kilometers. Wide area monitoring of landslides provides up to several square kilometer monitoring at once. The measurement is not dependent on weather conditions and high accuracy of 0.01–1 mm can be

reached. Rödelberger et al. (2010) also list advantages and disadvantages of the radar technology compared with various common monitoring techniques as follows

Advantages:

- Remote sensing instrument
- Simultaneous monitoring of all targets within the beam
- Independence of daylight and weather
- High accuracy and spatial resolution

Disadvantages:

- High dependence on atmospheric effects
- Relative displacements in line of sight (LOS) only
- Difficult point localization (georeferencing of target points)

Gentile and Bernardini (2009) also discusses advantages, limitation as well as possible applications of the radar based measurements. He also reminds about following issues that are good to keep in mind when handling with radar techniques to bridges and large structures:

As a consequence of the one-dimensional imaging capabilities of the radar sensor, measurement errors may arise when different points of the structure are place at the same distance to the radar

The radar provides a measurement of the variation of the target position along the sensor's line of sight (i.e., the radial displacement); hence, the evaluation of the actual displacement requires the prior knowledge of the direction of motion.

Manhattan Bridge from 1909 suffered of fatigue cracks and was stiffened under several contracts after 1980. Mayer et al. (2010) describes monitoring of the vertical and torsional displacements of the midspan of the Manhattan Bridge using Interferometric Radar and Global Positioning Systems (GPS). Figure 1.8 shows the Interferometric Radar equipment



FIGURE 1.8 Interferometric radar equipment in front of the Manhattan Bridge (Courtesy of IDS).

in front of the Manhattan Bridge. The bridge testing provided valuable information about the dynamic characteristics of the structure, effects of the stiffening and also verified that the deflections and resonant frequencies measured with these different technologies compared well with one another.

1.7 GLOBAL POSITIONING SYSTEM

The global positioning system (GPS) systems are space-based global navigation satellite systems that provide reliable location and time information anywhere on or near the Earth. A comprehensive introduction to global positioning system (GPS) can be seen in El-Rabbany (2002). According to Rizos (2002), relative positioning techniques can be divided into following classes: (1) static and kinematic surveying techniques, (2) differential GPS (DGPS), and (3) real-time kinematics (RTK) techniques. Plentiful applications can be found in land, marine, and air navigation. Nowadays, GPS has also found its way to civil engineering applications. It is weather independent but needs an unobstructed line of sight to five GPS satellites in order to perform SHM of large structures; long-span bridges, high-rise buildings, high industrial chimneys and towers are some of the current targets.

The major distinction between static and kinematic baseline measurements involves the method by which the carrier wave integer cycle ambiguities are resolved. A network or multiple baseline approach for positioning is normally used in static applications. Static GPS system can also have multiple receivers, multiple baselines, multiple observational redundancies, as well as multiple sessions. Adjustment of observations is performed by least squares; therefore, the system is very accurate. The disadvantage is that the observation time is also longer: around an hour to several hours.

Real-time kinematic (RTK) positioning requires two receivers that record observations simultaneously. Postprocessing of the data to obtain a position solution is not required and real-time surveying in the field can be performed. Advantage is that the measured data can be checked directly. Modern GPS systems use real-time kinematic surveying; direct and absolute measurements in real-time are achieved.

Improvement to GPS is called differential global positioning system and it uses a network of fixed, ground-based reference stations to broadcast the difference between the positions indicated by the satellite systems and the known fixed positions. DGPS needs to receive a correction signal using a separate radio receiver in order to perform. A “base station” antenna is placed near to the target on a stable reference place with line of sight to at least five orbiting GPS satellites so no signals are reflected. Antennas that are placed on the structure in monitoring points are called rover antennas. The base station receiver calculates its position based on satellite signals and compares this location to the known location. The difference is applied to the GPS data recorded by the roving GPS receiver. When the base station calculates and broadcasts corrections for each satellite as it receives the data, it becomes a real-time application.

Accuracy of the order of 1 mm can be obtained using GPS, although not necessarily with real-time systems. The alternative to the real-time mode is the processing of data saved by separate receivers.

Brownjohn et al. (2004) reports about testing the feasibility of GPS for building performance monitoring for Republic Plaza high-rise building in Singapore with a dual-rover RTK GPS system in order to provide ultralow frequency response data. This chapter

describes the system integration for the present monitoring system and presents some results on performance of the building and the GPS system.

Roberts et al. (2010) do report about monitoring bridges by Global Navigation Satellite Systems (GNSS). The work reported investigates the use of kinematic GPS and compares results to the FE modeling. Seven different bridges, of various sizes and types are investigated, and the results show that GNSS is valuable tool to monitor the movements, and the characteristics of the movements of bridges.

Example of monitoring earthquakes by GPS monitoring systems is the instrumentation of the medium span cable-stayed Naini Bridge in India. GPS rovers have been installed at the pylon tops (see Figure 1.9) of the bridge in order to monitor for settlements of the pylons as these have been constructed by use of an unusual foundation method based on thinking wells. By using the advanced GPS monitoring program DEMON by Eiva, Denmark, it has been possible to carry out very accurate RTK GPS measurements with accuracy down to 2 mm lateral and 5 mm vertical by postprocessing data based on special algorithms. These include computation of absolute positions from pseudorange (code) observations, preliminary validation of single difference baseline observations, the computation of relative positions using code and carrier or carrier only observations and the computation of the GPS satellite orbit repeatability period, which is used to reduce or eliminate effects due to multipath in the baseline results (Bhanushali et al., 2006). The Kashmir Earthquake October 8, 2005 was recorded with this system, even though the earthquake occurred 800 km from the bridge. Figure 1.10 shows output from the bridge SHMS with lateral pylon vibrations with amplitudes up to 50 mm.

Modern GPS provides high performance and real-time measurements, they are easy to install and have shown to be suitable for measurement and SHM purposes for various civil engineering structures.



FIGURE 1.9 GPS at pylon top of Naini Bridge, India (Courtesy of COWI).

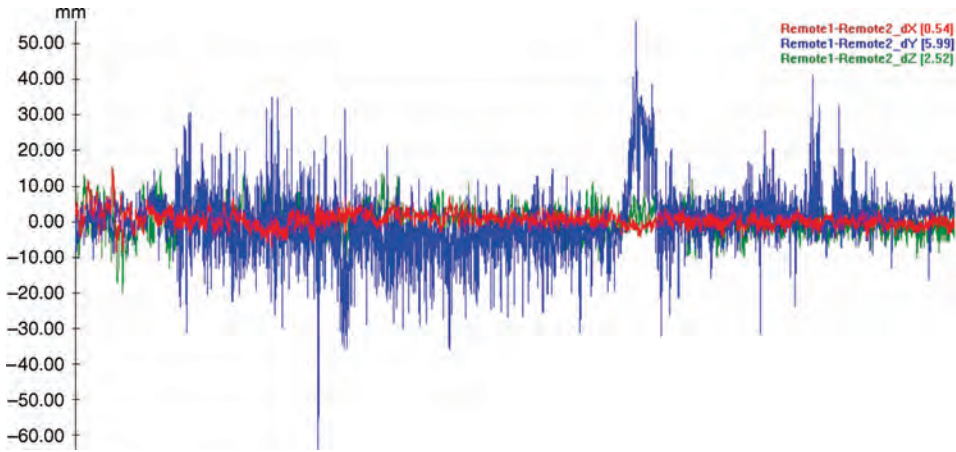


FIGURE 1.10 The Kashmir earthquake that occurred 800 km away was recorded on the Naini Bridge by the COWI operated SHMS; Figure 1 shows output from the bridge SHMS with lateral pylon vibrations with amplitudes up to 50 mm.

1.8 CORROSION MONITORING SYSTEMS

Corrosion is a huge problem and causes damage, failure, and uncertainty in structures. The loss of capital caused by premature deterioration in structures is large. Undetected corrosion may also cause collapse of the structure and thereby lost of human lives. Corrosion is present in aggressive environments and structures in these environments should be provided with adequate SHMSs for corrosion monitoring. “Techniques for corrosion monitoring” (Yang, 2008) gives a review of corrosion fundamentals and provides a four-part comprehensive analysis of a wide range of methods for corrosion monitoring, including practical applications and case studies. The book reviews electrochemical techniques for corrosion monitoring, such as polarization techniques, potentiometric methods, electrochemical noise and harmonic analyses, and galvanic sensors, differential flow through cells and multielectrode systems. In addition, the physical or chemical methods of corrosion monitoring such as gravimetric, radioactive tracer, hydrogen permeation, electrical resistance, and rotating cage techniques are then presented.

As part of a durability strategy for reinforced concrete (RC) structures in aggressive environments (marine or polluted) corrosion monitoring sensors can be installed. Sensors can be embedded in concrete from the first beginning so that the moisture content and temperature of the structure can be monitored during the curing process to ensure maximum strength of the concrete. Once construction is complete, the instrumentation can be used to conduct long-term monitoring of corrosion conditions over time (Dunn et al., 2010). Another issue with reinforced concrete infrastructure in order to prevent deterioration is cathodic protection. It controls the corrosion of a metal surface by making it the cathode of an electrochemical cell. It is important to know the timing for applying the cathodic protection and there are various measuring devices for that purpose. Many electrochemical measurements only record the corrosiveness of the environment at the time of taking the reading. In order to evaluate the effectiveness of the cathodic protection, it is common to use devices that measure cumulative metal losses. The whole corrosion profile

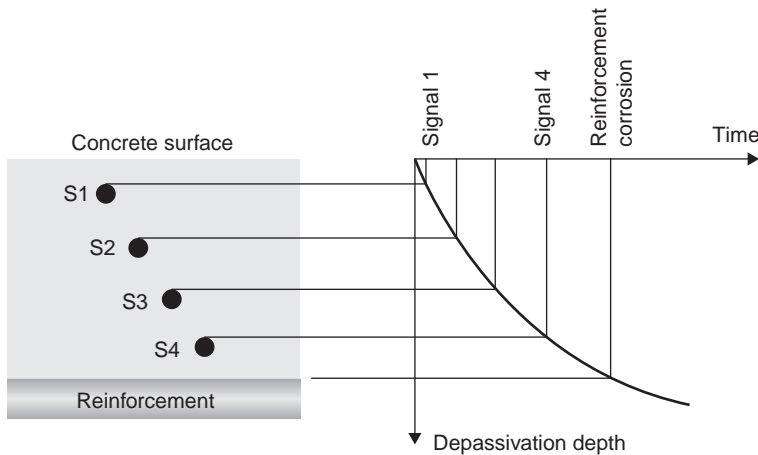


FIGURE 1.11 Electrodes placed at successive depths can sequentially be measured to locate the depth to which the chloride had ingressed into the reinforcement and thereby the corrosion front can be established (Buhr et al., 2010) (Courtesy of COWI).

in time is recorded by the probe and more accurate decisions in maintenance planning can be taken.

Traditional corrosion monitoring sensors for RC are typically a set of carbon steel electrodes A1–A4, which are small pieces of reinforcement with a specified concrete cover placed within the structural steel cover. The carbon steel electrodes are intended to be used as anodes. They are galvanically discontinuous, that is, they are not in electrical contact with the structural reinforcement.

These embeddable probes monitor ingress of aggressive substances, particularly chlorides. Figure 1.11 illustrates the depth to which the chloride had ingressed into the reinforcement and thereby the corrosion front can be established.

An interesting study of corrosion monitoring in concrete structures of Copenhagen Metro can be seen in Buhr et al. (2010). In all 170 sets of corrosion probes were installed to have a prewarning system for corrosion in the concrete cover. The article describes the monitoring system and highlights the results from the pilot installation. Another study (Dunn et al., 2010) presents a reinforced concrete corrosivity monitor including linear polarization resistance, open circuit potential resistivity, temperature, and a potential related to chloride ion concentration.

There is a wide variety of fiber-optic sensors for corrosion and humidity measuring purposes. Many of these sensors are aimed at reinforced concrete measurement and a lot of research in ongoing (Srinivasan et al., 2009).

Corrosion problem also occurs in steel constructions and the most common methods to protect the steel are coating, painting, and weathering steel. The coating/paint must be robust and properly applied in order to provide protection in the existing environment. Monitoring corrosion of the external surface of steel structures using potential or voltage techniques is also common. Weathering steel is steel with special chemical constellation that forms a protective layer on its surface under the influence of the weather. The surface of the steel corrodes by forming a dense and tightly adherent oxide barrier that seals out the atmosphere and retards further corrosion

(Roberge, 2006). Weathering steel is commonly used in sculptures, marine transportation, chimneys, and buildings.

Monitoring corrosion of some delicate structures such as steel oil pipelines to determine the state of corrosion is required by law in the United States. Many bridge structures also do suffer of combined corrosion and fatigue effects and do require monitoring for safety and maintenance reasons. The Silver Bridge over the Ohio River failed from the corrosion cracks in an eyebar killed 46 people in 1967 (Kulicki et al., 1990), and Oakland Bay Bridge in San Francisco also had problem with cracks that may have been caused by vibrations from wind and rumbling traffic united with corrosion.

Acoustic Emission methods can also be used for corrosion monitoring. The emission from corrosion usually releases much less energy than emission from crack growth and the sensor frequencies used to detect these signals are determined more on the basis of the environmental noise under test conditions than on the frequency spectrum of the emission (Cole and Watson, 2005).

Newer structures use optimized steels and coating materials that can provide much longer service lives to improve the performance of the structures and hopefully more reliable structures can be offered in the future while capital is also saved.

1.9 WEIGH-IN-MOTION (WIM) SYSTEMS

1.9.1 Weigh-in-Motion

Weigh-in-motion is a technology for measuring the weight of moving vehicles, and it is commonly used for trucking terminals, seaports, toll roads and bridges, and border crossing. Many heavy vehicles are weighty and do also exceed legal limits and can cause severe damage to roads and bridges. Accurate information about vehicle axle loads is important in order to make prognoses for traffic development as well as in construction and maintenance planning (Quilligan, 2003). Seaports and trucking terminal do acquire exact weight of cargos for loading purposes.

Various WIM systems have been developed around various technologies. Emerging WIM systems are being developed around new technologies such as fiber optics.

The most common weigh-in-motion systems include

1. Load cell
2. Bending plate
3. Piezoelectric
4. Fiber optic

Load Cell WIM systems use a single load cell with two scales to detect an axle and a weigh both the right and the left side of the axle at the same time. The system records the weights measured by each scale and summarizes them in order to count the axle weight while the vehicle passes over the load cell. Bending Plate WIM systems use with strain gauges bonded to the underside of the plates. The system records the strain measured by the strain gauge and calculates the dynamic load while a vehicle passes over. Piezoelectric WIM systems use piezoelectric sensors to detect the voltage change caused by the pressure exerted on the sensor by an axle while passing over. The system records the electrical charge created by the sensor and calculates the dynamic load.

Several WIM solutions with fiber-optic sensors have been presented in recent years. The advantage of these sensors is that they can be embedded deeper in the road surface that provides highly protective environment. These sensors are also simple to install and maintain and their lifetime is equal or better compared with conventional sensors (Kunzler et al., 2003). Also, the cost of fiber-optic sensors is competitive nowadays and as the new technology will improve, it might eventually replace the ones currently in use.

Bending-plate and piezoelectric WIM systems work to reasonable accuracies for traffic moving at motorway speeds. These technologies provide sufficiently accurate data for the statistical review of traffic and the development of structure specific assessment live loading. These sensors will detect vehicles without disrupting traffic flow. These sensors are ideal for installation on main thorough-fares such as motorways or dual-carriageways.

High-precision weight measurement for law enforcement is achieved well with weigh-bridges based on load cells. However, weigh-bridges tend to require slow transition by the vehicle being weighed in order to achieve high accuracy. Clearly, the installation of such a device for the measurement of all vehicles would require all vehicles to slow down in order to be weighed. This is only feasible at gated toll stations and would also require a large number of devices to cover the entire toll lanes provided.

For structures as very large bridges, it is important to track the total traffic load on the bridge. An accurate assessment of total traffic load is best achieved by the measurement of weight of all vehicles. A WIM system that can operate without interfering traffic flow is therefore also appropriate. With the weighing of all vehicles, accuracy on measurement of $\pm 10\text{--}15\%$ would be sufficient for determining the total traffic load on a bridge, and would also be sufficient for the purposes of poststatistical review for the development of bridge specific assessment live loading.

Typical WIM systems will be designed in accordance with the design requirements as presented in ASTM Standards (ASTM E1318-09). Other considerations to the choice of WIM system include maintenance requirements and system lifetime.

1.9.2 Railway Weigh-in-Motion

As for the WIM systems, the railway weigh-in-motion (R-WIM) market is under development. Old systems that are available frequently require the slow passage of trains by the measurement of loads transferred from the rails to the sleepers, or from the sleepers to support structures. Emerging in-line wheel load weighing systems are now available and can detect overloaded wagons as well as damaged wheels and flat spots. These systems are now used by railroad companies to check the operational safety and to identify possible unsafe wagons. It is not suitable for trains to be slowed down for weighing prior to crossing any structures to be monitored. Therefore, a system that has minimal impact on the railways shall be selected. Kistler Rail WIM based on quartz sensor technology is currently providing such a system (see Figure 1.12).

1.9.3 Bridge Weigh-in-Motion

Bridge weigh-in-Motion (B-WIM) was first developed in the late 1970s (Moses, 1979). A lot of development took place in the latest decade (Quilligan, 2003; Leming and Stalford, 2003), and there is also a commercial system on the market. B-WIM is the process by which axle and gross vehicle weights of trucks traveling at highway speeds can be



FIGURE 1.12 KISTLER—Rail WiM (using quartz sensor technology) (Courtesy of Kistler).

determined from instrumented bridges. B-WIM systems involve attaching strain sensors to the soffit of a bridge and placing sensors for detecting axles in order to provide information on vehicle velocity, axle spacing's and position of each vehicle. This is done either with axle detectors on the road surface axle detectors or by placing additional strain sensors under the bridge. Obrien et al. (2008) do give a comprehensive overview about latest development as well as applications in Europe and Japan.

1.10 COMPONENTS OF STRUCTURAL HEALTH MONITORING SYSTEM

It is important to have an appropriate comprehension of the structure and elements of a SHMS. In order to be able to design sustainable SHMS solutions as well as to integrate the advantages offered by a SHMS into structural design and operation, a short introduction to SHMS components is given.

Following common outline highlights the elements integrated in the SHMS solution

- Sensory system located on the structure; It consists of various types of sensors depending on the nature of the structure; Signal collection, synchronization, conditioning, and digitization units may also be included;
- Data acquisition systems to execute preprocessing and local buffering for sensors distributed in a limited geographical area on the structure; Portable and/or fixed systems are possible;
- Data communication system for the transfer of the collected data to a remote computer;
- Data processing and control system with database application that collects, stores, and processes the raw data in real time;
- User interface;
- Maintenance tools;
- Interfaces to external systems.

The task of the SHMS is to monitor the in situ behavior of a structure accurately and efficiently, to assess its performance under various service loads and environmental conditions, to detect damage or deterioration, and to help determining the health or condition of the structure. The SHMS system should be able to provide reliable information regarding safety and integrity of a structure. This information can be incorporated into maintenance and management strategies of the structure, as well as in improved design guidelines.

1.10.1 Sensory System

A sensor is a type of transducer that converts a physical property into a corresponding electrical or optical signal. As the number of various sensing technologies is large, it is important to have a sound general knowledge about them, in order to select the most suitable for a specific structure. It is favorable to choose commercially available sensors that are well proven in similar circumstances as the intended structure.

On the other hand, emerging technologies do lack that experience in many cases. The recommendation is to establish a thorough testing schedule and procedure that will proof the feasibility, functionality, installation issues as well as the long-term function of the complete sensory system, hardware, software, and other data acquisition equipment for the completed installation. Qualified, experienced, and competent testing team with representatives from stakeholders is to be established. External experts with both long-term theoretical knowledge and practical experience need to be part of the testing team. Testing starts with a focused technical specification: measurement parameters and prospective measuring techniques are specified. The results shall be verified against the technical specification and the testing aims. It is complimentary to start in small scale if anything unexpected would happen so no unnecessary capital is wasted.

The sensory system will include the sensors and their corresponding interfacing units for input signals gathered from various monitoring equipments and sensors. Detailed guidelines about SHM issues are given in Aktan et al. (2001). This report is directed to bridges but gives also general appropriate knowledge about sensor performance characteristics and helps the reader with sensor selection criteria.

1.10.2 Data Acquisition System

Data acquisition system converts the physical phenomenon measured by sensors into digital numeric values that can be operated by a computer. Modern monitoring systems have a tendency to grow rather large and over-instrumentation is common (Brownjohn, 2007). All data acquisition modules need also to be synchronized and preferable a fiber-optic network is used for communication. A local area network (LAN) is commonly used in SHMSs. The network connects computers and devices and ensures a high level of redundancy. Communication protocol is also specified.

Data acquisition system consists of hardware, software, and secondary equipment such as connectors, cables, and cabinets. Many SHMS have need for unique solutions and the data acquisition system can be adapted for these special needs, that is, so-called tailored solution can be offered.

1.10.3 Data Processing and Control System

Data Processing encompasses automatic conversion of the raw data into useable information. It consists of data collection, transmission, preprocessing, analysis, postprocessing, and storage of the data in a format that is easy to access and present.

Larger SHMSs do have a control room where supervisory control and data acquisition (SCADA) are normally located. The master screen in a control room can be either a desktop display with the ability to transfer data between different functions quickly and easily, or a large display wall. This wall can be divided into various monitoring areas with different functions.

Various operators will work with the SHMS either permanently or at chosen intervals. The man-machine interface may be established for the SCADA and will be used for display of general interest's views from the SHMS monitored systems. The system will visualize in real time all collected information, in the most suitable way for an immediate and efficient representation (graphs, tables, videos), and will allow the search, visualization, and elaboration related to user specified periods. Example of the control room design with SCADA can be seen in Figure 1.13.

1.10.4 User Interface

The conventional SHMS solutions have been forced to use local, workstation based user interfaces due to various limitations. The modern solutions though, do use a universal and standard based web interface. Generally, the goal is to create a user interface that makes it

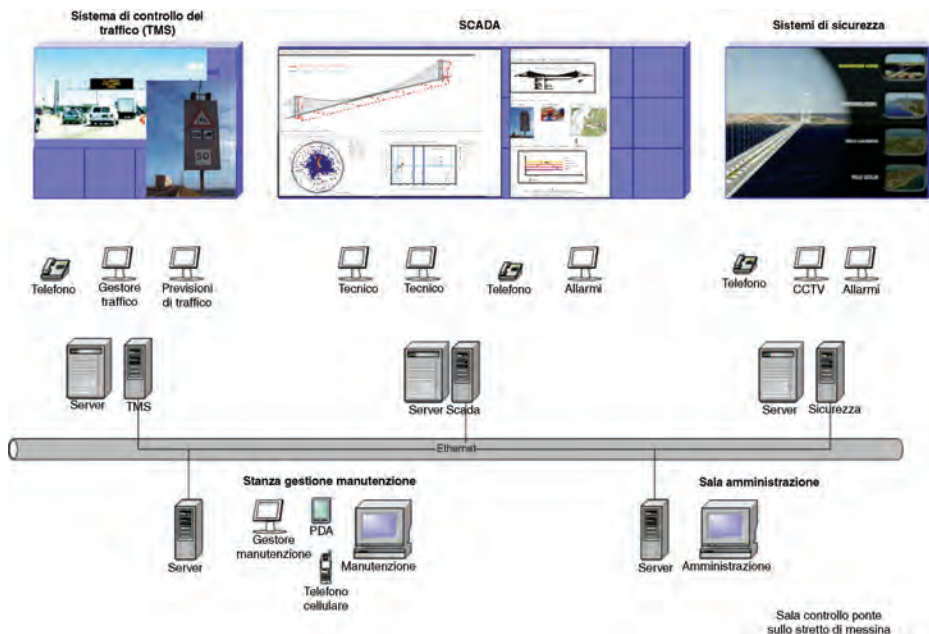


FIGURE 1.13 Example of control room design with Traffic Control, SCADA and Security Control (Courtesy Stretto di Messina).

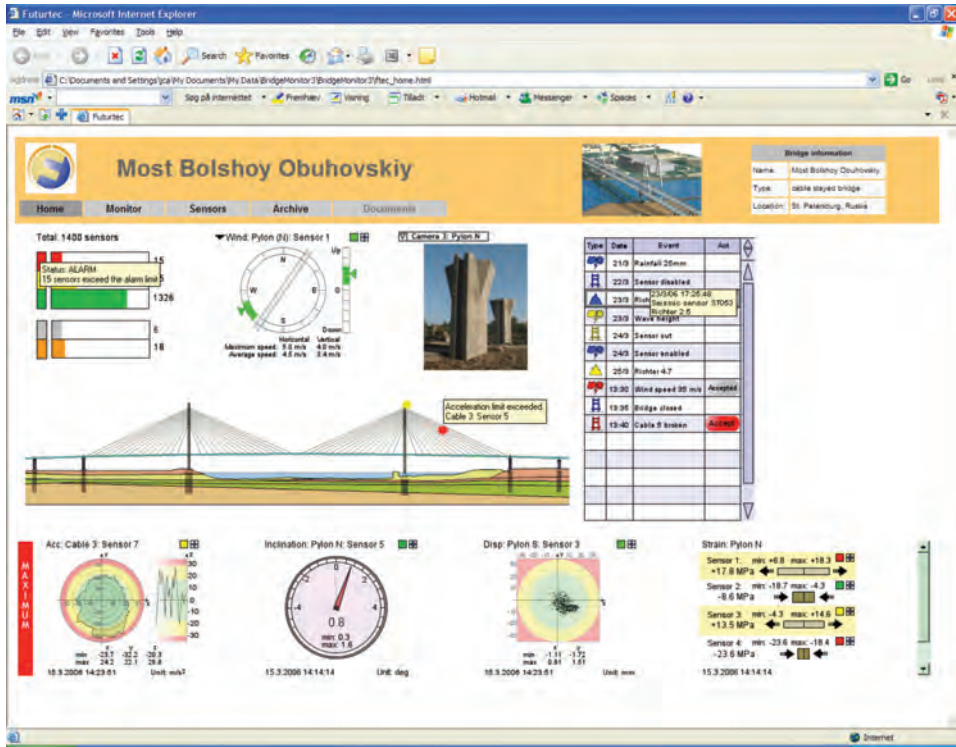


FIGURE 1.14 Optimized Internet user interface at the Neva Bridge (Courtesy of Futuretec OY, Finland 2006).

easy, efficient, and pleasant to operate a system while producing the desired results. The user interface has to be optimized to the different phases of the lifetime of the structure; construction, testing, operation, and extreme (unplanned) events. The measurement system should be redundant and indeed error resistant.

The role of the user interface might change over the years. The user interface should be as simple as possible as even small user interface problems when repeated day after day get to be intolerable in long term. The interface must be planned to give the minimized output and modifications to the monitoring system or to the user interface should be few and cautiously considered to minimize the need for personnel retraining. The management system of the structure must work so that it enables operators to understand and trust the information provided by the system and rapidly act on it regardless of tools, time and space. See Figure 1.14 for an example of the optimized Internet user interface.

1.10.5 Maintenance Tools

Maintenance tools are required for the inspection and maintenance purposes of the whole SHMS. A Portable inspection and maintenance system is located on the structure and can be used if problems should occur. Many emerging technologies do need some special tooling and furthermore training in problem finding and solving.

1.11 STRUCTURAL HEALTH MONITORING SYSTEM DESIGN

An appropriate SHMS for a structure should be able to provide, on demand, reliable information pertaining to the safety and integrity of a structure. Critical parameters, load conditions, and environmental circumstances are monitored, preferably with the minimum amount of sensors. The more straightforward the system is the simpler it is to perform data processing and reliable information about the real structural conditions of the structure in real time can be extracted and presented.

1.11.1 Structural Analysis for New Structure

The structural analysis is a preliminary task and the core of a favorable monitoring. Identification of the structure and the parameters that are to be monitored is done cautiously in order to choose the most important features to be monitored. Feature identification concludes a review of drawings and any relevant information about design uncertainties, special care is given to new complicated design as well as new material constellations.

A complex structure often requires a complex SHMS. Measurements need to be comparable with the calculations, analytical models and FEM in order to allow for calibration. Technical requirements for desired sensors are established.

1.11.2 Structural Analysis for Existing Structure

Existing structures do provide a lot more information; drawings, inspection protocols, maintenance actions, retrofitting/strengthening of the structure, noted problems, test sample results, measuring data, concerns, or verified structural weakness. It is also possible to visit the structure and perform a visual inspection. The structure can also be tested and classified in order to find the most relevant features for monitoring.

Technical requirements for desired sensors are established.

1.11.3 Sensor Selection

Sensors that will fulfill the technical requirements established in the structural identification are chosen with care. The chosen sensors need to be commercially available and proven technology and installation procedure for all chosen sensors are studied in detail. The choice of sensors may involve collaboration with various experts in several fields of engineering depending on project complexity. If dealing with emerging technologies where experience lacks, and some uncertainties do exist, a small structural test or test installation can be performed in order to clear these uncertainties.

Locations of the sensors, cables, connection boxes, data loggers, central units, and communication systems are decided and drawings are established. If temporary monitoring devices under the construction are needed, they are also listed in detail and the list of equipment and the schedule of temporary monitoring is established.

1.11.4 Data Acquisition Issues

Data acquisition components that are compatible with selected sensors and which will fulfill the technical requirements established in the structural analysis are chosen. Measurements schedule for every sensor is established. Output signal of a

sensor is either analogue or digital. The conversion of an analogue to a digital signal is performed by an analogue-to-digital converter (ADC). Many electrical sensors need signal conditioning and filtering in order to allow for transmission and postprocessing of the signal. In general, methods for data quality assurance, analysis, and processing are invented for all types of sensors. Civil engineering SHMSs tend to be large in size and signals need to be transmitted over long distances, many times up to few kilometers. Therefore, the data transmission needs to be well planned, to prevent data corruption or any loss of data.

Many projects are unique with unique needs: if the commercial software for the sensors is not adequate, it has to be calibrated and tested. System servicing and troubleshooting procedures after the installation are also described and planned in detail. The criterion for the database is established.

If the safety purposes are included in the designed monitoring system, the warning and alarm subsystem is designed and created. Criteria for decision making with alarm or warning, reporting schedule, and the form of documentation are established.

1.11.4.1 Data Analysis and Data Mining Comprehensive data analysis is a key issue to successful application of a SHM. Data mining is the process of extracting patterns from large data sets by combining methods from statistics (statistical pattern recognition) and artificial intelligence. Data analysis may allow for damage location and quantification as well as for condition assessment. Comparison between the measured values and FEM is also part of the scope and allows for model calibration and updating, and maintenance planning. Different approaches such as frequency and time domain analysis for evaluation and interpretation of measurement data can be found in FIB Guidelines (<http://www.ishmii.org/wp-content/uploads/2010/07/FIB-Task-Group-Europe.pdf>).⁸

The SHMS will collect synchronized measurement data, status of sensors and fault notice generated by the individual subsystems and equipment. The on-site operator shall have the possibility to request displays of all current measured points, values or status, initiate logs and reports on events, notice and confirm alert, and alarm description. The SHMS shall also be designed to be as user friendly as possible: information that is available need to be straightforward as the operators may lack the ability to understand and judge procedures behind the analysis. Emerging technologies do set new requirements for operators and education is needed in order to avoid mistakes and malfunctions. Requirements for training and education should be presented and scheduled.

1.11.4.2 Data Access Accessibility and easy localization of measured point, time and data is an important challenge in large projects, because the size of the collected raw data can be up to terabytes. The access requirements differ for various stakeholders. The designer desires to find data that can be compared to his original design calculations. The construction engineer desires values that describe what is happening in the structure at the moment of construction and testing. The operator's desires values that describe how the structure is operating under load during the operation and the unit responsible for maintenance planning desires values that describe how the structural behavior changes in the long term.

In order to design an efficient platform for all stakeholders, all specific needs are collected and evaluated. Technical requirements are established and procedures for different groups are defined. The key point about access to the data archive is that all access

⁸Accessed January 27, 2010.

methods have to use open protocols and to have a documented key feature with “ease of access” in mind, then just documenting the data access protocols is not sufficient.

1.11.4.3 Data Storage and Data Archival If possible, all raw data are stored and later archived as it is uncertain what types of information will be needed or useful in the future, during the expected lifetime that could be up to 200 years. Data storage consideration will reflect the model for data organization. Depending on the extent and purpose of data storage, the monitoring system can be structured to fulfill the specified requirements as follows:

- None
- Ad hoc reports at random/selected periods
- Statistic
- All raw data
- Processed data

For small monitoring systems designed to support the user safety of the bridge, one can choose only to save alarms and warning events when these are exceeding preset trigger values; consequently, storage of any other measured data will not take place. Limited data saving can also be the strategy for those monitoring systems based on ad hoc, random or selected period reports, thereby reflecting the periods when the monitoring system is in operation.

Ad hoc measurement based on portable SHMS is often subject to discard many measured data and archive only those that are needed for the analysis work, as the measurement points of these systems are often mobile and the measurement from the sensors can differ from time to time.

Data archival is essential in large projects and inactive or noncritical data will be transferred to a particular type of long-term storage media. Archived data generally consist of primary copies of the data being stored.

1.11.5 Responsibilities and Installation Planning

Responsibilities for the incorporated partners are well defined in order to avoid malfunction, inaccuracy, and delays. All deliveries, installation, and testing are planned in detail, and timetables and schedules are established. The opportunity to highlight requirements, standpoints and expectations for the system is given to all incorporated partners so that changes can be made upon request.

After the final decision is taken about the SHMS components, it is necessary to make a detailed installation plan. It is recommended to pay a visit to the existing structure on-site in order to make a visual inspection and recognition. Attention is paid for accessibility to the structure from the installation point of view. The condition of surfaces, need for fixing the concrete surface, grinding the steel surface free from corrosion and paint, need for scaffolding for installation to be able to reach the installation points and be able to fasten the sensors in a proper and safe way to the structure, possible position for equipment and cables, access to electricity and any other concern is carefully investigated onsite. For structures under construction, the timetable of the construction steps is studied carefully, and a timetable for the installation of the sensors is established. A visit is paid to the construction site if possible, and all insecure installations are considered once more and

tested if needed. If any divergence from the planning is noted, all the partners are to be contacted in order to find a new solution. In complicated concrete constructions, the timetable for reinforcement need to be studied carefully for easy and rapid installation if the sensors are to be fixed on the reinforcement.

The complete installation plan is distributed to the owner, contractor, installation team, and any other part that is involved in the process so that they are able to request changes. If required, the installation plan is revised until all parts have approved it (Enckell, 2006).

1.12 SYSTEM PROCUREMENT AND INSTALLATION

1.12.1 System Procurement

Construction lead times for minor and portable systems are typically calculated in weeks or months and most of their components are available immediately. Mid to large size system engineering and construction can take from a few months to several years including a number of complete subdeliveries from the construction period to the operation phase.

Following tasks are recommended

- Application development has to be carried out under a verified quality system.
- All hardware components need to undergo component level testing prior to integration.
- All application modules have to undergo component level testing prior to application testing and resulting into the actual SHMS hardware.
- Integration and testing of all required specifications are performed prior to the final factory acceptance test that is a requirement for shipping authorization. A client representative is typically present at FAT that generally takes a few days to complete.

1.12.2 Commissioning

Commissioning assures that all subsystems and components of a SHMS are designed, installed, tested, operated, and maintained in accordance to the technical requirements of the owner. Various techniques and procedures to check, inspect, and test all components from individual sensors to data analysis subsystems are used in order to allow quality assurance.

The SHMS manufacturer is also responsible for all other supplies such as installations performed by local subcontractors. Sensor and processor installations, component connections, and testing are always the responsibility of the SHMS supplier's specialist personnel.

Significant part of the commissioning is the operator and maintenance training. Written user instructions such as Operation and Maintenance manuals in local language as well as clear and fast on-line search and help functions are required. On the large and long-term projects lasting for years or even decades, it is challenging to ensure integrity of the system due to additions and changes to the initial scope during the project. The possibility for modification of the existing technology and integration of the new technology is a requirement. The SHMS supplier should be capable to offer remote service for software maintenance, updates and debugging and these requirements have to be part of the initial operating system and application design.

1.12.2.1 Factory Acceptance Tests (FAT) A factory acceptance test demonstrates that the sensors and related hardware and software operate in accordance to technical specifications. FAT is documented and does test functionality, communications, support systems, error management, and interface requirements.

The client is generally present and shall approve FAT. Any unapproved equipment shall be rejected or subject to repair. FAT is especially important for emerging technologies due to the uncertainty of the potential issues that may exist.

1.12.2.2 Site Acceptance Test (SAT) The site acceptance test is an important verification of the SHMS and important test for manufacturer. This test shall be performed after the installation and will guarantee that the system works in accordance to the design parameters and technical requirements. It is advisable that the SAT is done in three phases:

1. Functional test of the whole system (no loading on structure).
2. Calibration of the system parameters during the load tests of the structure.
3. Verification of the calibrated parameters during the operation.

Unapproved sensors shall be rejected or subject to repair. If installed sensors are rejected, alternative measurement techniques shall be established.

1.12.3 Installation

Installation procedures that are described in the literature seem often to point out that the installation is an easy process. Anyhow, in the practical case it is often the opposite. This misleading information is causing a lot of expectations and may cause malfunction and might decrease the quality of the system. Installation is a complex matter and related issues are highlighted in this section.

A good collaboration with the contractor is vital as it is important to get information about delays and changes in construction or schedule rapidly. The workers on the construction site must be informed about the sensor installation so that a positive attitude is created in order not to damage the sensors. Need for the necessary equipment and personal is controlled.

Installation of the sensors and devices is performed following the installation plan and with the presence of experienced personal.

The works to be done before actual installation are the following (Enckell, 2006):

- The sensors and devices to be installed are delivered, checked, and quality controlled.
- The personnel installing the sensors are well informed about the installation procedure and trained so it has a good understanding of the sensor devices. When installing novel sensors, it is necessary to give personnel additional education about the sensors and installation procedure.
- The necessary equipment is procured and if necessary tested.
- Data acquisition systems are calibrated and the software is programmed and tested in the office.
- A lot of practical details often delay the project; therefore, it is good to check the following: access to the construction site, safety regulations, other activities that

might collide with the installation at the building site, access to electricity and facilities for the personnel, and so on.

- Passages for the cables, and so on, are realized in advance if possible. This is especially important in new concrete structures where the plastic pipes or similar facilities need to be cast in advance.

The works to be done in actual installation are the following ones (Enckell, 2006):

- The sensors and devices are installed in accordance to the installation plan and drawings.
- If any changes are made, they must be noted so that the drawings can be revised and updated.
- The installation, as performed, is described in detail in a diary and documented with photographs.
- The sensors are tested, measured, and calibrated if possible or necessary.
- If there is a risk for damages, the sensors are protected or marked.
- The sensors are connected to the data logger for temporary or permanent measurement, or if they are not to be used directly, they are protected or set in a safe place.
- If temporary measurements are performed during some stage of the construction the measurement equipment is protected against damage and marked clearly.
- In the case of a new concrete construction and sensors are to be embedded, it is recommended that the responsible engineer is present during the casting in order to supervise the casting and ensure the survival of the sensors.

The works to be done after installation of the sensors and cables are the following ones (Enckell, 2006):

- The sensors and devices are connected to connection boxes, data acquisition systems main units, and so on, according to installation plan and drawings.
- The communication system is established.
- The cables are fixed temporarily or permanently on the cable rack.
- If any changes are made, they must be noted so that the drawings can be revised and updated.
- The installation as realized is described in detail in a diary and documented with photographs.
- The system is tested.
- The system and monitoring results are verified by other systems, models, or calculations.

Correct installation procedure in the field is highly important for trustworthy results and for the long-term quality of the monitoring system, especially with the installation of the new and emerging technologies. Many projects have unique requirements; therefore, it is not possible to describe a general procedure that covers all details for all the projects. Good and detailed planning saves time and thereby the costs for the installation and ensures qualified installation.

1.12.4 Lifetime Support

The SHMS component availability is typically guaranteed for 10–30 years. Operating system, user interface and application support services in the future could be highly variable, depending on the supplier and can pose a serious risk to the stakeholder if an improper choice is made.

As the lifetime expectation for the solutions in question is over a decade and potentially it should last for several decades, the safest choice is to require the system to be based on an open source operating system such as Linux, user interface to be built into a standard web browser and internal communications utilizing extensible markup language (XML). Ensuring support for the application itself can only be achieved by requiring a third party software escrow arrangement as even the largest vendors cannot provide guarantees for the required competence availability after several years. SHMS should be as flexible as possible in order to guarantee the long-term function and upgrades.

1.12.5 System Efficiency and Redundancy

The system has to provide full and detailed information storage under disaster conditions to enable fast recovery and postdisaster off-line analyses. In fact, the most critical usage pattern for the monitoring system is during disaster recovery. This dictates the requirements for user interface simplicity and intuitiveness as well as a remote data storage service at the supplier's server.

1.12.6 Dismantling Environmental Issues

A plan for dismantling and waste collection needs also has to be established. The materials used should be environment friendly and chemicals used in the installation procedures as well. Safety instructions need to be clearly stated if any risks are present.

1.13 APPLICATION OF STRUCTURAL HEALTH MONITORING SYSTEMS

Authors of this chapter have been working with emerging technologies: both in theory and in practice. Several both small-scale and large-scale installations and measuring campaigns are completed and SHMSs are designed, mostly for international major bridges but also for other structures such as high-rise buildings and heritage structures.

In the 1970s, the first structural monitoring was mainly based on simple instrumental loops for automatic data acquisition from some critical points of the structure. This trend has slowly been developed from relatively small systems logging only semi-static parameters to today's modern turnkey SHMS distributed data acquisition systems with a high degree of build-in structural evaluation and decisional support.

Some SHM projects are presented here. These projects give an image of SHM projects with advantages, disadvantages, malfunctions, and uncertainties through the past decades.

1.13.1 High-Rise Building, Singapore—2001

Singapore is a cosmopolitan city-state often described as a gateway to Asia with a city landscape of tall buildings. The Housing and Development Board (HDB), as Singapore's public housing authority, has an impressive record of providing a high standard of public

housing for Singaporeans through a comprehensive building program. As part of quality assurance of new HDB tall buildings, it was decided to perform long-term structural monitoring of a new building of a project at Punggol East Contract 26 (Glisic et al., 2003).

1.13.1.1 Quick Facts

Name and Location: Punggol East Contract 26, Block 166A, Singapore

Owner: Multiple

Structure Category: High-rise building

Start of SHMS: 2001 (during construction)

Number of Sensors Installed: 10 discrete long-gauge strain sensors

Instrumentation Design by: SMARTEC and HDB

1.13.1.2 Description of the Structure The Punggol EC26 project consists of six blocks with pile foundations; each block is a 19-storey building, with six units supported on more than 50 columns at ground level. The block called 166A was selected for monitoring. The photographs taken during construction and upon completion are shown in Figure 1.15.

This monitoring project is considered as a pilot project with two aims: to develop a monitoring strategy for column-supported structures such as buildings, and to collect data related to the behavior of this particular building providing rich information concerning their behavior and health conditions. The monitoring is to be performed during whole life span of the building, from construction to the in use. Thus, for the first time; the sensors are used in a large-scale life cycle monitoring of high-rise buildings.

1.13.1.3 Purpose of the Instrumentation The monitoring strategy was developed based on the different criteria presented in detail in the literature (Glisic et al., 2003). The main requirements were to perform long-term SHM of a representative number of critical structural members, and that is why fiber-optic sensors were selected. Ten ground floor columns were selected as being the most critical elements in the building, whereas the



FIGURE 1.15 Building at Punggol EC26: (a) during construction and (b) upon completion (Courtesy of SMARTEC, HDB, and Sofotec).

number of sensors was adapted to the available budget. The identification of columns was based on numerical modeling. The dominant load in each column is an axial compression force; therefore, it is supposed that the influence of bending on deformation can be neglected. Consequently, a single sensor per column was installed parallel to the column axis. The gauge length of the sensors was determined with respect to the available height of the column (3.5 m) and on-site conditions; hence, 2 m long sensors were used.

Several columns at the same level (storey) in a building of simple topology create a so-called scattered simple topology (Glisic and Inaudi, 2007). The analysis of a scattered topology consists of several algorithms that are to be applied in order to assess the global structural behavior. Supposing the floor slab to have high stiffness, the columns supporting the slab are expected to deform by similar absolute values and the total strains in different columns are expected to be in mutual linear correlation. The column with foundation subject to settlement will elongate while the neighboring columns will shorten. In this case, the linear correlation between the elastic strains of the columns is lost.

1.13.1.4 Examples of Outcomes Average strain evolution is for 10 monitored columns presented in Figure 1.16. A period of 9 years is presented in the figure and several events are highlighted. Three periods in strain evolution can be identified in Figure 1.16: (1) May 2001–July 2002 is construction period and all the factors: load, creep, shrinkage, and temperature variations have visible influence to total strain; (2) July 2002–January 2005, creep and shrinkage have dominant influence to total strain (temperature influence is minimized by reduced exposure of columns to direct sun); (3) January 2005–May 2010, rheologic strain components stabilize slowly and the daily thermal strain fluctuations become more visible.

Detailed data analysis is presented in literature Glisic and Inaudi (2007) and Glisic et al. 2010b, and here only the summary of main findings is presented.

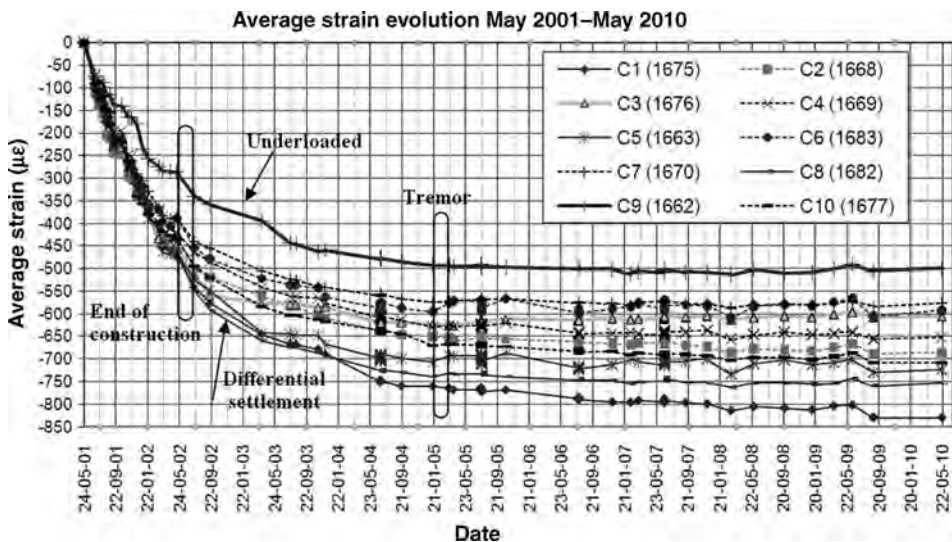


FIGURE 1.16 Nine-year record of strain evolution in the building (Courtesy of HDB, SMARTEC, and Sofotec).

During construction, Column C9 had the strain that is significantly lower than other columns, though its change is proportional to other columns in the period of construction. The explanation for this behavior is found in lower load level at this location.

In Singapore, the temperature is very stable over the year, and only small fluctuations of strain are registered on a daily and yearly basis.

Earthquake in neighboring Indonesia in 2005 generated a tremor that was felt by inhabitants of the building. A measurement session was performed immediately, and as the strain change measured after the tremor was within the ranges measured before and after the tremor, the conclusion was that the tremor did not affect the structural behavior of the columns.

Differential settlement of foundation is detected in column C3. The strain in this column had significantly lower increase in compression than the other columns, and in the past few years even decrease of compression is noticed. The evaluated differential settlement of the column is in range of 1 mm, and thus it is inoffensive for the building performance.

1.13.1.5 Benefits of Using SHMS Technologies in the Project Monitoring strategy has been developed and monitoring system applied for the first time in long term in high-rise buildings. The monitoring helped (1) to understand the real structural behavior of the building columns, (2) to evaluate creep and shrinkage, (3) to determine underloading (over-dimensioning) of columns C9 and improve structural design for the future, (4) to immediately evaluate the impact of tremor to the building, and (5) to detect differential settlement in column C3. The employed monitoring system performed very well and all the sensors function properly more than 9 years after embedding in concrete. Unusual behaviors were detected at early stage demonstrating high sensitivity of sensors and suitability of monitoring strategy.

1.13.2 The New Årsta Railway Bridge, Sweden—2005

The New Årsta Railway Bridge in Stockholm is an optimized, very slender and complex 11-span prestressed concrete structure. It is located to the west of the Old Årsta Bridge crossing the bay Årstaviken. The bridge was built as part of an upgrading from two tracks to four between Stockholm South and Årstaberg further south. The bridge accommodates two tracks for railway traffic, a service road and a pedestrian and cycle road. The slender design of the bridge and complicated construction methods were reason that initiated The Swedish National Railway Administration (Trafikverket, former Banverket) to install a monitoring system on the bridge. The bridge was opened for traffic on 2005 (Enckell, 2006; Wiberg, 2006).

1.13.2.1 Quick Facts

Name and Location: The New Årsta Railway Bridge, Sweden

Owner: Swedish National Railway Administration (Trafikverket, former Banverket)

Structure Category: Continuous prestressed concrete girder bridge

Spans: 11 spans: 1×65 m, 9×78 m, 1×48 , 15 m

Structural System: Prestressed concrete

Start of SHMS: 2003 during construction



FIGURE 1.17 The New Årsta Railway Bridge during construction.

Number of Sensors Installed: 56 sensors

Instrumentation Design by: Collaboration between Royal Institute of Technology (KTH), BBK AB, SMARTEC SA, and BEMEK AB.

1.13.2.2 Description of the Structure The new Årsta Railway Bridge is very complex 11-span prestressed concrete girder structure (Figure 1.17). The bridge is 833 m long and 19.5 m wide. The headroom under the bridge is 26 m. The trains are running on two unballasted tracks and the bridge accommodates a pedestrian track on the west side and a service road in the east side of the bridge.

The cross-section has a very soft form: the design is very slender with a superstructure that is thickest above each pier and then gets thinner toward each mid span where it is as narrow as possible according to design calculations. Every pier has a 2-m thick transversal beam and manholes are located on the bridge deck in every span, between the rails.

Concrete is of red color as it was desired that the new bridge design would fit the old bridge design and the surrounding landscape. The slender design of the bridge required a high amount of reinforcement and in order to satisfy the needs for bearing capacity concrete of the quality class K60 was used. The total cost of the bridge is calculated to about SEK1.2 billion.

1.13.2.3 Purpose of the Instrumentation The Swedish National Railway Administration (Trafikverket, former Banverket) ordered a SHMS consisting of fiber-optic long-gauge sensors, thermocouples, strain transducers and accelerometers. A research project was connected to the project in order to study and understand the dynamic and the static behavior of the bridge. The main objectives were to monitor a chosen characteristic span of the bridge during 10 years including the construction phase and the testing phase in order to

- Verify that design assumptions concur with reality
- Improve knowledge of traffic and temperature effects
- Document changes in static strain and dynamic properties

- Ensure that elements have not been subjected to excessive loading
- Obtain information concerning the condition of the structure that has consequences for maintenance and repair operations
- Learn about SHM and related issues.

1.13.2.4 Examples of Outcomes As handling with emerging technologies in the field without previous experience, some problems occurred, especially during the construction period. Some sensors stopped working after installation or casting and were probably broken by reinforcement workers or vibrating during casting. Two sensors are not in pretension and therefore unable to measure: these sensors are not damaged but were either gliding after the installation or were not pretensioned enough during the installation procedure. Some cables were also damaged by workers after installation.

On the other hand, the measured results confirmed the behavior of the concrete from the casting to the operation: initial swelling, drying shrinkage, creep, shrinkage, loading, unloading, and prestressing are verified. Sudden strain level changes were also noted and changes in daily variations revealed cracking in the structure.

Figure 1.18 illustrates the prestressing steps from 5 to 8 of May 2003. Figure also shows when sensor AS4, that was either gliding or not prestressed enough in the installation phase, stops measuring after the last prestressing stage. Taking off the formwork; 12 of May 2003 can also be seen in the figure.

Figure 1.19 shows the results for sensor AS1 during a load test. Measure is taken about every 7 min that is the time to measure the whole system of 40 fiber-optic sensors and thermocouples. The measured strain is expressed as microstrain, meaning strain of 1×10^{-6} .

More results for early age, construction, long term, temperature, and cracking can be seen in Enckell (2006, 2007).

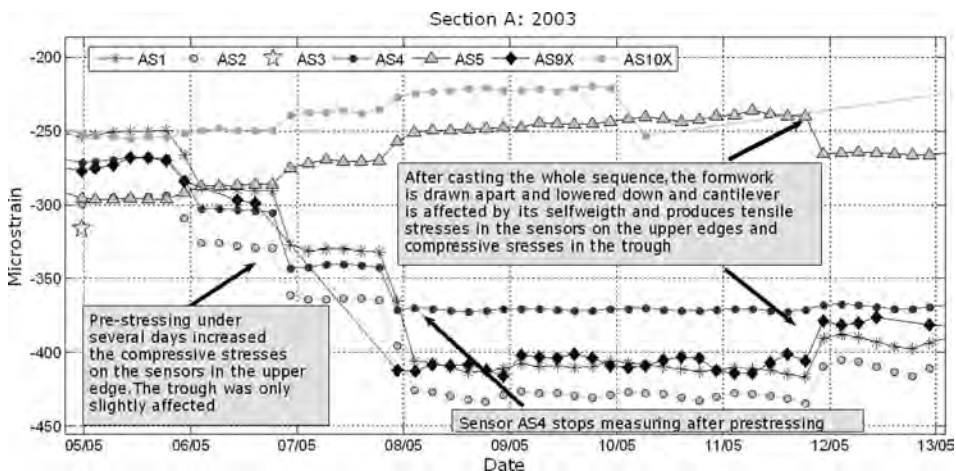


FIGURE 1.18 The figure illustrates measured results for the New Årsta Railway Bridge: the prestressing of the cross-section, the lowering down the formwork and the failure of the sensor function.

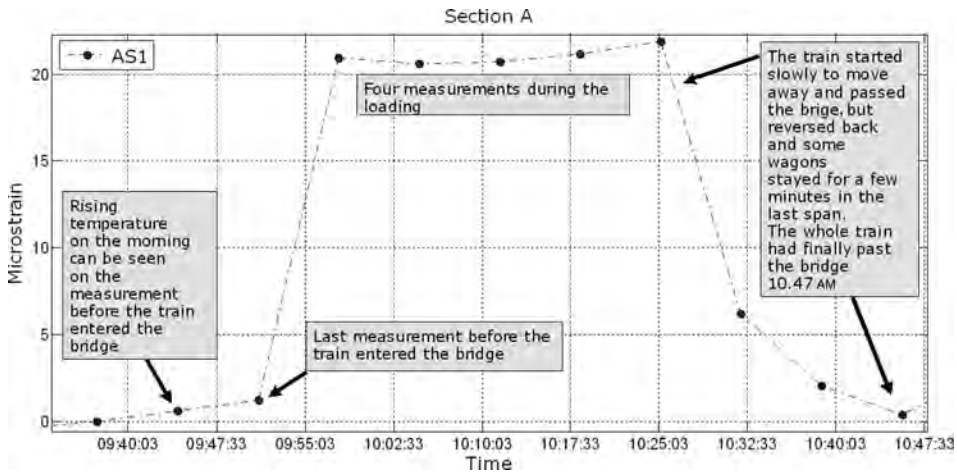


FIGURE 1.19 The figure illustrates measured results for the New Årsta Railway Bridge during a load test: the maximum strains that occurred for sensor AS1 during the train on the bridge were 20×10^{-6} in elongation.

1.13.2.5 Benefits of Using SHMS Technologies in the Project As the monitoring has carried out during several years, project has produced material and a lot of results are available. Two licentiate thesis, one doctoral thesis, conference papers, and few technical reports are published (Enckell et al., 2003a, 2003b; Enckell and Wiberg, 2005; Enckell, 2006; Wiberg, 2006; Wiberg and Enckell, 2008).

Installation, measurements, and data analysis has been done and a lot of heuristic knowledge has been learned. Significant phases of bridge life were registered including early age deformation, prestressing, removal of the forms, and long-term creeping and shrinkage. Unusual behaviors that revealed cracking were detected and can be used in maintenance planning. Comparison between fiber-optic sensors and strain transducers were also made and proofed the superior quality of the static fiber-optic measurements in long term.

1.13.3 Stonecutters Bridge, Hong Kong—2010

The Stonecutters Bridge is a cable-stayed bridge with a main span of 1018 m (see figure 1.20). The main span is supported from two single central towers both placed on land providing a clear entrance to the container port with a vertical clearance of minimum 73.5 m.

1.13.3.1 Quick Facts

Name and Location: Stonecutters Bridge, Hong Kong, SAR, China

Owner: Highways Department

Main Consultant: Ove Arup & Partners Hong Kong Limited

Structure Category: Main bridge: Cable-stayed bridge, double cable plane, and two girders.

Spans: Main span 1018 m.



FIGURE 1.20 Stonecutters Bridge, Hong Kong, SAR, China (Courtesy of COWI).

Structural System: Composite steel—reinforced concrete towers and steel box girders inter connected with cross girders. Back span as concrete girder

Start of SHM: 2010

Number of Sensors: 1420 sensors

System Concept and Functions Initiated by: Bridges and Structures Division of Highways Department, the Government of HKSAR.

Instrumentation Design by: COWI (as subconsultant to Arup).

System Operation by: Bridges and Structures Division of Highways Department, the Government of HKSAR.

1.13.3.2 Description of the Structure The 53.5 m wide bridge deck consists of twin box girders connected by cross girders. The stay cables connect to the outside edges of the deck only. The deck is in steel in the main span and 50 m into the first back span while the rest of the back spans are in concrete.

1.13.3.3 Purpose of the Instrumentation The Stonecutters Bridge Structural Health Monitoring System is an example of a distributed monitoring system with the provisions for all the aspects of health evaluation based on automatically processing of data.

In designing, the Structural Health Monitoring Systems for Stonecutters Bridge, COWI and Highways Department took full account of the valuable experience gained in operating the Wind and Structural Health Monitoring System (WASHMS) for Tsing Ma Bridge, Kap Shui Mun Bridge, Ting Kau Bridge and the cable-stayed bridge (Hong Kong Side) in Hong Kong—Shenzhen Western Corridor. In order to obtain a whole life health record of Stonecutters Bridge, a construction stage structural health monitoring system will be implemented in addition to a more conventional operation stage structural health monitoring system.

The design challenges for the WASHMS was primarily to monitor the aerodynamic loads free of the turbulence generated by the bridge, temperature loads and responses, stress monitoring of cable stays with record length over 500 m and advanced deformation monitoring based on GPS and accelerometer readings analyzed by operational model analysis.

Also, the great design effort was put into the user interface and data management system in order to prevent users of the system to be drowned by data overload from the +1500 sensors installed on the bridge.

1.13.3.4 Planned Outcomes of the SHMS In order to establish a link between the events logged by the WASHMS on the bridge and the Bridge Management System and Bridge Rating system for the structural, electrical and mechanical components of the bridge was defined, to give a clear and common basis for observations either logged by the WASHMS.

The bridge rating is based on a point ranking concept. The ranking is general which means that the method can be used on both structural components and installations. For maintenance management of structural components, it is necessary to consider the risk imposed by the possible failure modes of the individual component. For that purpose, the point ranking method has been used.

The bridge rating system will provide rational basis for prioritization of inspections and maintenance on primary and secondary structural components. The categories, primary and secondary components, are related to the load capacity analysis model. The system will be based on the results from the principal inspection and the WASHMS. By using these two in combination the additional inspections and maintenance work can be initiated in a proactive manner.

The rating system will ensure that the needed actions to be taken in order to keep structures safe will be taken in time. Such actions include structural repair and strengthening as well as protection against environmental actions.

The WASHMS has been designed divided into three subsystems, namely, the On-Structure Instrumentation System (OSIS), the Portable Data Acquisition System (PDAS), and the Computer System for Operation and Control (CSOC). Furthermore, the WASHMS includes a system for monitoring of traffic by digital video cameras, the DVC monitoring system.

The OSIS is based on a number of data acquisition units (DAQ units) that acquire data on the bridge and transmit these data to the OSIS Controlling Computer through a fiber-optic data network.

The PDAS equipment is compatible with the OSIS network architecture, ensuring that the PDAS can be connected to the OSIS system at the bridge for on-line or batch transfer of data to the central computer.

The CSOC is used for monitoring and analysis of the data acquired by OSIS. This includes alarms generated by OSIS when preset limit values for traffic, wind or structural loads are exceeded. Alarms can be used as inputs to the Bridge Rating system, and they can be transferred to the traffic police.

A bridge rating system is established to be used in the systematic approach described in the inspection and maintenance concept in the maintenance and management system (MMS).

The functioning of the bridge rating system is based on the results from the principal inspection and data from WASHMS. By using these data in combination, it shall be possible to initiate the additional inspections and maintenance works in a proactive manner.

In praxis, the bridge rating system will receive structural events from the monitoring system. This can either be events measured by the WASHMS or events logged manually in the event database based on visual inspections. By the properties of the event, the system will automatically update a bridge inventory according to the principals for the point

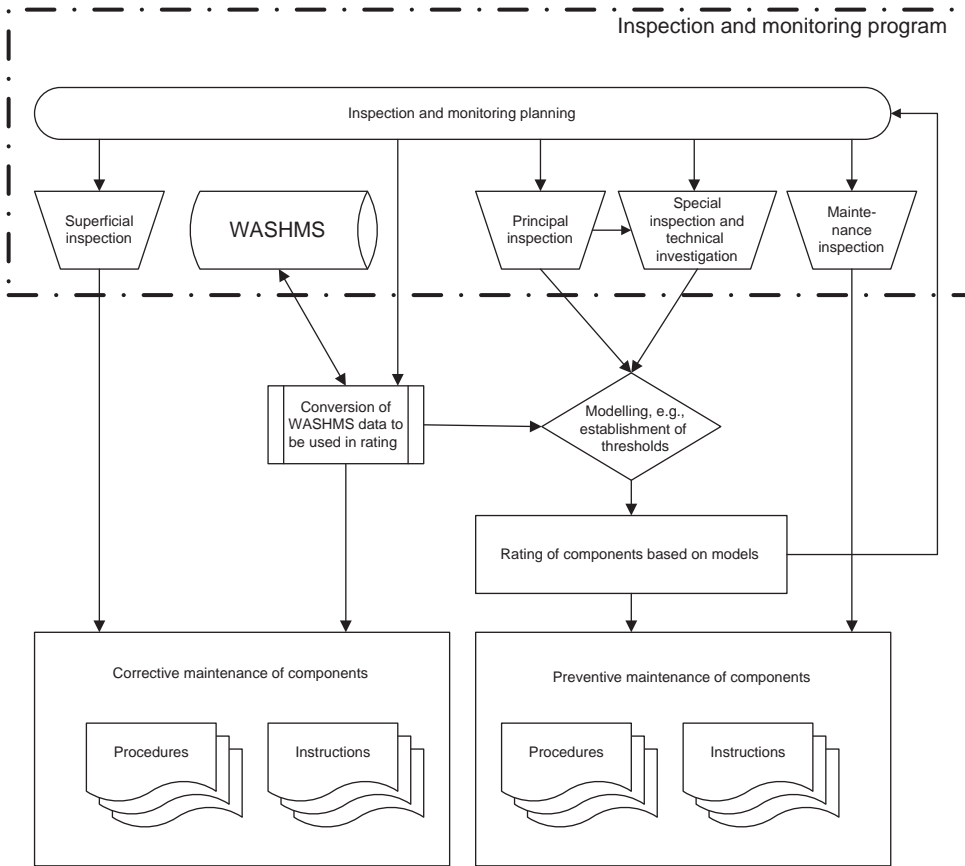


FIGURE 1.21 WASHMS as an integrated part of the management system.

ranking. This can be the planning tool for the maintenance managers in order to evaluate what work orders for maintenance inspections shall be issued; to optimize the works and to track and plan maintenance budgets.

Figure 1.21 shows WASHMS as an integrated part of the management system similar to the inspections. This means that the management may benefit from WASHMS in the long-term planning of preventive maintenance and in the day-to-day corrective maintenance. Both superficial and maintenance inspections are routine inspections that creates no information to be included in a rating process. It shall be emphasized that one should not rely fully on the data and results from WASHMS but on a combination of monitoring and inspections done by experts.

1.13.4 Severn River Crossing, UK—2010

Severn Suspension Bridge, which links England to South Wales in the United Kingdom is a conventional suspension bridge. It was opened to traffic in 1966. The total length of the bridge is 1597 m, and it accommodates also cycle tracks and footway. Traffic flows increased by 63% between 1980 and 1990, and further increases were expected in the years ahead. Therefore, the Severn Bridge crossing was strengthened and resurfaced in

the late 1980s. As there were severe congestion problems in the summer and at peak times each day, a Second Severn Crossing construction was started in 1992 and the second bridge was opened to traffic in 1996.

1.13.4.1 Quick Facts

Name and Location: Severn River Crossing, United Kingdom

Owner: Severn River Crossing for the UK Highway Agency

Structure Category: Suspension Bridge

Spans: Main span 9876 m, side spans 3048 m each; total length approximately 1597 m

Structural System: A suspension bridge with the orthotropic steel box girder deck supported by two main cables slung between two steel towers

Start of SHMS: 2008

Number of Sensors Installed: 90 passive acoustic emission sensors

Instrumentation Design and Supplied by: Physical Acoustics Limited (PAL).

1.13.4.2 Description of the Structure The Severn Bridge was opened in 1966 to replace the ferry service crossing from Aust to Beachley. The new bridge provided a direct link for the M4 motorway into Wales (<http://www.severnbridge.co.uk/>).⁹ Length of the main span is 9876 m and lengths of the side spans are 3048 m each. The high tensile steel towers rise to 136 m above mean high-water level and are of hollow box construction. The deck is an orthotropic steel box girder of aerofoil shape with cantilevered cycle tracks and a footway. Suspension cables carrying the deck are not vertical as in most suspension bridges, but rather arranged in a zigzag style in order to reduce vibration.

1.13.4.3 Purpose of the Instrumentation The Severn Bridge has carried more than 300,000,000 vehicles since it was opened in 1966. In 2008, Severn Suspension Bridge underwent invasive investigation of the main cable that revealed corrosion in the individual steel wires and also a small number of broken wires. A dehumidification system and a full AE wire break detection system for long-term evaluation of suspension bridge (see Figure 1.22). Main cables over the whole 1673 m of each cable were installed in order to ensure the continued integrity and safety of the structure (<http://www.ndtnews.org/> 2008). The AE monitoring system monitors continuously sounds for individual wires breakages within the main cable in order to keep track of the rate and location of any further wire breaks. The system enables evaluating the health of the structure and can also to identify regions that might require further cable investigation. The monitoring system consists of 90 passive acoustic emission sensors as well as six local data interrogators that are distributed along the bridge on an optic fiber network. Sensors are attached to cable clamps every 36 m along the whole cable. Signals from the main cable are detected, automatically processed by local interrogator and reported to a “base-station” on the network. There the data is processed and potential wire breaks are automatically reported via a secure website by e-mail.

1.13.4.4 Examples of Outcomes The wire break system has been successfully proven to detect artificial wire break sources at all points of the cable during the commissioning to

⁹Accessed February 25, 2011.



FIGURE 1.22 (a) Sensor installation. (b) Acoustic emission sensor mounted on cable band (Courtesy of Physical Acoustics Ltd.)

within 0.5 m or less. Unusually in this type of monitoring as the wire break is hidden from view, there has also been a third party verification of a real wire break. This took place in 1 m section of unwrapped cable and was witnessed by the maintenance engineers working on the bridge. The system identified, analyzed, and reported the wire break event automatically and calculated the location within 0.08 m of the actual observed break.

1.13.4.5 Benefits of Using SHMS Technologies in the Project The wire break monitoring system is capable of detecting and locating wire breaks between the cable saddles at either end of the main cable. The monitoring system offers valuable information regarding wire breakage where no other technology can. In comparison, opening of cable by wedging limited to a small percentage of the overall cable length, with the visual inspection at that point showing approximately 5% of wire. The data from the system would allow identification of any points of interest on the main cable, allowing focused inspection at that point. The system also acts as a safety system, monitoring overall cable integrity which would alert the bridge owners if wire break occurs and also if there was a change or significant increase in the number of wire breaks. After the initial installation costs, the long-term monitoring and maintenance is low and automated, this is needed for long-term monitoring.

1.13.5 A4 Hammersmith Flyover, UK—2010

Hammersmith Flyover, constructed in 1961, is 630 m long bridge with 16 spans varying between 33 and 47 m. It is one of the earliest examples of a post-tensioned segmental viaduct in the United Kingdom. Managed by Transport for London, it carries a high volume of traffic, including a large percentage of heavy goods vehicles, on the A4 over the

intersection of four major routes. Post-tensioned concrete has been an economically competitive form of construction for medium to long span structures over the past 50 years. During the 1980s, however, a number of defects were discovered in this form of construction around the world. Corrosion can remain hidden until the loss of strength is sufficient to cause failure without warning, such as in the case of Yns-y-Gwas bridge in Wales in 1985.

1.13.5.1 Quick Facts

Name and Location: Hammersmith Flyover, West London, England

Owner: Transport for London

Structure Category: Post-tension concrete bridge

Spans: 33–47 m spans with total length approximately 630 m

Structural System: Precast, segmental concrete box girder with post-tensioning.

Start of SHMS: 2010

Number of Sensors Installed: Approximately 300 passive acoustic emission sensors

Instrumentation Design and Supplied by: Physical Acoustics Limited (PAL).

1.13.5.2 Description of the Structure Hammersmith Flyover is an elevated roadway in west London. The superstructure of bridge consists of a series of 3 m long triple-cell spine box beam units with 75 mm in situ concrete joints to 300 mm cantilever diaphragm units. The longitudinal post-tensioning consisted of 2 sets of 4 tendons of 16, 19-wire strands either side of the two internal webs encased in in situ grout “boxes.”

Hammersmith Flyover was one of the first examples of an elevated road employing reinforced concrete balanced cantilever beam supports with a single central column.

1.13.5.3 Purpose of the Instrumentation A detailed special inspection was carried out on the Hammersmith Flyover by exposing the tendons at critical locations. Found section losses were used in a three-dimensional finite element load assessment model to determine the flyover’s current live load capacity. The results showed that despite the evident corrosion, the structure passed full 40t highway loading in accordance with BD21/01. While this gave reassurance that the structure was safe and “fit-for-purpose,” visual inspection at a limited number of locations was not considered to be sufficient by Transport for London. To manage the structure effectively, transport for London needed to know whether the post-tension cable deterioration was continuing and if so, at what rate and where wire breaks were occurring. Acoustic wire break monitoring is deployed over the east side of the flyover (nine spans) to detect and triangulate wire breaks linearly along the structure and define the location to specific tendon groups.

The specified accuracy of the system required a high-frequency acoustic emission sensor in each 3 m segment. Sensors are bonded to the surface of the concrete with structural adhesive and cabled back to the monitoring system in the instrument cabinet.

Detection of the wire breaks uses the signature data collected during third party Japanese Public Highways Corporation trials and those detected on other structures. The distinct signature and signal transmission from breaks allows automatic detection and instant wire break alerts. Signal location uses time of arrival of the energy wave (released during wire break) at acoustic sensors. These data are cross correlated with an acoustic map of the structure developed during the commissioning of the system that confirms the exact location, even where tendon groups are in close proximity. This method takes into



FIGURE 1.23 Blind wire break trials during system commissioning (Courtesy of Physical Acoustics Ltd.)

account complex wave paths caused by mortar condition and variable contraction (Bridge Design and Engineering Autumn 2010).

In addition to the acoustic emission system, a strain, displacement and rotational structural monitoring system and a total station system was also installed.

1.13.5.4 Examples of Outcomes Once installed, the wire break system operation was verified by cutting a single exposed wire in a blind test, see Figure 1.23. This was done by incrementally reducing the section of a single wire over a period of time using a grinder until the wire yielded and broke. The break was remotely detected and triangulated to within 100 mm of the physical break and to the correct tendon group. The system will operate for at least 5 years monitoring any change in the condition of the post-tensioning, highlighting any specific areas of concern and any significant increase in wire break activity.

1.13.5.5 Benefits of Using SHMS Technologies in the Project The system was operational from the end July 2010 to present and no results have been publicized. However, generally, acoustic emission monitoring of post-tension structures is widely acknowledged to be the only way to detect and locate wire breaks, which occur hidden inside ducts. The technique is nondestructive so it does not require any disturbance of ducts which, through the refreshing of oxygen, can reinitiate the corrosion process. The system once set up runs and reports automatically so costs are low. With historical sudden failures of post-tension structures, such as Yns-y-Gwas Bridge in Wales, monitoring of any post-tension bridge with a suspected problem is recommended.

1.13.6 Streicker Bridge, United States—2010

The Streicker Bridge is a new pedestrian bridge at Princeton University campus. Funded by Princeton alumnus John Streicker (class of 1964), the bridge was designed by renown Swiss engineer Christian Menn in collaboration with the HNTB architecture and engineering firm, whose lead engineers for the project were Princeton alumni Theodore Zoli

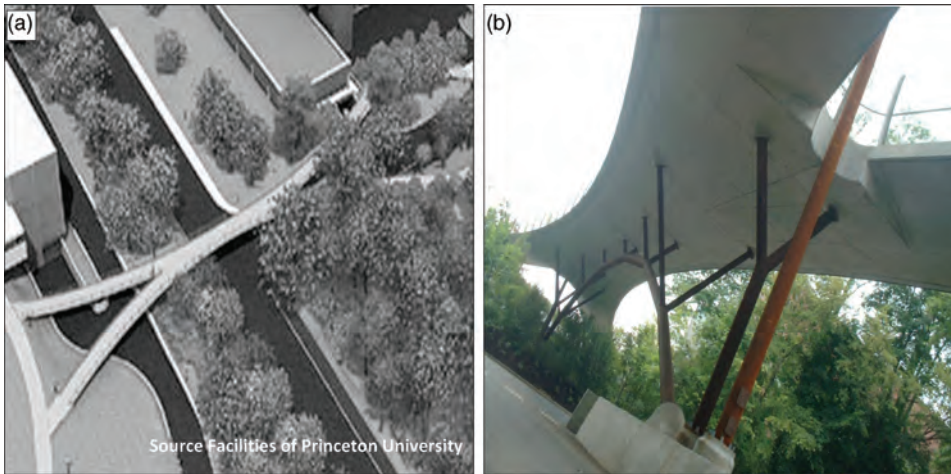


FIGURE 1.24 Rendering (a) and photograph (b) of Streicker Bridge.

(class of 1988) and Ryan Woodward (class of 2002). The rendering and the photograph of the bridge are shown in Figure 1.24.

Besides its primary aim, to provide and facilitate safe pedestrian crossing-over the Washington Road, the bridge has strong symbolic and aesthetic significance. Taking into account its scientific, social, and symbolic measures (Billington, 1983), the Streicker Bridge can be considered as a piece of structural art. Created as a part of Princeton University's Natural Sciences neighborhood, the bridge connects the Icahn Laboratory (Lewis-Sigler Institute for Integrative Genomics) and new neuroscience and psychology building (under construction) on the west with Jadwin Hall (physics), Fine Hall (mathematics), new Chemistry building, and Lewis library on the east. The bridge will "stand as a tangible symbol of the cross-disciplinary collaborations that are central to scientific research and teaching today" (President Tilghman in News at Princeton 2006). The bridge's "X" shape in plane symbolizes these cross-disciplinary collaborations, and the arch itself represents the south entrance gate to Campus.

1.13.6.1 Quick Facts

- *Name and Location:* Streicker Bridge, Princeton University, NJ, United States
- *Owner:* Princeton University
- *Structure Category:* Pedestrian bridge, deck stiffened arch, curved continuous girders ("legs")
- *Spans:* Main span 34.75 m, total length approximately 104 m, maximum "leg" length 40 m (maximum "leg" span 19 m)
- *Structural System:* Post-tensioned concrete deck and weathering steel arch and columns
- *Start of SHMS:* 2009 (during construction)
- *Number of Sensors Installed:* 53 discrete long-gauge strain sensors, 47 temperature sensors, 40 m of distributed strain sensor
- *Instrumentation Design by:* SHM lab at Princeton University.

1.13.6.2 Description of the Structure The Streicker Bridge has a main span and four approaching legs. Structurally, the main span is a deck-stiffened arch and the legs are curved continuous girders supported by steel columns. The legs are horizontally curved and the shape of the main span follows this curvature. The arch and columns are weathering steel while the main deck and legs are reinforced post-tensioned concrete. The slender and elegant deck-stiffened arch represents an efficient solution to bridge the span of 34.75 m (114 ft) keeping the deck thickness of only 578 mm (22.75 in.) and arch diameter of 324 mm (12.75 in.). The slender cross-section is maintained in all four legs. Each leg had different length and different span length that were imposed by the surrounding roads and the landscape. The longest is the 40-m long south-east leg, and it contains the longest span of approximately 19 m. Turner Construction Company was the main contractor and university's Office of Design and Construction was the supervisor.

1.13.6.3 Purpose of the Instrumentation SHM Lab at Princeton University instrumented the bridge with various SHM systems, with aim to transforming it into an on-site laboratory for various short- and long-term research and educational purposes (Glisic, 2011). The research will focus on addressing general SHM challenges such as bridging education gap between research and practice, collecting real structural behavior data sets, identifying changes in strain patterns caused by unusual behaviors, and characterizing the SHM contribution to sustainability of built environment. Research in domain of monitoring approaches, methods and instrumentation, and evaluation of lifecycle cost benefits in function of the long-term SHM approaches is planned as well (Glisic and Adriaenssens, 2010). A long-term objective of this project is to develop a holistic framework that will transform the current inefficient fragmented approach to SHM and eventually lead to the widespread, comprehensive, and beneficial application of SHM as a tool for optimized bridge management. The SHM of Streicker Bridge serves as a support to university courses on SHM and several other courses related to structural analysis and design. The bridge represents tangible demonstrator for students and practitioners, and it is open for visits of community members.

At the present stage, the bridge instrumentation is based on two monitoring approaches: global structural monitoring using long-gauge FBG strain sensors, and integrity monitoring using distributed fiber-optic sensing based on stimulated Brillouin scattering (Brillouin Optical Time Domain Analysis or BOTDA). Taking into account that the main aims of monitoring are related to research and education, not all the components of the bridge were instrumented. Assuming symmetry and similarity in structural performance, it was decided to equip only half of the main span and only one approaching leg, that is, south-east leg with sensors.

The fiber-optic technologies were used in this initial project phase because they have proven durability and feature very good long-term stability and insensitiveness to external environmental and man-made influences. In the future, the other monitoring systems and other approaches will be applied. The sensor network design is based on fiber-optic methods using loose structural analysis approach (Glisic and Inaudi, 2007). The total number of the sensors was determined as a trade-off between the performance and the cost. Installation was performed by undergraduate and graduate students of Princeton University (see Figure 1.25). The sensors were embedded in concrete during the construction. The sensors attached to rebars before pouring are shown in Figure 1.25.

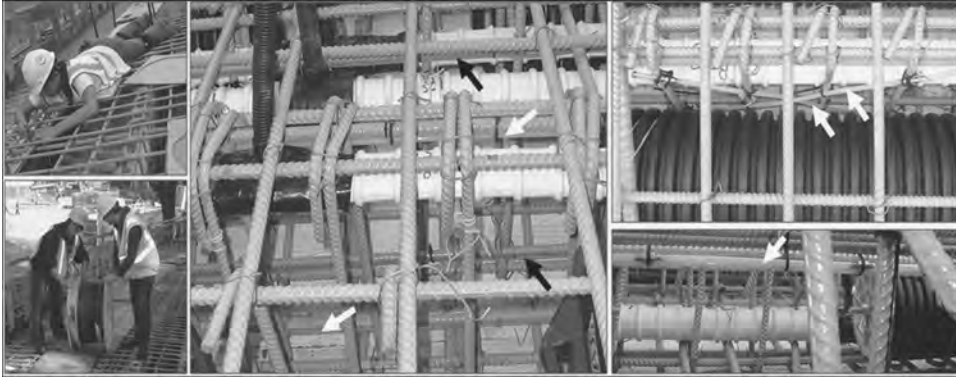


FIGURE 1.25 Photograph taken during the installation (from upper-left clockwise): student installing the sensors, parallel discrete and distributed sensors installed onto the rebar, crossed sensors, package of three sensors with different gauge lengths, and students installing extension cable.

1.13.6.4 Examples of Outcomes Static monitoring started for the main span approximately 6 h after the pouring of concrete, while for the south-east leg it started immediately after the pouring. The data were registered with the pace varying from 5 to 15 min for discrete FBG sensors, and approximately 1 h for distributed BOTDA sensor. Full presentation of results would exceed the contents of this chapter, which is why only the most illustrative findings are presented. Since the project is still in its initial stage, the analysis of the presented results is to be considered as preliminary.

The first month of measurement performed on the middle segment of the main span is given in Figure 1.26a. Various works and events were successfully detected and they are indicated in Figure 1.26a: thermal contraction due to hydration process (swelling was not registered since the monitoring started 6 h after the pouring), post-tensioning, de-centering, and removal of the forms. No unusual behaviors were detected in this period.

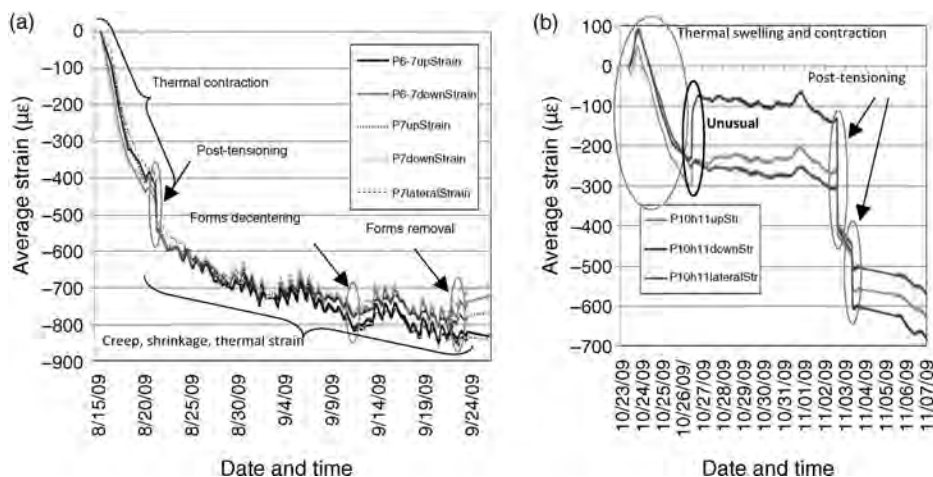


FIGURE 1.26 Early age measurement in the main span (a) and the south-east leg (b).

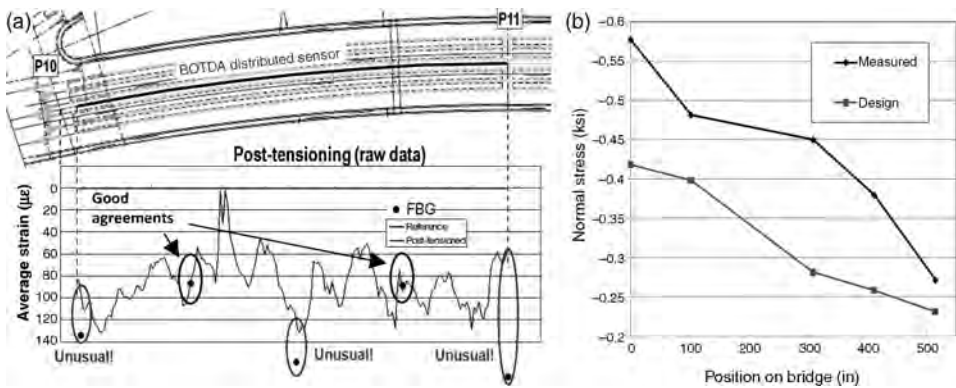


FIGURE 1.27 Comparison between distributed and long-gauge sensor measurements (a) and comparison between design and measured stresses in the cross-sections centroids of the main span (b) (Liew 2010).

The early age measurements for middle of the longest span the south-east leg are presented in Figure 1.26b. Thermal swelling and contraction due to hydration process and the post-tensioning of the deck were detected and indicated in the figure. However, besides these usual behaviors, an unusual event is detected—increase in strain—by all the three sensors. Similar unusual behaviors were noticed in some other sections of the south-east leg. The nature of this event is still under investigation, but preliminary study indicates the early age cracking as the cause. The unusual strain was rectified in all the sections by the post-tensioning. Further analysis of the detected event represents the topic of a graduate student master thesis.

Post-tensioning as measured with distributed sensor is given in Figure 1.27a. The comparison with long-gauge FBG sensors is shown in the same figure. Two monitoring systems have good agreement in quarter span sections where no unusual behavior occurred. Discrepancy is noticed in the section where the unusual behaviors were detected. This discrepancy is a consequence of 1-m spatial resolution of distributed system, which makes it impossible to accurately detect strain concentrations, unless a special software module is enabled. Comparison of two systems is the topic of an undergraduate student senior thesis.

Based on measurements, the actual post-tensioning stresses are calculated and compared with design as shown in Figure 1.27b. The measurements show that in reality, post-tensioning losses are likely to be smaller than expected (Liew, 2010).

1.13.6.5 Benefits of Using SHMS Technologies in the Project Important phases of bridge life were registered including early age deformation, prestressing, and removal of the forms. Unusual behaviors were detected in several sections of the south-east leg, and based on preliminary analysis they are attributed to the early age cracking. Post-tensioning closed the cracks successfully. The post-tensioning stresses in concrete were evaluated based on measurements and compared with design. The comparison shows higher stresses than designed, indicating smaller post-tensioned losses and better performance of the bridge (Figure 1.27).

Comparison between two monitoring systems was successful and at locations where no unusual behaviors were detected the agreement between the systems was within the error of measurements. At locations where unusual behaviors occurred a discrepancy

between the systems were observed due to 1-m spatial resolution of distributed monitoring system.

Undergraduate and graduate students were involved in various phases of the project including installation, measurements, and data analysis. Project generates material for an undergraduate and a graduate course at Princeton University and for several senior, master, and Ph.D. theses.

1.13.7 Messina Bridge, Italy—2018

The planned Messina Strait Bridge will connect the coasts of Sicilia and Calabria in southern Italy. It is planned to carry a four-lane highway with emergency lanes and a dual railway line. The bridge is a suspension bridge with a world record breaking 3300 m main span. The design life of the bridge is 200 years (Figure 1.28).

1.13.7.1 Quick Facts

- *Name and Location:* Messina Bridge, Strait of Messina, Italy
- *Owner:* Stretto di Messina Spa.
- *Structure Category:* Main bridge, suspension bridge, double main cables and three girders, two for road and one for rail traffic
- *Spans:* Main span 3300 m, total length 3666 m, span lengths 333 m–3300 m–333 m
- *Structural System:* Steel towers and steel girders
- *Planned Start of SHMS:* 2014, during construction
- *Number of Sensors Planned to be Installed:* 3500 sensors
- *Instrumentation Detailed Design by:* COWI A/S for Eurolink Spa.

1.13.7.2 Description of the Structure The suspended deck is arranged with the cross girders spaced at 30 m as the main elements whereas the two roadway girders and the central railway girder are taken as secondary elements spanning between the cross girders. Thereby, the Messina Strait Bridge will be the first bridge in the World to adopt the triple box concept for the deck, which is 68 m wide. The main cables consist of twin



FIGURE 1.28 Messina Bridge, Italy (Courtesy Stretto di Messina).

cables spaced 1.75 m—that is, a total of four cables are required for the bridge. The sag to span ratio of the cables is fixed as 1:11.

The towers are frame structures with slightly inclined legs (inclination of $\sim 2^\circ$) and three connecting cross-beams. They are constructed in steel. The tower top level is at 382.6 m.

1.13.7.3 Purpose of the Instrumentation The SHMS has been developed in parallel to the design of the bridge. The SHMS will facilitate various tasks including construction monitoring, operation monitoring, design validation, and maintenance planning. The SHMS will be integrated into a sophisticated management and control system for the bridge and approach (road and rail) networks. A key function of the SHMS will be to provide information that will support decisions for both operational and maintenance tasks. Operation monitoring will be performed through continuous real-time monitoring of environmental conditions, traffic (road and rail) loads, as well as structural response, providing real-time assessment of the status of the structure for the protection of user safety. A facility for extrapolating data and simulating predictions of future loading conditions will be provided to enable improved bridge operation capability. Data that are recorded by the SHMS will also be used to update service life calculation and point ranking assessment of structural components of the bridge to assist with long-term maintenance planning.

The detailed arrangement of sensors has been developed in collaboration with the structural designers to produce a monitoring plan that will focus on relevant critical areas of the bridge. The layout of sensors and the structure of the SHMS are also sympathetic to installation of the sensors during fabrication of the structure as well as progressive activation for effective construction monitoring. The SHMS will consist of approximately 3000 measurement locations and portable sensors, of which 90% will be integrated into the permanent SHMS data-stream.

The Messina Bridge SHMS has been engineered for the following objectives:

- to provide data for design, construction, and performance verification
- to provide data for review of design loading and development of assessment loading
- to provide current information on load conditions for effective bridge operation
- to provide current information on the condition of structural components
- to provide data for maintenance planning
- to provide data for troubleshooting unforeseen structural problems.

1.13.7.4 Examples of Planned Outcomes The Messina Bridge is a structure that will push the boundaries of engineering knowledge and experience. The significant increase in scale of the whole structure pushes all elements to their physical and dynamic limits. All aspects of the design have been reviewed and scrutinized by leading researchers and engineers, tested in the wind tunnel, and analyzed with Finite Element modeling. A detailed understanding of the structural behavior has been developed through significant research. The Messina Bridge SHMS has been strategically engineered to provide data for design, construction and performance verification. If unforeseen structural problems develop, important data for troubleshooting will be provided. The Messina Bridge SHMS will also provide live information on load and environmental conditions as well as deflections for effective bridge operation. Data recorded by the SHMS will be used to prepare and update simulations of load conditions expected in the subsequent hours, which will

further enhance the effectiveness of bridge operation and in particular traffic management. Data will be stored in accordance with data management protocols, which are designed to optimize data storage while retaining important information. This data will not only allow for the review of events and structural performance, but it will also allow for the future review of design loading and development of assessment loading.

1.13.7.5 Benefits of Using SHMS Technologies in the Project The objectives of the SHMS are not only limited to structural performance; the SHMS has also been engineered to provide data for monitoring the internal environmental conditions, condition of structural components as well as to provide data for maintenance planning. The Messina Bridge will be provided with an extensive dehumidification system to control the internal environment of the bridge boxes, towers as well as main-cables. The SHMS will provide monitoring of the internal environment and therefore function of the dehumidification system. Maintenance associated with fatigue can become costly to repair, particularly if cracks propagate extensively following crack initiation. Significant efficiency in maintenance can be achieved if the fatigue life of components is known. The monitoring of stress cycles and the development of fatigue utilizations is an integral part of the SHMS. The monitoring of the mechanical devices as well as elements that are prone to wear is particularly important to the effective maintenance of the structure. Buffers have been provided at the towers to control the deck movement relative to the tower, while allowing an important change in articulation under seismic conditions. The maintained function of these buffers is important to the success of the design. Expansion joints and bearings are the components of bridges that are most commonly subject to repair and replacement at intervals throughout the life of the structure. The excessive wear of the expansion joints and bearings may also lead to deterioration of other components that are not straightforward to maintain and can be expensive to repair, whether by direct association to the relevant component or by modifying structural behavior.

1.14 DISCUSSION

1.14.1 Development of New and Emerging Technologies

Tremendous development took place in emerging technologies in recent decades and it is still ongoing. A consequence of the process is that a new sensory technology with advance data acquisitions tools is now available. Completely new advantages and challenges were presented to the civil engineering society working with these emerging technologies.

Many sensors are accurate and reliable and also easy to comprehend and work with if some education is given. Fiber-optic sensor technology provides us with precise, sensitive equipment with well-established data acquisition facilities. Some fiber-optic technologies are already well recognized with long-term experience while others just entered the market and are still under development. Many commercial products exist in applications measuring discrete and distributed strain, temperature, and pressure, but more development is needed in the areas of monitoring humidity, corrosion, crack detection, and combination of parameters in the same sensor.

Noncontact NDE techniques such as interferometric radars, photogrammetry, and thermography open new areas and show potential. They do provide various advantages: no need to stop traffic, no need to install sensors, portable equipment, and easy setup.

Data interpretation is an important issue as it is not always delivered with the sensory system. Data interpretation is also almost unique for every single structure as even two similar structures could have different problems if built in problems arose during the construction period. Some time after installation is normally needed to learn about the structural behavior related to daily and annual variations, loading conditions and environmental effects. After that trial period, it is easier to calibrate the SHMS to fit the defined requirements.

1.14.2 Obstacles

By experience, a proper understanding is only gained with long-term experience and several obstacles have been observed while working with new and emerging technologies. The following ideas are identified as the most important and should be further investigated in order to find solutions and encourage the use of emerging technologies as they can bring to SHM numerous benefits.

When planning projects, the lack of understanding from different stakeholders can sometimes be a hinder for emerging technologies. The experts in different areas that do have proper understanding for the specific technology are hindered by decision makers that are conservative and do prefer to stuck to old, conventional techniques. And not only that, new and emerging technologies are also rejected by civil engineers as some civil engineers do lack understanding of variability and advantages that these techniques can offer. Economical limitations are also a problem as they set limit for authorities when choosing the systems. Authorities, in their turn, also have to deal with political decisions that do control projects and influence its pace.

The quantity of collected data must be adapted to the needs of the project. Capital is wasted with too complicates systems with either sensor and/or data overload. No proper analysis of the results can be performed if the amount of data is huge and the time of interpretation is limited. These kinds of projects contribute to bad status and give negative picture of chosen technologies. Therefore, it is important to give general and simple information to all parts so that the field becomes more adaptive. To demonstrate performed projects and achieved goals is a good way to highlight the advantages.

1.14.3 Need for Education and Collaboration

As the subject is large and development ongoing, there is also need for education. The people involved (engineers, owners, financiers, and researchers) working and making decisions concerning advanced fiber optic, laser, and other techniques need to be skilled to be able to judge the reliability of these systems. Engineers working with these techniques should also be unassuming and take their time to learn and understand and thereby be able to use these techniques properly and efficiently.

Special needs emerge if any deviation is notified when working with installations of emerging technologies. Multiple skills including the ability to solve practical problems are required in harsh environment in the field. Capability to judge problems, find solutions and see consequences of decisions made is most favorable. Complexity of existing problems is not easy to understand if the heuristics lacks. The systems need also to be flexible so that modifications can be made during the process. Universities are providing courses in various different technologies but as the subject is large, it is not easy to distinguish what is most preferable. An expert working with emerging technologies does need a wide

education, understanding and experience for various techniques and needs also be continuously updated in relevant subjects as development is ongoing.

Emerging techniques are many times characterized by their advantages and the disadvantages, and challenges are totally left out. The importance of international collaboration cannot be neglected: advantages, disadvantages, and challenges need to be discussed to save resources and capital. Mistakes and malfunctions are also to be reported in order not to be repeated.

1.14.4 Future Use and Development

It is good to keep in mind that very high quality is needed in sensory technology. In order to reach good results, sensor production must be optimized and the sensors have to be quality controlled, preferable in a standardized way. Some standards do exist for a few technologies like for example acoustic emission that does have a long-term experience in other engineering applications and testing procedures.

Less established technologies need to be tested and go through defined procedures as FAT, SAT, or any other required testing if implemented in real structures as they do lack long-term function verification. Implementing emerging techniques also includes taking a certain risk, and these projects should therefore be performed initially in smaller scale so in case of malfunction the loss of capital is kept in acceptable frames. With gained experience larger projects can be gradually realized. Also, the documentation and presentation need to become more generalized so that the different stakeholders have easier to understand and compare the products that are available on the market.

New challenges for fiber-optic sensors lie on cable applications such as prestressed or poststressed cables in various applications; stay cable and main cable applications in bridges. It would be convenient to install these cables with embedded sensors; either discrete or distributed sensors that do measure humidity, strain, and temperature as the inspection of these cables afterwards are indeed complicated. Distributed strain sensors need packaging improvements that will make them easier to install on the surfaces of large structures.

Many of the technologies aforementioned also compete with each other, and it is many times difficult to tell which one is the most suitable for a specific application. Our experience is also that complex structures do require the combination of several of these technologies in order to meet all technical requirements and provide a SHMS with functionality, quality as well as redundancy.

1.15 CONCLUSION

This chapter is based on research, literature surveys, design work on major international bridges and long-term experience of practical use of emerging technologies in SHM on various bridges as well as on other structures. Major conclusions are presented here in order to point the direction for future focus and development.

Emerging technologies do provide high precision, high accuracy, stability, and redundancy. They perform real-time monitoring that saves capital and improves the safety of the structure. Possible problems are faced at the beginning and an instant measure can be taken in order to correct the malfunction in monitored structures.

Emerging technologies together with adequate data acquisition contribute to accurate SHM methods in harsh environments. Nevertheless, emerging technologies do require a new kind of philosophy as they are characterized with certain uncertainty and intricacy. Interpretation of results is not always as easy as described in literature and experienced staff is essential for successful SHM. Need for wider understanding for relevant stakeholders is also significant. Necessity for new education is thereby emerged: engineers with practical skills are needed when working in the field with emerging technologies.

Experienced experts need to evaluate FATs, SATs, various installation issues and other related issues in order to judge the reliability and thereby provide high-quality systems that fulfill the initial technical requirement. A simple system consisting of smaller amount of sensors with proper data analysis is better than a complicated system with numerous sensors but with poor analysis. SHMS for complicated structures also requires combination of several techniques in order to be able to measure all desired parameters. As the development is ongoing, continuous update and education for people working with emerging technologies are required.

International collaboration and openness are needed for faster development: both advantages and disadvantages are presented and highlighted in truthful manner to avoid repetition of mistakes. Sensor production and especially quality insurance issues need to be optimized in order to provide high quality sensors.

Standardization of vocabulary and methods of a bit more established techniques need to get started so that it will be easier for new stakeholders to judge the function and reliability of these systems. General development in various areas of science and technology combined with the use of novel construction materials and structural systems will certainly bring both needs for further developments new SHM technologies, and knowledge and resources for their successful developments. The world is full of deteriorated, malfunctioning structures and as our infrastructures are aging there is an accelerating need for emerging technologies from that point of view too. Thus, we are confident that the future for the new and emerging technologies is bright.

ACKNOWLEDGMENTS

We would like to thank all the participants in Punggol EC26 project and also thank them for their kind collaboration and making available the data presented in this book: Er. Lau Joo Ming, Er. Fong Chor Cheong, and their crews in the Housing and Development Board (HDB), Singapore—the pioneer of large-scale implementation in long-term structural health monitoring for residential buildings; Nicoletta Casanova, Daniele Inaudi, and their teams in SMARTEC SA, Switzerland, and Rocrest Ltd., Canada; Mr K.P. Kwan, Mr Jeffery Low, and Sofotec Singapore Pte Ltd, Singapore.

The Strecker bridge project has been realized with great help, and kind collaboration of several professionals and companies. We would like to thank Steve Hancock and Turner Construction Company; Ryan Woodward and Ted Zoli, HNTB Corporation; Dong Lee and A.G. Construction Corporation; Steven Mancini and Timothy R. Wintermute, Vollers Excavating SMARTEC SA, Switzerland; Micron Optics, Inc., Atlanta, GA. In addition the following personnel, departments, and offices from Princeton University supported and helped realization of the project: Geoffrey Gettelfinger, James P. Wallace, Miles Hersey, Paul Prucnal, Yanhua Deng, Mable Fok; faculty and staff of Department of Civil and Environmental Engineering and our students: Maryanne Wachter, Jessica Hsu, George

Lederman, Jeremy Chen, Kenneth Liew, Chienchuan Chen, Allison Halpern, David Hubbell, Morgan Neal, Daniel Reynolds, Konstantinos Bakis, and Daniel Schiffner.

Concerning the Stonecutters Bridge, Hong Kong, great thanks for the input received by K. Y. Wong at Bridges & Structures Division of Highways Department, the Government of Hong Kong S.A.R. P.R. China. For the Messina Straight Bridge, thanks for kind help and fruitful discussions provided by the owner, Stretto di Messina represented by Eng. Enzo Vullo, the Contractor Eurolink SpA and COWI's designer group with help from Simon de Neumann from Flint & Neil Ltd.

Great thanks go also for: Paolo Papeschi and Paolo Farina from IDS Ingegneria Dei Sistemi S.p.A (Italy) and Kevin Banks from IDS Ingegneria Dei Sistemi (UK) Ltd; Leif Norman (Sweden) and Jon Watson (UK) from Physical Acoustics Ltd and Birit Buhr Jensen from COWI A/S, Denmark for their valuable help in the areas where author's did lack in experience.

We would also like to humbly thank all involved people in projects, measuring campaigns, and testing that are not mentioned above. Working with new and emerging technologies has been exciting, demanding, and enlightening and we are extremely satisfied over the goals we have reached so far and would like to continue together against new challenges.

REFERENCES

- Ackermann F. *Digital image correlation: performance and potential application in photogrammetry*. The Photogrammetric Record 1984;11:429–439.
- Aktan AE, Catbas FN, Grimmelsman KA, Tsikos CJ. *Issues in infrastructure health monitoring for management*. Journal of Engineering Mechanics 2000; V126:711–724.
- Aktan AE, Catbas FN, Grimmelsman KA, Pervizpour M. *Development of a Model Health Monitoring Guide for Major Bridges*. Drexel Intelligent Infrastructure and Transportation Safety Institute, USA; 2001.
- ASTM E1106-07. Standard test method for primary calibration of acoustic emission sensors. 2007. DOI: 10.1520/E1106-07.
- ASTM E1781-08. Standard practice for secondary calibration of acoustic emission sensors. 2008. DOI: 10.1520/E1781-08.
- ASTM E1318-09. Standard specification for highway weigh-in-motion (WIM) systems with user requirements and test methods. DOI: 10.1520/E1318-09.
- Aufleger M. Innovative dam monitoring systems. Dam Safety: Proceedings of the International Symposium on New Trends; 1998. ISBN 9054109742.
- Bao XY, Ravet F, Zou LF. Non linear strain response of the concrete column to detect the debonding and cracks using distributed Brillouin sensor. Proceedings of Second International Conference Structure Health Monitor Intelligent Infrastructure; 2005;1:235–242.
- Belli R, Glisic B, Inaudi D, Gebreselassie B. Smart textiles for SHM of geostuctures and buildings, SHMII-4. The 4th International Conference on Structural Health Monitoring of Intelligent Infrastructure. Paper on conference CD Zurich, Switzerland; 2009.
- Benning W, Lange J, Schwermann R, Effkemann C, Görtz S. Monitoring crack origin and evolution at concrete elements using photogrammetry. In Proceedings of XXth Congress of ISPRS ISPRS XXXV Part B;2004. p. 678–683.
- Bhanushali V, Andersen JE, Christensen SC. Structural Health Monitoring System, Naini Bridge, India. Proceedings of IABSE Report Vol. 91;2006.

- Billington DP. *The Tower and the Bridge*. New York: Basic Books, Inc., Publishers; 1983.
- British Standard. BS EN 13477-2:2001 Non-Destructive Testing. Acoustic Emission. Equipment Characterization. Verification of Operating Characteristic. British Standards Institution; 2001.
- British Standard. *BS EN 13554:2002 Non-Destructive Testing—Acoustic Emission—General Principles*. British Standards Institution; 2002.
- British Standard. *Non-Destructive Testing. Terminology. Terms Used in Acoustic Emission Testing*. British Standards Institution; 2009.
- Brownjohn JMW, Rizos C, Tan GH, Pan TC. Real-time long-term monitoring of static and dynamic displacements of an office tower, combining RTK GPS and accelerometer data. 1st FIG International Symposium on Engineering Surveys for Construction Works and Structural Engineering, Nottingham. UK. <http://www.fig.net/nottingham/>. 2004.
- Brownjohn JMW. *Structural Health Monitoring of civil infrastructure*. Philosophical Transactions of the Royal Society A 2007;365:589–622.
- BSI. *Mechanical Vibration and Shock-Performance Parameters for Condition Monitoring of Structures*. BS ISO 15587:2004; British Standards Organisation; 2004.
- Buderer R. *The Invention that Changed the World*. New York: Simon & Schuster; 1996.
- Buhr Jensen B, Schultz K, Grønvold F. *Management System for Corrosion Monitoring in the Copenhagen Metro*. CORROSION 2010, March 14–18, 2010; San Antonio, TX: NACE International; 2010.
- Clark P, Boriniski J, Gunther M, Poland S, Wigent D, Watkins S. *Modern fibre optic sensors*. Smart Materials Bulletin 2001;6:8–11.
- Cole PT, Watson J. *Acoustic emission for corrosion detection*. Advanced Materials Research 2006;13–14:231.
- Colombo S, Main IG, Forde MC, Halliday J. AE on bridges: experiments on concrete beams. On Proceedings of the 25th European Conference on AE Testing, Prague, Czech Republic; 2002.
- Conyers LB. *Ground-Penetrating Radar for Archaeology*. Walnut Creek, CA: AltaMira; 2004.
- Conyers LB, Lucius JE. *Velocity analysis in archaeological ground-penetrating radar studies*. Archaeological Prospection 1996;3(1):25–38.
- Conyers LB, Goodman D. *Ground-Penetrating Radar: An Introduction for Archaeologists*. Walnut Creek, CA: AltaMira; 1997.
- Conyers LB. Ground Penetrating radar Published Online: 15 Jan; 2002.
- Dadpay C, Sivakumar NR, Mrad N. Strain distribution and sensitivity in fiber Bragg grating sensors. Proceedings of SPIE 7099, 70992E; 2008.
- Day GS, Schoemaker PJH, Gunther RE. *Wharton on Managing Emerging Technologies*. New Jersey: Wiley; 2000. ISBN: 978-0-471-68939-3.
- Del Grosso A, Torre A, Inaudi D. Monitoring system for a cable-stayed bridge using static and dynamic fiber optic sensors. 2nd International Conference on Structural Health Monitoring of Intelligent Infrastructure (SHMII 2), Shenzhen, China; 2005. p. 415–420.
- Dunn RC, Ross RA, Davis GD. Corrosion Monitoring of Steel Reinforced Concrete Structures Using Embedded Instrumentation. NACE International: CORROSION 2010, Paper No. 10173; 2010.
- El-Rabbany A. *Introduction to GPS: The Global Positioning System*. Norwood: Artech House Inc.; 2002. ISBN: 1-58053-183-0.
- EMPA, CityU, COWI, LTU, NFBC, OU, UMINHO, USTUTT, USAC, WUT. *Evaluation of Monitoring Instrumentation and Techniques. Technical report, the sustainable bridges project co-funded by the European Commission within the Sixth Framework Programme*; 2004.
- Enckell M. Structural Health Monitoring using modern sensor technology—long-term monitoring of the New Årsta railway bridge [Licentiate thesis]. Royal Institute of Technology, KTH; 2006.

- Enckell M. Structural Health Monitoring of bridges in Sweden. On proceedings CD of the 3rd International Conference on Structural Health Monitoring of Intelligent Infrastructure—SHMII-3 Paper No. 117; 2007.
- Enckell M, Glisic B, Myrvoll F, Bergstrand B. *Evaluation of a large-scale bridge strain, temperature and crack monitoring with distributed fibre optic sensors*. Journal of Civil Structural Health Monitoring 2011;1:37–46.
- Enckell M, Karoumi R, Lanaro F. Monitoring of the New Årsta Railway Bridge using traditional and fibre optic sensors. Proceedings of the SPIE, Smart Structures and Materials, NDE for Health Monitoring & Diagnostics; Vol. 5057; 2003a; p. 279–288.
- Enckell M, Karoumi R, Wiberg J. Structural Health Monitoring for an optimized pre-stressed concrete bridge. Proceedings of the ISHMII-1 V 2; 2003b; p. 993–996.
- Enckell M, Larsson H. Monitoring the behaviour of the Traneberg Bridge during retrofitting. Proceedings of the ISHMII-2, V P2; 2005; p. 1631–1635.
- Enckell M, Wiberg J. Monitoring of the New Årsta Railway Bridge. Instrumentation and preliminary results from the construction phase. Technical Report Royal Institute of Technology, KTH; 2005.
- Farrar C, Darling T, Migliori A, Baker W. *Microwave interferometers for non-contact vibration measurements on large structures*. Mechanical Systems and Signal Processing 1999;13:241–254.
- Farrar C, Worden K. *An introduction to Structural Health Monitoring*. Philosophical Transactions of the Royal Society A 2007;365:303–315.
- Feng Q. Novel methods for 3-D semi-automatic mapping of fracture geometry at exposed rock faces [Doctoral thesis]. Royal Institute of Technology (KTH). 2001; ISBN 91-7283-113-8.
- Fowler TJ, Yopez LO, Barnes CA. Acoustic emission monitoring of reinforced and prestressed concrete structures. Structural Materials Technology III: An NDT Conference; 1998; Vol. 3400, p. 281–298.
- Gentile C. *Deflection measurement on vibrating stay cables by non-contact microwave interferometer*. NDT & E International 2010;43:231–240.
- Gentile C, Bernardini G. *An interferometric radar for non-contact measurement of deflections on civil engineering structures: laboratory and full-scale tests*. Structure & Infrastructure Engineering 2009; DOI: 10.1080/15732470903068557.
- Glisic B. Fibre optic sensors and behaviour in concrete at early age [Dissertation]. Ecole Polytechnique Federale de Lausanne (EPFL); 2000.
- Glisic B. Streicker Bridge: an on-site SHM laboratory at Princeton University campus. First Middle East Conference on Smart Monitoring, Assessment and Rehabilitation of Civil Structures. Paper No. 306, Dubai, UAE; 2011.
- Glisic B, Adriaenssens S. Streicker Bridge: initial evaluation of life-cycle cost benefits of various Structural Health Monitoring approaches, IABMAS2010. The 5th International Conference on Bridge Maintenance, Safety and Management, Paper No. 0396; Philadelphia, PA; 2010.
- Glisic B, Enckell M, Myrvoll F, Bergstrand B. Distributed sensors for damage detection and localization. On Proceedings CD of the 4th International Conference on Structural Health Monitoring on Intelligent Infrastructure-SHMII-4 Paper No. 393; 2009.
- Glisic B, Inaudi D. *Fibre Optic Methods for Structural Health Monitoring*. Chichester, England: Wiley; 2007.
- Glisic B, Inaudi D, Casanova N. SHM process as perceived through 350 projects. SPIE Smart Structures and Materials/NDE Symposium, Paper No. 7648-26; 2010; San Diego, CA, USA; 2010a.
- Glisic B, Inaudi D, Hoong KC, Lau JM. Monitoring of building columns during construction. 5th Asia Pacific Structural Engineering and Construction Conference; Johor Bahru, Malaysia; 2003; p. 593–606;

- Glisic B, Inaudi D, Lau JM, Fong CC. Real strain evolution in concrete columns, monitoring results and CEB-FIP Model Code 1990. 3rd fib International Congress; 2010; Washington DC, USA; 2010b.
- Glisic B, Posenato D, Persson F, Myrvoll F, Enckell M, Inaudi D. Integrity monitoring of old steel bridge using fiber optic distributed sensors based on Brillouin scattering. On Proceedings CD of the 3rd International Conference on Structural Health Monitoring of Intelligent Infrastructure—SHMII-3 Paper No. 112; 2007.
- Golaski L. Acoustic emission monitoring of reinforced concrete elements—an overview. Proceedings of the First Workshop of COST 534 of NDT Assessment and New Systems in Prestressed Concrete Structures; 2004. p. 7–23.
- Golaski L, Gebksi P, Ono K. *Diagnostics of reinforced concrete bridges by acoustic emission*. Journal of Acoustic Emission 2002;20:83–98.
- Hamstad MA, McColskey JD. *Detectability of slow crack growth in bridge steels by acoustic emission*. Materials Evaluation 1999;57:1165–1174.
- Highways Agency (UK). Agency Highway structures: Inspection and maintenance. Inspection. Advice notes on the non-destructive testing of highway structures. DMRB Volume 3 Section 1 DMRB Vol. 3 Section 1 Part 7 (BA 86/06); 2006. ISBN 9780115527784.
- Holford KM, Cole PT, Carter DC, Davies AW. The non-destructive testing of steel girder bridges by acoustic emission. 14th World Conference on Non-Destructive Testing; 1996; 4, p. 2509–2512. ISBN 8120411269.
- Imai M, Nakano R, Kono T, Ichinomiya T, Miura S, Mure M. Crack detection application for fiber reinforced concrete using BOCDA-based optical fiber strain sensor. On Proceedings of the ISHMII-4, Paper No. 509; 2009.
- Inaudi D. Fiber optic sensor network for the monitoring of civil structures [Ph.D. thesis]. No. 1612, EPFL, Lausanne; 1997.
- Inaudi D, Glisic B. Application of distributed fiber optic sensory for SHM. Proceedings of the 2nd International Conference on Structural Health Monitoring of Intelligent Infrastructure (SHMII-2'). Vol. 1; 2005. p. 163–169.
- Inaudi D, Glisic B. Fiber Optic sensing for innovative oil & gas production and transport systems. 18th International Conference on Optical Fiber Sensors, Paper No. 14; 2006.
- Inaudi D, Glisic B. Distributed Fiber Optic Sensing for long range monitoring of pipelines. On proceedings CD of the 3rd International Conference on Structural Health Monitoring of Intelligent Infrastructure—SHMII-3 Paper No. 56; 2007.
- ISO 12713. Non-destructive testing—Acoustic emission inspection—Primary calibration of transducers; ANSI; 1998.
- ISO 12714. Non-destructive testing—Acoustic emission inspection—Secondary calibration of acoustic emission sensors; 1999.
- Jaffrey D. *Acoustic Emission Monitoring*. Metropolitan Forum 1982;5:154–157.
- Karashima T, Horiguchi T, Tateda M. *Distributed temperature sensing using stimulated Brillouin scattering in optical silica fibers*. Optics Letters 1990;15(18):1038–1040.
- Kikuchi K, Naito T, Okoshi T. *Measurement of Raman scattering in single-mode optical fiber by optical time-domain reflectometry*. IEEE Journal of Quantum Electronics 1988;24(10):1973–1975.
- Krohn DA. *Fiber Optic Sensors: Fundamentals and Applications*. Triangle Park, NC: Instrument Society of America; 2000.
- Kulicki JM, Prucz Z, Sorgenfrei DF, Mertz DR, Young WT. *Guidelines for Evaluating Corrosion Effects in Existing Steel Bridges*. National Cooperative Highway Research Program Report 333 (Transportation Research Board, December 1990); 1990.

- Kunzler M, Udd E, Taylor T, Kunzler W. Traffic Monitoring using fiber optic grating sensors on the I-84 freeway & future uses in WIM. *Proceedings of SPIE* 2003;5278:122.
- Kwong AKL, Kwok H, Wong A. Use of 3D laser scanner for rock fractures mapping. *Proceedings of TS 8C—Terrestrial Laser Scanning II, Strategic Integration of Surveying Services, FIG Working Week; China: Hong Kong SAR; 2007.*
- Lanticq V, et al. *Soil-embedded optical fiber sensing cable interrogated by Brillouin optical time-domain reflectometry (B-OTDR) and optical frequency-domain reflectometry (OFDR) for embedded cavity detection and sinkhole warning system.* *Measurement Science and Technology* 2009;20:034018.
- Leming SK, Stalford HL. Bridge Weigh-in-Motion System using superposition of Dynamic Truck/-Static Bridge Interaction. *Proceedings of the American Control Conference; Denver, Colorado; 2003.*
- Leondes CT. *Mems/Nems Handbook Techniques and Applications Design Methods, (2) Fabrication Techniques, (3) Manufacturing Methods, (4) Sensors and Actuators, (5) Medical Applications and MOEMS; Springer V 1. ISBN: 978-0-387-24520-1; 2006.*
- Liehr S, Lenke P, Wendt M, Krebber K, Seeger M, Thiele E, Gebreselassie B, Munich C. *Polymer optical fiber sensors for distributed strain measurement and application in Structural Health Monitoring.* *IEEE Sensors Journal* 2009;9(11):1330–1338.
- Liew K. *Bridging art and engineering: structural art and performance analysis of the Streicker Bridge [Senior thesis]. Princeton University; 2010.*
- Lindley TC, Palmer IG, Richards CE. *Acoustic emission monitoring of fatigue crack growth.* *Materials Science and Engineering* 1978;32:1–15.
- Lynch JP. *An overview of wireless Structural Health Monitoring for civil structures.* *Philosophical Transactions of the Royal Society A* 2007;365:345–372.
- Matsuyama K, Fujiwara T, Ishibashi A, Ohtsu M. *Field application of acoustic emission for the diagnosis of structural deterioration of concrete.* *Journal of Acoustic Emission* 1993;11: S65–S73.
- Mayer L, Yanev B, Olson LD, Smyth A. Monitoring of the Manhattan Bridge for Vertical and Torsional Performance with GPS and Interferometric Radar Systems. *Transportation Research Board Annual Meeting* 2010. Paper No. 10-3183; 2010.
- McCann DM, Forde MC. *Review of NDT methods in the assessment of concrete and masonry structures.* *NDT & E International* 2001;34:71–84.
- Measures MR. *Structural Monitoring with Fibre Optic Technology.* San Diego: Academic Press; 2001.
- Messervey TB, Zangani D, Withiam JL. Smart textiles and their application in bridge engineering, IABMAS2010—The Fifth International Conference on Bridge Maintenance; Philadelphia, PA: Safety and Management; 2010. p. 2135–2139.
- Mikhail EM, Bethel J, McGlone JC. *Introduction to Modern Photogrammetry.* Wiley; 2001. ISBN: 978-0-471-30924-6.
- Miller RK, McIntire P. *Nondestructive Testing Handbook. 2nd edn. Vol. 5: Acoustic Emission Testing.* American Society for Nondestructive Testing; 1987.
- Miyata T. *Historical view of long-span bridge aerodynamics.* *Journal of Wind Engineering and Industrial Aerodynamics* 2003;91:1393–1410.
- Miyata T, Yamada H, Katsuchi H, Kitagawa M. *Full-scale measurement of Akashi-Kaikyo Bridge during typhoon.* *Journal of Wind Engineering and Industrial Aerodynamics* 2002;90:1517–1527.
- Moses F. (1979), *Weigh-In-Motion system using instrumented bridges.* *ASCE Transportation Engineering Journal* 1979;105:233–249.
- Mufti A. *Guidelines for Structural Health Monitoring. ISIS Design Manual No. 2. ISIS; 2001. Canada. www.isiscanada.com.*

- Myrvoll F, Bergstrand B, Glisic B, Enckell M. Extended operational time for an old bridge in Sweden using instrumented integrity monitoring. On Proceedings of the Fifth Symposium on Strait Crossings; 2009; p. 397–401.
- Nikles M, Thévenaz L, Robert PA. Simple distributed temperature sensor based on Brillouin gain spectrum analysis. Proceedings of 10th International Conference on Optical Fiber Sensors OFS 10, SPIE Vol. 2360; 1994; p. 138–141.
- Nikles M, Thévenaz L, Robert PA. *Brillouin gain spectrum characterization in single-mode optical fibers*. Journal of Lightwave Technology 1997;15(10):1842–1851.
- Nikles M, Vogel B, Briffod F, Sauser F, Luebbecke S, Bals A, Pfeiffer T. Leakage detection using fiber optics distributed temperature monitoring. 11th SPIE's Annual International Symposium on Smart Structures and Materials; March 14–18; San Diego, USA; Vol. 5384; 2004. p. 18–25.
- Nöther N, Wosniok A, Krebber K, Thiele E. A distributed fiber-optic sensing system for monitoring geotechnical structures. On Proceedings of the ISHMII-4, Paper No. 509; 2009.
- Obrien E, Žnidaric A, Ojio T. Bridge Weight-In-Motion-Latest developments and applications world wide. Proceedings of the International Conference on Heavy Vehicles; 2008. p. 25–38.
- Ohtsu M. *Acoustic Emission Characteristics in Concrete and Diagnostic Applications*. Journal of Acoustic Emission 1987;6:99–108.
- Ohtsu M, Yuyama S. Recommended practice for in situ monitoring of concrete structures by acoustic emission. Proceedings of the 15th International Acoustic Emission Symposium; 2000; p. 263–268.
- Physical Acoustics Corporation. *Acoustic Emission for bridge inspection*. Report No. FHWA-RD-94-prepared for FHWA and U.S. Department of Transportation; 1995.
- Pollock AA, Smith B. *Acoustic Emission Monitoring of a military bridge*. Nondestructive Test 1972;5(6):164–186.
- Posey R. Jr., Johnson GA, Vohra ST. *Strain sensing based on coherent Rayleigh scattering in an optical fibre*. Electronics Letters 2000;36(20):1688–1689.
- Pullin R, Holford KM, Evans SL. Confidence of detection of fracture signals using acoustic emission. 5th International Conference on Advances in Experimental Mechanics, Manchester, 4th–6th Sept; 2007.
- Quilligan M. Bridge weigh-in-motion—development of a 2-D multi-vehicle algorithm [Licentiate thesis]. Royal Institute of Technology, KTH; 2003.
- Ramadan S, Gaillet L, Tessier C, Idrissi H. *Detection of stress corrosion cracking of high-strength steel used in prestressed concrete structures by acoustic emission technique*. Applied Surface Science 2008;254(8):2255–2261.
- Ravet F, Briffod F, Glisic B, Nikles M, Inaudi D. *Submillimeter crack detection with Brillouin-based fiber-optic sensors*. IEEE Sensors Journal 2009;9(11):1391–1396.
- Rizos C. Making sense of the GPS techniques. In: Bossler J, Jensen J, McMaster R, Rizos C, editors. *Manual of Geospatial Science and Technology*. London, UK: Taylor & Francis; 2002. p. 146–161.
- Roberge PR. Corrosion Basics: An Introduction. 2nd edn; 2006. ISBN: 1-57590-198-0.
- Roberts GW, Brown C, Ogundipe O. Deformation Monitoring of Structures using GNSS. On Proceedings of the FIG Congress; 2010.
- Ross RM, Matthews SL. *In-service structural monitoring—a state of the art review*. The Structural Engineer 1995;73:23–31.
- Rödelsperger S, Läufer G, Gerstenecker C, Becker M. *Monitoring of displacements with ground-based Microwave Interferometry: IBIS-S and IBIS-L*. Journal of Applied Geodesy 2010;4:41–54.
- Salawu OS, Williams C. *Review of full-scale dynamic testing of bridge structures*. Engineering Structures 1995;17:113–121.

- Sarri A, Manacorda G, Miniati M. Novel GPR system for high-resolution inspection of walls and structures. Proceedings of the Ninth International Conference on Ground Penetrating Radar; Koppenjan S, Lee H, editors, Vol. 4758; 2002. p. 498–502.
- Shi B, Sui HB, Liu D, Zhang W, Suo WB. *Applications of distributed fiber optic sensing technologies in geotechnical engineering monitoring*. Proceedings of the ISHMII-2 2005;1:299–306.
- Simi A, Manacorda G, Miniati M, Bracciali S, Buonaccorsi A. Underground asset mapping with dual frequency dual-polarized GPR massive array. 13th International Conference on Ground Penetrating Radar (GPR); 2010. p. 1–5.
- Simmons G, Strangway DW, Bannister L, Baker R, Cubley D, La Torraca G, Watts R. In: Geophysics L, Kopal Z, Strangway DW, editors. *The Surface Electrical Properties Experiment*. Dordrecht: D. Reidel; 1972. p. 258–271.
- Sinclair ACE, Connors DC, Formby CL. *Acoustic emission analysis during fatigue crack growth in steel*. Materials Science and Eng 1977;28:263–273.
- Srinivasan S, Taylor SE, Basheer PAM, Smith BJ, Sun T, Grattan KTV. Fibre optic relative humidity sensors for long-term monitoring of concrete structures. In Proceedings 4th International Conference on Structural Health Monitoring on Intelligent Infrastructure (SHMII-4); 2009.
- Stern W. *Versuch einer elektrodynamischen Dickenmessung von Gletschereis*. Gerlands Beitrage zur Geophysik 1929;23:292–333.
- Stern W. *Über Grundlagen, Methodik und bisherige Ergebnisse elektrodynamischer Dickenmessung von Gletschereis*. Zeitschrift für Gletscherkunde 1930;15:24–42.
- Sundquist H, James G. Monitoring of shear cracks and the assessment of strengthening on two newly-built light-rail bridges in Stockholm. On proceedings of the Second International Conference on Bridge Maintenance Safety and Management, IABMAS, Kyoto, Japan; 2004.
- Tarchi D, Casagli N, Fanti R, Leva DD, Luzi G, Pasuto A, Pieraccini M, Silvano S. *Landslide monitoring by using ground-based SAR interferometry: an example of application to the Tessina landslide in Italy*. Engineering Geology 2003;68:15–30.
- Thevenaz L, Nikles M, Fellay A. Truly distributed strain and temperature sensing using embedded optical fibers. SPIE Conference on Smart Structures and Materials, Vol. 3330; San Diego, USA; 1998. p. 301–314.
- Udd E. *Fiber Optic Sensors*. New York: Wiley; 1991.
- Udd E. *Fiber Optic Smart Structures*. New York: Wiley; 1995.
- Udd E. *Fiber Optic Sensors: An Introduction for Engineers and Scientists*. New York: Wiley; 2006.
- Veldhuijzen Van Zanten R, editor. *Geotextiles and Geomembranes in Civil Engineering*. Rotterdam, The Netherlands: Balkema; 1986.
- Vurpillot S. Analyse automatisée des systèmes de mesure de déformation pour auscultation des structures. Dissertation, Ecole Polytechnique Federale de Lausanne (EPFL); 1999.
- Wait PC, Hartog AH. *Spontaneous Brillouin-based distributed temperature sensor utilizing a fiber Bragg grating notch filter for the separation of the Brillouin signal*. IEEE Photonics Technology Letters 2001;13(5):508–510.
- Washer GA. *Developments for the Non-destructive Evaluation of Highway Bridges in the USA*. Non-Destructive Testing & Evaluation, Elsevier Science Ltd 1998;31(4):245–249.
- Watson JR, Cole PT, Holford KM, Davies AW. Damage assessment using acoustic emission. Proceeding of 4th International Conference on Damage Assessment of Structures, Key Engineering Materials; 2001. p. 204–205.
- Watson JR, Cole PT, Kennedy-Reid I, Halliday J. Condition assessment of concrete half joints. Proceedings of the First International Conference on Bridge Maintenance, Safety and Management, IABMAS; 2002.

- Wiberg J. Bridge monitoring to allow for reliable dynamic FE modelling: a case study of the New Årsta Railway Bridge [Licentiate thesis]. Royal Institute of Technology, KTH; 2006.
- Wiberg J, Enckell M. Monitoring of the New Årsta Railway Bridge. Presentation of measured data and report on the monitoring system over the period 2003–2007. Technical Report Royal Institute of Technology, KTH; 2008.
- Yang L, *A Techniques for Corrosion Monitoring*. Cambridge: Woodhead Publishing Ltd; 2008.
- Yuyama S, Okamoto T, Shigeishi M, Ohtsu M, Kishi T. A Proposed Standard for Evaluating Structural Integrity of Reinforced Concrete Beams by Acoustic Emission, Acoustic Emission: Standards and Technology Update, ASTM STP 1353. Vahaviolos JJ, editor. American Society for Testing and Materials; 1999; p. 25–40.
- Zhang H, Wu Z. *Performance evaluation of BOTDR-based Distributed Fiber Optic Sensors for crack monitoring*. Structural Health Monitoring 2008;7:143–156.

2

APPLICATIONS OF GIS IN ENGINEERING MEASUREMENTS

GARY S. SPRING

- 2.1 Introduction
- 2.2 Background
 - 2.2.1 Measurement-based GIS
- 2.3 Basic principles of GIS
 - 2.3.1 GIS data
 - 2.3.2 Spatial modeling
- 2.4 Measurement-based GIS applications
- 2.5 Implementation issues
 - 2.5.1 Data
 - 2.5.2 Technology
 - 2.5.3 Organizations and people
- 2.6 Conclusion
 - 2.6.1 Metadata
 - 2.6.2 Data visualization
 - 2.6.3 Enterprise GIS/asset management
- References

2.1 INTRODUCTION

The concepts that drive geographic information systems (GIS) have become commonplace in today's everyday activities—witness the use of guidance systems in our cars, our use of online mapping programs (such as Google maps, Yahoo maps, etc.) to explore destinations. All of these are based upon basic GIS functions, namely, linking data to maps. The use of geographic information systems has consequently become widespread in the past 20 years or so. All 50 states have institutionalized their GIS activities and 45 of them provide clearinghouses for their geospatial data (AASHTO, 2008). For a full list of

GIS acronyms and jargons visit the National Center for Geographic Information and Analysis web site (Padmanabhan et al., 1992), and download report number 92-13.

GIS offers a data management and modeling platform capable of integrating a vast array of data from various sources, captured at different resolutions, and on seemingly unrelated themes. The objectives for this chapter are to define and describe the basic elements of GIS, explore its application to engineering measurements (in the context of surveying and geodesy), provide some example applications for the technology, and present issues involved in implementing these systems. The discussion begins with some background on GIS providing definitions and some discussion regarding the “why” of using them.

2.2 BACKGROUND

The GIS is a computerized database management system that provides graphic access (capture, storage retrieval, analysis, and display) to spatial data. GIS software provides a map display that allows thematic mapping of data and its graphic output overlain onto a map image. Figure 2.1 depicts the generic framework used by virtually all-current GIS software packages.

The key element that distinguishes GIS from other data systems is the manner in which geographic data are stored and accessed. GIS packages used for engineering measurements applications store geographic data using topological data structures (objects’ locations relative to other objects are explicitly stored and therefore are accessible) that allow analyses to be performed that are impossible using traditional data structures. The addition of this spatial dimension to the database system is, of course, the source of power of GIS. Without this dimension, the GIS is merely a database management engine. Linked with the spatial dimension, its database management features enable GIS to capture spatial and topological relationships among geo-referenced entities even when these relationships are not predefined. Standard GIS functions include thematic mapping, statistics, charting, matrix manipulation, decision support systems, modeling algorithms, and simultaneous access to several databases.

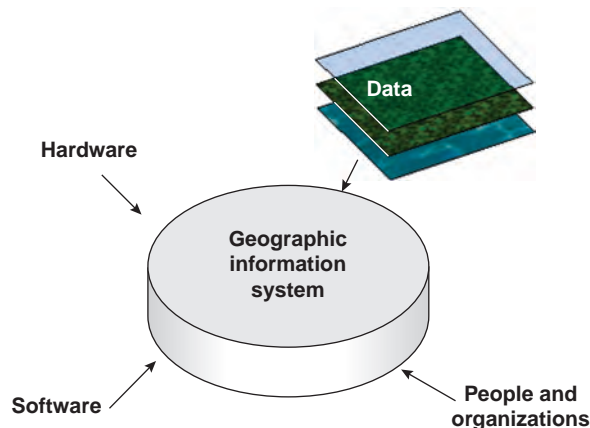


FIGURE 2.1 Generic framework GIS.

2.2.1 Measurement-Based GIS

In 1999, Goodchild proposed a new and expanded focus area for GIS software, namely, one whose geometry is based upon measurements rather than coordinates. The traditional approach to storing geographic information is with coordinate data. Mapping has not, normally, been considered a measurement science. Instead, information that links GIS layers to original measurements is commonly removed, making conventional error analysis, within the GIS, impossible. The data-based nature of GIS, however, allows this information to be included.

Sadiq and Duckham (2009) examined a technique for storing and retrieving spatially varying data quality information in a relational spatial database. Rather than storing global data quality statements, the system enables data quality information to be referenced to a spatial framework, individual spatial objects, or even parts of spatial objects. Their relational model allows flexible storage of spatially varying data quality information, and seamless querying irrespective of the underlying storage model. They found that this system is practical and efficient for a wide range of queries, and outlined the performance trade-offs associated with different data quality storage models. Kuhn (2009) proposed a taxonomy of measurements that allows for data integration and the application of computational techniques, such as least squares, to measurements. Positions and their uncertainties, for example, could more reliably be determined if original measurement data were maintained in a system.

2.3 BASIC PRINCIPLES OF GIS

2.3.1 GIS Data

Data and the modeling of those data are crucial elements of a successful system. The GIS derives information from raw data—one of its primary strengths. Although data and information are often used interchangeably, they can be quite different. Data consist of facts or numbers representing facts whereas information derives from data and gives meaning to those facts. This section describes the nature, types and sources of data, and the modeling techniques and data models which are used to convert the raw data to information.

2.3.1.1 The Nature of Data There are two types of data in general—spatial data and attribute data. The former describes the physical geography represented in a database whereas the latter, attribute data, linked to the spatial data, describe the “attributes” of the spatial objects. For example, a point could represent a tree location and its attributes might be species, diameter, or height. If a line feature represents a road segment, attributes might include pavement type, number of lanes, or speed limit. Apropos to measurement-based GIS, attributes may also include survey measurement information. Area attributes might include soil type, vegetation cover, or land use. An attribute is the generic descriptor for a feature, while each feature has a specific value. In the point example above, likely values for the species attribute would be Pine, Fir, and Aspen. Attributes may be described as the questions that would be asked and values as the answers. Attributes to be collected would be determined in the software planning process, while their specific values would be entered in the field. Again, these may include survey measurements and/or their history.

The two most common ways of representing spatial data are the raster and vector formats. A familiar analogy for these formats is bit-mapped photos (such as jpeg) corresponding to the raster format and vector-based ones (such as tiff) representing the vector format. Each of these approaches offers advantages and disadvantages based on the application for which the data are to be used. Data structure, that is choice between raster and vector, is in general one of the first major decisions to be made in establishing a GIS. With regard to measurement-based GIS, this decision is fairly straightforward. Field measurements are typically concerned with measuring and locating features and artifacts rather than processing images, for example. These are all well-defined lines and shapes lending themselves better to the vector data structure. This notwithstanding, there are situations in which the raster format may apply and, so, the next few paragraphs provide a description of both structures including some discussion of their relative advantages and disadvantages.

The Raster Data Structure Raster data consist of rows of uniform cells coded according to data values. For example, a landscape scene would be gridded, and each cell in the grid would be given a single landscape identity, usually a code number that refers to a specific attribute measure (e.g., a particular land use or type of land cover). The number might also be an actual measurement value, such as an amount of rainfall. These cells are akin to pixels on a computer monitor. Indeed, carrying the analogy further, pixels have values associated with them as well—color, hue, and shades of gray—and the degree of resolution provided relates to the size of the pixel. Similarly, the degree of approximation in a landscape relates to the size of the raster cell. A grid comprised of small cells will follow the true location of a boundary line more closely, for example. However, as with photo resolution, there is an overhead cost, namely, the size of the raster file containing the data will increase as cell size decreases. In general, raster provides a simple data model and fast processing speed at the expense of the excellent precision provided by the vector model with its higher data needs. The nature of raster data lends itself to natural resources applications whose data elements tend to be continuous in nature—such as soil type.

The Vector Data Structure Vector data consist of points, lines, and closed polygons or areas, much the same as the drawing elements in computer-aided drafting (CAD) programs. The lines are continuous and are not broken into a grid structure. In the vector model, information about points, lines, and polygons is encoded and stored as a collection of x - y coordinates. The location of a point feature, such as a manhole, is described by a single x - y coordinate. Linear features, such as roads, are stored as a string of point coordinates. Area features, such as census tracts or traffic analysis zones, are stored as a closed loop set of coordinates. The vector model represents discrete features such as buildings better than continuous features such as soil type. The vector format, in general, provides a more precise description of the location of map features, eliminates the redundancy afforded by the raster model and therefore reduces mass storage needs, and allows for network-based models. Vector models are, however, often computationally intensive—much more so than their raster counterparts.

Topology describes how graphical objects connect to one another by defining relative positions of points, lines, and areas. For example, topology allows queries about which street lines are adjacent to a census tract area, or what intersection node points form the end points of a street segment (link), for example. The latter information is essential for

routing applications. Only vector-based data include topological information which is in part why these data may have higher costs associated with them—the presence of topology implies the maintenance of additional databases whose purpose is to store information on the connectedness of points, lines, and areas in the database.

Scale and Accuracy Scale and accuracy are important data considerations, especially when using more than one data source, and is at the heart of any discussion involving measurement-based GIS.

Historically, a main indicator of accuracy and precision is the scale used to describe the data. Simply defined, scale is the relationship between distance on the map and distance on the ground. A map scale usually is given as a fraction or a ratio—1/10,000 or 1:10,000. These “representative fraction” scales mean that 1 unit of measurement on the map—1 inch or 1 centimeter—represents 10,000 of the same units on the ground. Small-scale maps have smaller representative fractions than large-scale maps; for example, 1 in 500,000 versus 1 in 24,000, respectively, that is, large is small. As scale size increases, the ability to depict detail also increases. For example, a map with scale 1 in 500,000 would necessarily represent only main roads and only by centerline whereas with a scale of 1 in 20,000, details such as ramps, collectors, and direction of travel could be shown as well. See Figure 2.2 for examples. On much larger scales, such as 1 in 600 (1 in = 50 ft), features such as pavement markings, actual lane designations, and specific design elements can be depicted. The latter scale is the level of resolution often used in roadway design work.

GIS employs a wide range of data sources reflecting the varied goals of the systems themselves. Since GIS may involve applications as varied as archeological analysis,

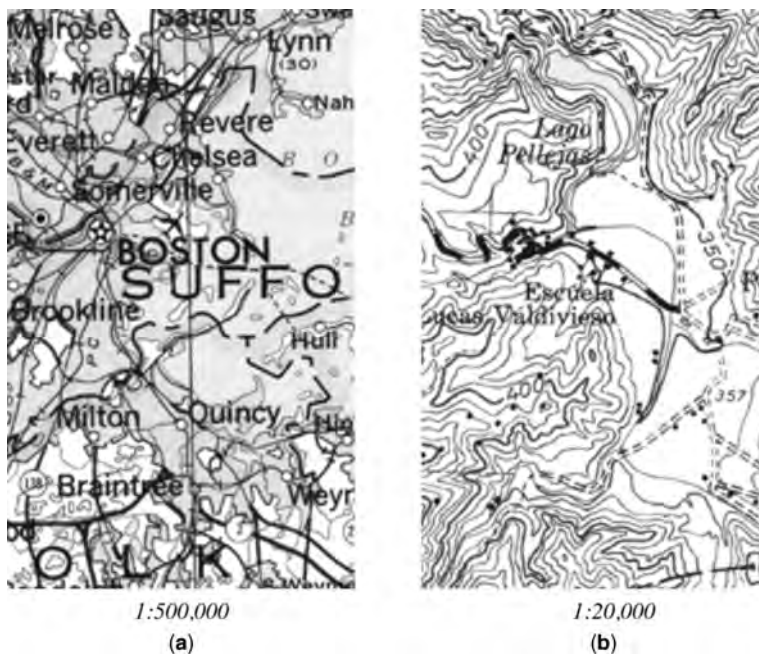


FIGURE 2.2 Effects of scale.

marketing research, and urban planning; the source materials can be difficult to inventory and classify comprehensively. The fact that many different scales may be encountered adds to the complexity of the problem. Even within a single GIS project, the range of materials employed can be daunting. If multiple datasets are contemplated, and they do not have common scales, the process of conflation (the fusion or marrying together of data) may be used in some cases to spatially integrate the datasets to create a new master coverage with the best spatial and attribute qualities. In situations such as this, the analyst should use extreme care in the use of these conflated data.

Uncertainty and Errors Goodchild (1998) described uncertainty as the Achilles' heel of GIS: "the dark secret once exposed, perhaps through the arguments of clever lawyers, will bring down the entire house of cards." Although this may seem extreme, he points out that GIS analysis results are often presented as more accurate than they really are.

In the fields of surveying and geodesy, error analysis based on known measurement error distributions is routine, are based on well-developed theory. For example, errors resulting from individual measurements that are combined arithmetically may be predicted based largely on knowledge of the error distributions of the measurements.

As stated earlier, GIS data normally separate original measurement information from geographic layers. Error modeling for these data has developed rapidly over the past 15 years (see, e.g., Zhang and Goodchild, 2002), but the exclusion of data about measurements constrains its development; error models are "retrofitted" to data, and key information such as error covariance structure often goes missing. In contrast, the fields of surveying and geodesy, as described above, where error analysis based on known measurement error distributions is routine, based on well-developed theory.

A stochastic view of the world, such as that taken by surveyors, assumes that it is not possible to know location on the Earth's surface perfectly (no measurement of a continuous quantity can ever be exact). Herein lies a major advantage of the measurement-based GIS for engineering measurement applications: they adopt a very different approach to GIS architecture, where positions are established not in some absolute Earth coordinate system, but by a hierarchy of relative measurements, anchored to the Earth only at well-defined points (Goodchild, 1999). Preserving the measurements, and the transformations that convert them to positions (and the inverses of such transformations), solves several longstanding GIS-related problems. For example, it is possible to do partial updates, for example, where positions of some features can be made more accurate without at the same time replacing the entire database. It is also possible to perform comprehensive analysis of the propagation of error, and thus of the errors in results of GIS analysis.

2.3.1.2 Geocoding The linkage between attribute data and spatial data as stated previously is the key advantage of using GIS. The process by which this is accomplished is called geocoding. There exist a great many tools to accomplish this process. Some involve matching a nonmap-based database, such as a list of names and addresses, with a map-based database, such as one containing zip code centroids. Using a common field, and the relational model (described in the next section), coordinate information is attached to the name and address.

Other common geocoding tools involve the use of distances along reference lines and offsets from these lines to place objects on a map, thus requiring some sort of linear referencing system. For example, going back to the nonmap-based names and addresses database, one could use a street centerline file which includes address ranges on its links

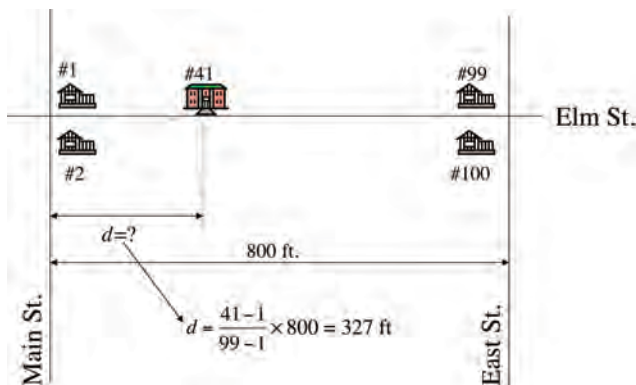


FIGURE 2.3 Geocoding using address ranges.

to geocode addresses by estimating the address locations based on the length of the segment and the address range assigned to the segment. For example, an address of 41 Elm Street would plot as shown in Figure 2.3.

2.3.1.3 Data Processing For the processing and management of data, the two paradigms most commonly used are the relational and object-oriented models. It should be noted that the database is the operation center of the GIS where much of the primary work is done. Graphics may or may not be necessary for an application but, almost without exception, the database is a key part of any analysis.

Relational database management systems (RDBMS) are well suited for ad hoc user queries—an important aspect of GIS analysis. The relational model uses tables of data arranged as columns (categories of data) and rows (each observation entry). Columns are called fields and rows are called records—as shown in Figure 2.4. Consider, for example, the table of courses shown in Figure 2.6. Note that each course’s attributes are read across, on a row. Queries may be made from these tables by specifying the table name, the fields of interest and conditions (e.g., course credit hours greater than 3). The rows that meet those conditions are returned for display and analysis. The power of RDBMS lies in its ability to link tables together via a “unique identifier” that is simply a field

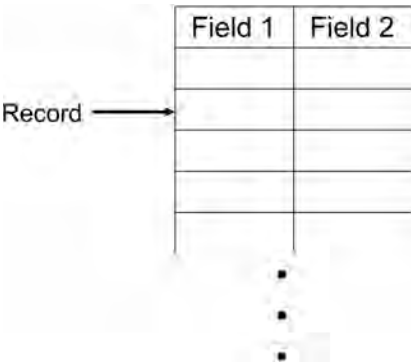


FIGURE 2.4 DBMS—The relational model.

Professor	Course	Hours	Course ID
Adams	CE341	3	44356
Burken	CE441	3	77877
Nanni	CE223	4	33645
Spring	CE210	3	65645
Zhang	CE446	4	48655

FIGURE 2.5 Table in the RDBMS.

common to both tables and whose values appear only once. Thus, using the current example, one could, in addition to querying about courses from Figure 2.5, query about which students take courses from which professor, number of students in each class who hail from Missouri, geographic distribution of students, and so on. This is done via the common field, student ID, which links the two data tables, as shown in Figure 2.6. Within GIS, this capability to link tables is used to link spatial objects with tables containing information relating to those objects—as shown in Figure 2.7.

These systems may also be used within the GIS to manage survey measurement data. They make it possible to store these data explicitly as an integral part of the geodatabase.

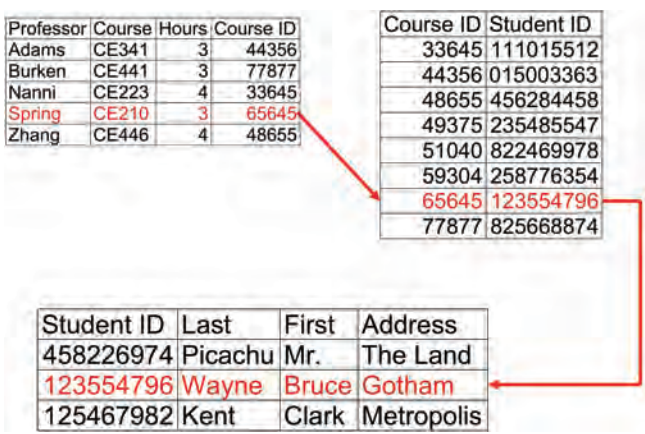


FIGURE 2.6 Linking tables in the RDBMS.

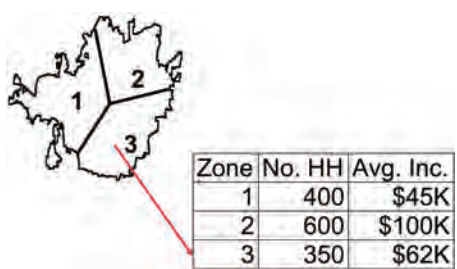


FIGURE 2.7 RDBMS in GIS.

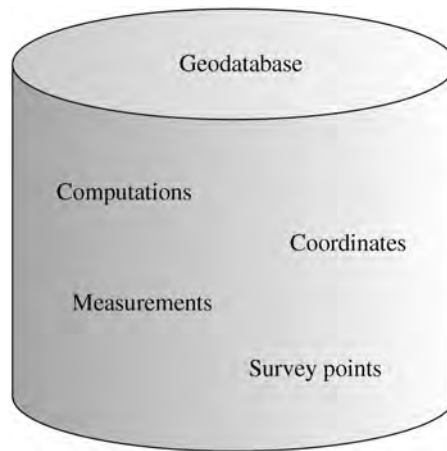


FIGURE 2.8 Survey data in the GeoDatabase.

An ESRI technical paper (ESRI, 2002) describes the data model used to solve problems related to the processing of survey data. Figure 2.8 depicts the conceptual framework.

Object-oriented database management systems (ODBMS) offer the ability to integrate the GIS database with object-oriented programming languages such as Java and C++. Martin and Odell (1992) state that the object-oriented approach “models the world in terms of objects that have properties and behaviors, and events that trigger the operations that change the state of the objects. Objects interact formally with other objects.” In short, the database has a set of objects with attached attributes. At this simplest level, ODBMS is analogous to the RDBMS model in that its objects represent the latter’s “rows” in tables, and its attributes the RDBMS fields. The ODBMS model is much more powerful than this simple definition implies, however. Three concepts are crucial to understanding the ODBMS: abstraction, encapsulation, and inheritance.

Data abstraction is the process of distilling data down to its essentials. The level of abstraction indicates what level of detail is needed to accomplish a purpose. For example, on a small scale, a road network may be presented as a series of line segments, but on a large scale, the road network could include medians, edges of roadways, and roadway fixtures (Rumbaugh et al., 1991). Encapsulation includes procedures with the object data. In other words, code and data are packaged together. The code performs some behavior that an object can exhibit; for example, calculating the present serviceability index of a section of road. Thus, encapsulation allows the representation of an object to be changed without affecting the applications that use it. Modern software programming provides a familiar example of this concept. The application Excel may be called as an object in a Visual Basic routine. The routine would then have access to the application’s full functionality through its attributes (built-in functions).

Finally, inheritance provides a means to define one class of object in terms of another. An object class represents a group of objects with common operations, attributes, and relationships (Fletcher et al., 1995). An object is a specific instance of a class. Each object can have attributes and operations (or code, as explained above). For example, a conifer is a type of tree. There are certain characteristics that are true for all trees, yet there are

specific characteristics for conifers. An attribute is a data value held by objects in a class (Rumbaugh et al., 1991), for example, for the attribute “tree color” the value might be “green.” An operation is a function that may be applied to or by objects in a class. Objects and classes of objects may be connected to other objects. That is, they may have a structural relationship called an association with these other objects. The primary purpose for inheritance is software reuse and consistency (Lewis and Loftus, 1998).

ODBMS hold great promise as GIS tools. Their power, with respect to their abilities to interface with object-oriented programming languages and to provide “smart” data, and their utility when applied to extremely complex databases all support their eventual pre-eminence in the GIS industry. Indeed, in the late 1980s most GIS vendors were moving from RDBMS to ODBMS models (Sutton, 1996). However, today most states continue to use the RDBMS models to manage their geospatial data.

2.3.1.4 Data Sources Digital data may be obtained in a variety of ways and from a large number of sources, all of which fall into one of the five categories listed below. The data sources of interest for measurement-based GIS are those that involve direct measurement of geographic information, such as remote sensing, GPS, and survey data. However, it is often the case that other sources of data are used as the basis for, or supplement to directly measured data, and are therefore included in this discussion.

- Digitizing source materials
- Remote sensing
- GPS for road centerline data
- Existing digital sources
- Survey data

Digitizing Source Materials Digitizing source materials involve digitizing data elements from aerial photographs (considered to be primary source material) or from hardcopy maps (considered to be secondary source material). The process involves collecting the x - y coordinate values of the line features by tracing over each one using a digitizing tablet with a cursor or puck as the input device to locate and input map features into the computer from a paper map. This may also be done using a mouse with a digital raster image map. This latter approach is called “heads up” digitizing. This type of manual production requires planning, source material preparation, and production setup, in addition to postprocessing of data (Davis, 1996). The costs for this type of data acquisition often represent the majority of system startup costs. Semiautomated methods using map scanning and line-tracing technologies are sometimes used to lower the cost and improve the accuracy of the digitizing process.

Remote Sensing Remote sensing is the science (and to some extent, art) of acquiring information about the Earth’s surface without actually being in contact with it. This is done by sensing and recording reflected or emitted energy and processing, analyzing, and applying that information. Remote sensing provides a convenient and efficient means of identifying and presenting planimetric data. Imagery is available in varying scales to meet the requirements of many different users. In general, there are two sources for remotely sensed data: photogrammetry, which uses cameras on-board airplanes, and satellite imagery. The former produces photographs that may either be in paper form, in which case

they must be digitized, as was described in the previous section, or in digital form, which may then be used directly as raster images. Cameras and their use for aerial photography are the simplest and oldest of sensors used for remote sensing of the Earth's surface. Cameras are framing systems that acquire a near-instantaneous "snapshot" of an area of the surface. Camera systems are passive optical sensors that use a lens (or system of lenses collectively referred to as the optics) to form an image at the focal plane, the plane at which an image is sharply defined. Aerial photographs are most useful when fine spatial detail is more critical than spectral information, as their spectral resolution is generally coarse when compared to data captured with electronic sensing devices. The geometry of vertical photographs is well understood and it is possible to make very accurate measurements from them, for a variety of different applications (geology, forestry, mapping, etc.). The use of satellites involves an interaction between incident radiation and the targets of interest. Note, however that remote sensing also involves the sensing of emitted energy and the use of nonimaging sensors. With the advent of online mapping software (as mentioned before, such as Google and Yahoo provide), satellite imagery for the entire United States and most countries around the world is available. These maps are fairly small scale and are therefore are mainly useful for planning purposes only.

Global Positioning Systems These systems use a constellation of 31 actively broadcasting satellites and spherical geometry calculations to determine geo-positions (Wikipedia, 2009). These data, in the form of coordinates, are transformed into path-oriented or linear locations. These measurements are referenced to a standard ellipsoid model instead of base maps or linear field monuments. This allows for new economies in the field and the continued use of existing legacy, and it supports path and network models, applications, and displays.

Existing Digital Data from Other Sources A cost-effective alternative to digitizing is to acquire digital data from a third-party source, such as those described below.

In the recent past, most GIS projects have had to rely almost exclusively upon data available only in printed or "paper" form. Much of these data are now available in digital form while continuing to be published on paper. The ever-increasing pace of this transformation from paper to digital sources has many repercussions for GIS. Inexpensive, and in many cases free, access to high-quality data will enhance the use of GIS in the coming years. The Internet and Worldwide Web are being used more and more to distribute data and information, thus requiring users to know where to look and how to search networks.

All data sources have strengths and limitations. Digital sources are no different. It is important to understand their characteristics, costs, and benefits before using them. Learning a little about commonly employed digital formats will save much work in the long run. Although the types of materials will vary greatly from project to project, GIS practitioners know something of the characteristics and limitations of the most commonly available data sources. These are materials collected and published by a variety of government agencies and commercial interests and are used quite widely.

Local, state, and federal government agencies are major suppliers of digital data. Better than 90% of states maintain databases with state and county route centerlines, and 98% of these also maintain other geospatial databases, as well. The majority of the states (62%) distribute these data free of charge and the others have policies in place that allow data to be shared with other public agencies. Finding the data appropriate for a given system may often require significant online research. This is perhaps less the case at the

federal level mainly because certain key agencies such as the Bureau of the Census, United State Geological Survey, Soil Conservation Service, National Aeronautics and Space Administration (NASA), and Federal Emergency Management Agency, provide standard sorts of information for the entire nation.

Key federal data sources and helpful indexes to those sources include the following.

The Bureau of Transportation Statistics (BTS,) provides links to state GIS resources, national transportation atlas data files, which consist of transportation facilities, networks and services of national significance throughout the United States. The files are in shape-file format, which is the proprietary format for ESRI data. There are also links to Dyna-Map/1000 files which contain centerline data and features for nearly every street in the Nation as well as the National Highway Planning Network which is a comprehensive network database of the nation's major highway system. It consists of over 400,000 miles of the nation's highways comprised of Rural Arterials, Urban Principal Arterials, and all National Highway System routes. The dataset covers the 48 contiguous States plus the District of Columbia, Alaska, Hawaii, and Puerto Rico. The nominal scale of the dataset is 1:100,000 with a maximal positional error of ± 80 m. The network was developed for national level planning and analysis of the U.S. highway network. It is now used to keep a map-based record of the National Highway System and the Strategic Highway Corridor Network.

The United States Geological Survey (USGS, 2010) produces topographic maps for the nation, as well as land use and land cover maps, which include information about ownership and political boundaries, transportation, and hydrography. For access to these data as well as an extensive index of other federal online data sources see the USGS publications web site. The index includes links to the Global Land Information System and tends to concentrate on data created through the USGS.

The Bureau of the Census (Census, 2010) provides socioeconomic and demographic data, census tract boundary files, and street centerline networks (TIGER files) for the entire nation. TIGER comes from the acronym Topologically Integrated Geographic Encoding and Referencing and was developed by the U.S. Census Bureau to support its mapping needs for the Decennial Census and other Bureau programs. The topological structure of the TIGER database defines the location and relationship of streets, rivers, railroads, and other features to each other and to the numerous geographic entities for which the Census Bureau tabulates data from its censuses and sample surveys. It has a scale of approximately 1 in equals 100 ft and is designed to assure no duplication of these features or areas.

The National Aeronautics and Space Administration (NASA, 2010) provides remotely sensed data from all over the world. These data are typically in raster format and are 30 m resolution.

The Federal Geographic Data Committee (FGDC, 2010), established by the Office of Management and Budget, is a clearinghouse for digital data, which is an excellent resource for online federal information of any kind.

As mentioned above, 45 states provide clearinghouses for their geospatial data, significantly easing the task of finding the data desired. Additionally, every state now has a contact person who takes responsibility for the state's GIS. Most data of this type are now available in standard formats, come with adequate documentation and data quality reports, and but still may need to be checked as to origin and quality. Furthermore, the spatial data holdings themselves have limitations. As mentioned earlier different data sources may have different scales associated with them and require conflation to integrate

them. Conflation, however, may not yield the information required. For example, the most widely held data, the USGS 7.5 min quadrangle data (1:24,000 scale), are increasingly becoming unsuitable for the uses of a transportation agency. For example, when collecting driveway data with a GPS unit at 1 m accuracy, centerlines based on 40 ft accuracy do not work well. Updating base maps with greater accuracy has the potential to be an expensive issue.

A primary advantage offered by government datasets is that most are in the public record and can be used for free or for a small processing fee. Some of the agencies that do not provide their data free of charge have found that this practice seriously impedes data sharing.

There also exist many private sources of information. Commercial mapmaking firms are among the largest providers, but other firms have for years supplied detailed demographic and economic information, such as data on retail trade and marketing trends. Some of this information can be quite expensive to purchase. Also, it is important to check on restrictions that might apply to the use of commercially provided data.

Many software vendors repackage and sell data in proprietary formats as well. These data are usually checked and corrected during the repackaging process. The use of these converted datasets can save time. Some firms will also build datasets to a user's specifications. These are often termed "conversion" firms. They are usually contracted to build special purpose datasets for utility companies and some government agencies. These datasets are often of such special purpose that they cannot be assembled from existing publicly available sources, say when an electric utility wishes to digitize its maps of its service area.

Survey Data Ground survey datasets have been developed over a long period of time. Some cadastral datasets, for example, were first created as many as 200 years ago. Given the dramatic improvements in the quality of measurement equipment over the years, integration of modern measurements into geodetic networks presents problems. Storage of coordinates only, as is done in conventional GIS packages, means that adding new, higher quality, measurements to datasets from ground surveys does not improve the quality of the dataset. Indeed, it is necessary to "fit" the results of the new measurements into the old dataset, resulting in a loss of quality. Several authors have described measurement-based systems (Buyong and Frank, 1989; Buyong et al., 1991; Hintz et al., 1996; Goodchild, 1999). According to Joffe (2003) the search for software is over and the ArcGIS extension Survey Analyst is exactly what we need: with ArcGIS Survey Analyst, the quest is over; these tools are now available to surveyors and GIS analysts alike. As a result, a government agency's GIS map base can now be built as a measurement-based multipurpose cadastre, and existing GIS map bases can gradually be transformed into this much desired data model (Joffe, 2003). Navratil et al. () explored the use of this tool with a small test dataset and found that it performs as claimed.

Data Quality It has been said that an undocumented dataset is a worthless dataset. If a dataset's pedigree and quality are unknown, for example, the user must spend time and resources checking the data. Vendors should provide a data dictionary that provides a description of exactly what is in the file (data types and formats), how the information was compiled (and from what sources), and how the data were checked. The documentation for some products is quite extensive and much of the detailed information may be published separately, as it is for USGS digital products.

Some characteristics to consider when evaluating datasets are as follows:

- Age
- Origin
- Areal coverage
- Map scale to which the data were digitized
- Projection, coordinate system, and datum used
- Accuracy of positional and attribute information
- Logic and consistency
- Format
- Reliability of the provider

In summary, sometimes the costs of using and converting publicly and commercially available digital files outweigh their value. No matter how much data becomes available publicly, there is no guarantee that it will contain exactly the sorts of information necessary for specific projects.

2.3.1.5 Data Standards For a variety of reasons, GIS users often want to share or exchange data, usually having to do with overlapping geographic interests. For example, a municipality may want to share data with local utility companies and vice versa. GIS software developed by commercial companies store spatial data in a proprietary format protected by copyright laws. Although vendors often attempt to provide converters, in general, one company's product may not read the data stored in another's format. Applications of digital geospatial data vary greatly, but generally users have a recurring need for a few common themes of data. These themes include transportation, hydrography (rivers and lakes), geodetic control, digital imagery, government boundaries, elevation and bathymetry, and land ownership (or cadastral) information. A lack of investment, common standards, and coordination has created many situations in which these needs are not being met. As a result in some cases, important information may not be available and in others, datasets may be duplicated. A means to maintain and manage the common information being collected by the public and private sector does not exist. This results in increased costs and reduced efficiency for all involved.

Thus, data standardization is a fundamental consideration in developing GIS for integration with existing databases. All users of GIS data depend upon the establishment of standards that should address simple integration and processing of data within the GIS. Standards should include spatial data modeling and scale, accuracy, resolution, and generalization, and datum and projection mapping. As illustrated by the Internet and its interoperability, hardware platforms, operating systems, network environments, database systems, and applications software are generally *not* an issue. The standards for data definitions are much more important than these latter items for providing reliable and portable systems and applications. The GIS software standards, however, can add to the complexity of sharing datasets, since not all GIS share a common route system that is easily transferred from one vendor-specific application to another.

During the 1980s the USGS worked with academic, industrial, and federal, state, and local government users of computer mapping and GIS to develop a standard for transfer and exchange of spatial data. In 1992, after 12 years of developing, reviewing, revising,

and testing, the resulting standard—SDTS—was approved as Federal Information Processing Standard (FIPS) Publication 173-1 (NIST, 1994). The SDTS requires a two-step process for transferring data from one platform to another. The source data are exported by the first GIS to the SDTS format, then the second GIS imports the transfer file, creating a target dataset in its own file format. This approach, while extremely useful in enabling data sharing, is cumbersome at best. Recognizing the criticality of data sharing The National Spatial Data Infrastructure (NSDI) was established by President Clinton's Executive Order 12906 on April 11, 1994 to implement the recommendations of the National Performance Review published by his Administration in the Fall of 1993 [superseded in 1998 by ANSI NCITS 320-1998 (ANSI, 1998)].

The Order states that

Geographic Information is critical to promote economic development, improve our stewardship of natural resources, and protect the environment. . . . National Performance Review has recommended that the Executive Branch develop, in cooperation with state, local, tribal governments, and the private sector, a coordinated National Spatial Data Infrastructure to support public and private sector applications of geospatial data . . .

The concept of this infrastructure was developed by representatives of county, regional, State, Federal, and other organizations under the auspices of the Federal Geographic Data Committee (FGDC, 2010). Rechartered in August 2002, the Committee, an interagency committee, promotes the coordinated development, use, sharing, and dissemination of geospatial data on a national basis. This nationwide data publishing effort is known as the National Spatial Data Infrastructure (NSDI).

The private sector is also actively engaged in building the “National Spatial Data Infrastructure” to meet market place needs defined as business opportunities.

The Open GIS Consortium of 388 private sector, public sector, universities, and not-for-profit organizations representing technology users and providers is addressing the issue of easy access to spatial information in mainstream computing. OGC is working to develop open software approaches that facilitate the development and use of location-dependent software applications using spatial data to increase farm productivity, identify disease and health threats, assist police and law enforcement in identifying crime patterns and many more. In other words, GIS users will be able to access one another's spatial data across a network even if they are using different GIS software programs (OGC, 2009). The International Standards Organization adopted its standard for geospatial metadata, ISO 19115, in 2003, with the intention that standards in member states would eventually be brought into compliance (ISO, 2003).

2.3.2 Spatial Modeling

The benefits of GIS are well established. It provides

- The capability of storing and maintaining large datasets of spatial and tabular information
- Display and analytical capabilities that model the physical proximity of spatial features
- Flexibility in modeling spatial objects to suit the particular needs of the user or application

- Database integration
- Image overlay capabilities
- Network analyses (e.g., shortest path routing).

These capabilities have developed as the technology has matured. In addition, GIS provides a programming environment that allows users to develop specific analysis programs or customize existing programs. All functions for display and analysis can be employed in a single-system design using common programming languages, such as Visual Basic, C++, and Java (Smith et al., 2001). With the advent of object-oriented programming, GIS can be integrated into more mainstream enterprise applications, as well as web-based client applications. GIS provides abilities broader than simply mapping data and includes several types of analytical capabilities that may be broadly categorized into five groups:

- Display/query analysis
- Spatial analysis
- Network analysis
- Cell-based modeling
- Dynamic segmentation

2.3.2.1 Display/Query Analysis The primary appeal of GIS to many is its graphical capabilities as the adage “a picture is worth a thousand words” bespeaks. Maps are the pictures GIS uses to communicate complex spatial relationships that the human eye and mind are capable of understanding. The computer makes this possible, but still, it is the GIS user that determines what data and spatial relationships will be analyzed and portrayed, or how the data will be thematically presented to its intended audience. In short, the GIS allows analyses at a higher level of abstraction than do standard database tools. Using the database capabilities of GIS, the analyst can query the database and have the results displayed graphically. This query analysis, when spoken in everyday conversation, takes on the form of a “show me” question, such as “Can you show me sections of road that are in poor condition?” However, query analysis in GIS can also be used for other purposes, such as database automation, which might be used for error checking and quality control of coded data. As an example, the GIS roadway database could be queried automatically during the crash data entry process to verify the accuracy of speed limit and other crash report variables coded by an officer.

2.3.2.2 Spatial Analysis Several analytical techniques, grouped under the general heading “overlay analysis,” are available in GIS for spatial analysis and data integration. GIS provides tools to combine data, identify overlaps across data, and join the attributes of datasets using feature location and extent as the selection criteria. For example, the number of acres of wetlands impacted by a proposed highway corridor could be obtained by this overlay process. Overlay techniques may also be used in combining data features by adding (or by applying some other function) one dataset to another, or by updating or replacing portions of one dataset with another—thus, creating a new spatial dataset. For example, the analyst could use these techniques to combine number of households and the average number of school age children with pedestrian-related crashes, in order to derive risk factors for the total number of pedestrian-related crashes relative to the total number of school age children per road segment, for pedestrian-to-school safety analysis.

Proximity analysis represents the fundamental difference of GIS from all other information systems. Buffering is a means of performing this practical spatial query to determine the proximity of neighboring features. It is used to locate all features within a prescribed distance from a point, line, or area, such as determining the number of road crashes occurring within one-half mile from an intersection, or, the number of households that fall within 100 ft of a wetland boundary.

2.3.2.3 Network Analysis Unlike proximity analysis that searches in all directions from a point, line, or area, network analysis is restricted to searching along a line, such as a route, or throughout a network of linear features, such as the road network. Network analysis can be used to define or identify route corridors and determine travel paths, travel distances, and response times. For example, network analysis may be used to assess the traffic volume impact of a road closure on adjacent roadways. GIS networking capabilities can also be used for the selection of optimal paths or routes; for example, finding shortest paths between zonal centroids. The network may include turning points, avoid improper turns onto one-way streets, represent posted traffic control restrictions, and include impedance factors to travel (such as mean travel speeds, number of travel lanes, and traffic volumes) to enhance the network analysis. It is this capability that one sees in the on-board navigation tools currently available in some models of automobile.

2.3.2.4 Cell-Based Modeling Cell-based modeling, also referred to as “grid-based” analysis, uses a grid or cells to aggregate spatial data for discrete distribution. Although similar to raster-based systems, these are vector-based. In cell-based modeling, the spatial data are developed as tiles of a given dimension, or points of a uniform distribution, as defined by the user, for display and analysis. Cell-based modeling is effective in displaying patterns over larger areas, such as representing groundwater contamination levels using statistical models as the mathematical base. Since cell-based modeling aggregates data at a specified grid resolution, it *would not* be appropriate for site-specific spatial analysis. In cell-based modeling, special tools are available to merge grid data for overlay analysis. Cell-based overlay analysis is similar to the GIS overlay analysis previously discussed; however, the techniques and functions available in cell-based modeling are somewhat different. When the cells of different datasets have been developed using the same spatial dimensions, they can be merged on a cell-by-cell basis to produce a resulting dataset. The functions and processes used in cell-based modeling to merge grid data are referred to as “map algebra,” because the grid datasets in cell-based modeling are merged using arithmetic and Boolean operators called “spatial operators.”

2.3.2.5 Dynamic Segmentation Dynamic segmentation (dynseg) is a process that allows the association of multiple sets of attributes to a portion of a linear feature without having to modify feature geometry or topology. Although implementation of dynamic segmentation varies by GIS vendor, GIS uses dynamic segmentation to locate and display linear features along a route and/or to segment the route itself. The process consists of interpolating the distance along the measured line of the GIS route from the beginning measure to the ending measure of the line using attributes as the interpolation criteria. In short, the dynseg process allows temporary modification of feature geometry “on the fly” based upon the attribute data being considered. Consider, for example, a vector in a

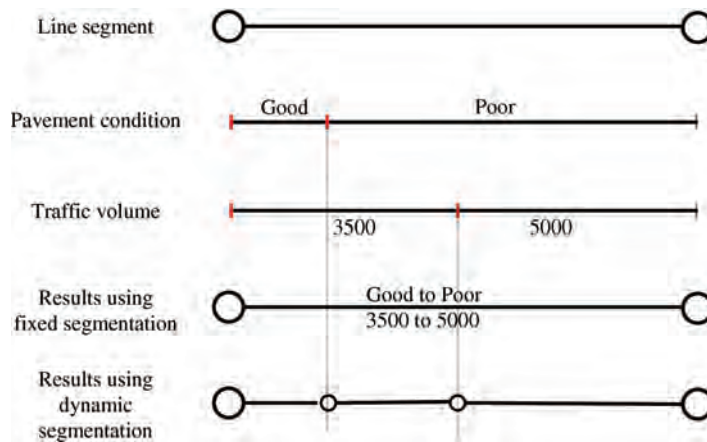


FIGURE 2.9 Dynamic segmentation example.

database representing a segment of road. As explained earlier, this line object is actively linked to a row in a table of attribute data—say road condition. The issue addressed by dynamic segmentation is the inability of the GIS to represent features that do not begin at the beginning of the line and end at its end. In other words, using the road condition example, a road segment's geometry may run from station 0 + 00 to station 5 + 00 but the road condition may (and probably does) change somewhere in between. Figure 2.9 depicts a situation where pavement condition and traffic volume levels are shown along the line segment and demonstrates the difficulty in determining various combinations of pavement condition and traffic volume using fixed segmentation. As the exhibit shows, dynamic segmentation allows representation of these nonuniform attributes without requiring the physical segmentation of the link; that is, no modification of geometry is necessary.

This robust method to represent model links allows the dynamic management of many-to-one relationships between GIS segments attribute features. The method offers several advantages over fixed segmentation. First, the editing of the routes is preferable to editing the underlying base map. Second, the routes may be stored in simple correspondence tables that can be easily read by planning models, for example. Third, depicting the transportation network as linear features (or "linear events") in the GIS means that route tables and associated transit line tables can be managed automatically through dynamic segmentation. Dueker and Vrana (1992) describe dynamic segmentation in great detail.

2.4 MEASUREMENT-BASED GIS APPLICATIONS

Just 10 years ago, GIS was primarily used in single applications. Now it has evolved into serving enterprise wide information through multiplatform applications. Additionally, GIS application development has grown in the past few years significantly. More than one-third of states now spend in excess of a quarter of a million dollars solely on GIS

application development, almost half of which is outsourced. There has been a significant increase in recent years in the number of state departments of transportation developing more accurate road base maps using digital imagery and kinematic GPS. There has also been significant growth in web-based GIS applications and the use of GIS for managing infrastructure inventories (AASHTO, 2009).

Fletcher and Elfick (2009) developed an apparatus that will define a parcel network in a control coordinate system formed from a number of interconnected parcels. This allows automatic creation of GIS databases from existing boundary map data.

Survey Analyst by ESRI integrating survey measurement data into the GIS has been used for a variety of applications to good effect. New South Wales (NSW), Australia, owns and operates the electrified metropolitan railway network throughout Sydney and surrounding commuting areas and is upgrading parts of that system. It used the software to compile survey plans of different ages, dimensions, and orientations resulting in a survey-accurate cadastral database with property attributes to underpin the project GIS database in the ArcGIS environment. The City of Encinitas, CA used Survey Analyst to integrate existing survey documents and records into a new cadastral fabric.

A measurement-based GIS approach is also being used by Encora as part of its European Action Plan for Coastal Management. The software handles data on all spatial scales (entire regional coast vs. a single harbor) and allows comparison of measurements from different years, as well as overlay analysis of measurements and modeling results (Encora, 2007).

Duckham (2002) and Heuvelink (2005) have described efforts to build error-aware systems, and data quality is now an important element of metadata. But, as Goodchild (2008) points out, the mainstream GIS products continue to report the results of calculations to far more decimal places than are justified by any assessment of accuracy, and to draw lines whose positions are uncertain using line widths that are in no way representative of that uncertainty. Indeed, GIS practice seems still to be largely driven by the belief that accuracy is a function of computation, not representation, and that the last uncertainties were removed from maps many decades ago.

2.5 IMPLEMENTATION ISSUES

Implementing GIS is not as simple as merely installing a new piece of software and then operating it. In addition to the information technologies intrinsic to GIS, successful GIS implementation involves elements, such as personnel and their GIS skills, the organizational structure within which they work, and the institutional relationships that govern the management of information flow (Hall, 2006).

GIS is an enabling technology and serves as a platform for integrating various types of data, systems, and technologies. It may be applied to a host of different applications, as demonstrated by the preceding limited review of current applications. Among these are systems that allow for near real-time assessment of conditions, simultaneous sharing of large databases and integration, and enhancement of existing models—all complex and challenging applications. The fact that the adoption rate of GIS technologies has grown exponentially makes successful implementation of these technologies even more challenging. Key areas that must be addressed when implementing GIS include data, people and organizations, and technology.

2.5.1 Data

The collection, maintenance, and use of data are challenging issues that must be addressed early in the implementation process. Who will collect and maintain data, who “owns” the data, and who will serve as its custodian are all questions that must be answered.

The main reason that these questions are important is that they are key in determining if one can take advantage of GIS’s main strength, namely, its ability to share data. That is, as stated previously, GIS serves as a logical and consistent platform in a common location reference system and allows diverse databases to be integrated and shared among different divisions of a department, for example. Integration standards should therefore be established to integrate different databases, some spatial, some not. Answers to the questions also address barriers, such as institutional and organizational arrangements, to the implementation process.

Technical questions, such as what is the nature of the data, how is it to be spatially referenced, and what is its accuracy must be considered—as well as the means by which data from various sources and applications can be integrated. To understand the various data elements in a complex system, metadata (information about data, such as when collected, by whom, scale) should be part of a system. Additionally, spatial data requires establishment of some sort of linear referencing system as was described earlier.

All GIS data, spatial data, and associated attribute data, suffer from inaccuracy, imprecision, and error to some extent. Data quality assurance and quality control rules ensure the delivery of high-quality data. Use of a data steward to collect, maintain, and disseminate GIS data would facilitate this.

2.5.2 Technology

While GIS applications no longer bear the stigma of “new technology” and have been accepted into the mainstream of professional practice, there nevertheless exist technological issues that must be addressed when implementing them. Among these are the identification of critical technologies to be used, and keeping abreast of technological innovations. These include technologies that support interoperability (sharing of data and processes across application/system boundaries), web-based GIS (given its significant growth in the past few years), and a rapidly expanding range of GIS applications. Of course, chosen technologies must match the architecture environment as well. The fact that industry standards are in their infancy exacerbates the problem of choosing the appropriate technologies for a given system. It is crucial to identify limiting technologies, statements by software vendors notwithstanding.

In making this decision one should look to the future, but not too far into the future. One should also keep in mind that there exists an increasing pressure to make geo-referenced data more accessible (and understandable) to the general public.

The introduction of new information technologies is necessarily accompanied by a change in organizational structures and institutional arrangements.

2.5.3 Organizations and People

Organizational and people issues are perhaps the most difficult to address. They continue to be more critical and more difficult to solve than technological issues. Personnel at both management and technical levels must be involved in implementing the system. Yet, this

1. Integration of information for better decisions
2. Decreased risk of poor decisions due to incomplete data
3. Access to more accurate data
4. Reduction of duplicate data
5. Ability to quickly visualize interrelationships of various data and projects
6. Ability to develop public information maps
7. Assist with revenue initiatives
8. Quick response to internal and external queries
9. Easy access to information and maps
10. Significant time savings for information access
11. Unprecedented analysis of information
12. Identification of trends
13. Assist in public relations
14. Better cooperation with regulatory agencies
15. Increased safety for personnel
16. Ability to monitor department commitments

FIGURE 2.10 Intangible benefits of GIS.

has been identified consistently as a problem area—especially maintaining support of upper management. Convincing decision-makers to accept the idea of GIS is key to a system's success. A top-down, rather than bottom-up GIS management strategy should be adopted for GIS planning and implementation.

Of primary interest to decision-makers are two questions: Who pays (for the hardware, software, and personnel required) and from where will the resources be drawn? As with any investment decision, GIS must have economic justification. Hall et al. (2000) conducted a benefit cost analysis of GIS implementation for the state of Illinois. They calculated not only a ratio of slightly less than one but also identified several intangible benefits for GIS project, shown in Figure 2.10, which indicate that GIS benefits outweigh implementation costs.

Champions (at both the management and “grass-roots” levels) are critical to a system's success as well. They can facilitate a positive decision to purchase and enhance the chances for system success.

With regard to the latter, qualified GIS personnel must be available. What training and education are necessary for minimum acceptable qualifications are based upon the level of GIS knowledge needed. There are likely to be three levels of GIS staff and users: local GIS users, local GIS specialists, and a GIS application/data steward. Training for the GIS support personnel is extremely important to the success of GIS.

Early implementation of GIS is often more dependent on vendor supplied training. However, in the long term, the GIS support group should develop specifications for in-house training. Several GIS entities (Zhang et al., 2001) have proposed that a certification for GIS professionals be established to address this important issue. Getting and retaining adequately trained GIS staff continue to be problems. Indeed, there will be increased reliance on outside experts (consultants) for more complex GIS analyses (because of the difficulty in finding and keeping expert staff on internal payroll).

Additional problems of which to be aware relate to the nature of GIS data, namely, that, to be successful, it must be shared. This leads to turf battles and questions about who should have access and how much to the GIS.

2.6 CONCLUSION

Emerging issues and applications at the forefront of current attention include the following:

- Metadata
- Data visualization
- Enterprise GIS

2.6.1 Metadata

Metadata, a description of the contents of a dataset, serves several independent purposes, all related to the ability of users to assess the suitability of a body of information for a specific use. Metadata provides information about quality, the degree to which contents match requirements, technical details about using the information, the legal and ethical constraints on use, and sources of further information.

With advancements in technology, our ability to share data has expanded, as has the concomitant complexity of metadata. Sharing of data with a colleague in the same department is relatively easy, since both the custodian and the potential user probably share a common discipline and language and common set of expectations. Sharing data over the Internet with potential users in other countries, cultures, and disciplines, on the other hand, introduces problems similar to those probably faced by early explorers in communicating with the native inhabitants of the Americas. Thus, there is the need for establishing good data standards governing quality.

The U.S. National Spatial Data Infrastructure, described earlier, includes the Content Standard for Digital Geospatial Metadata (CSDGM). Quality is a major part of the standard but falls short for current needs. Comber et al. (2007) argued for a more user-centric approach to metadata, and one that places greater emphasis on spatial data quality. Goodchild (2009) sets forth seven areas for improvement to the Federal standards relative to the use of metadata. They are summarized below.

2.6.1.1 Decoupling of Content from Positional Accuracy As described in the data section of this chapter, the representative fraction of an original map is often used as a measure of positional accuracy. However, in principle, representative fraction is not defined for digital data, since there are no distances in digital media to compare to distances in the real world. Goodchild and Proctor (1997) consequently argue for the decoupling of content, spatial resolution, and positional accuracy under this single surrogate measure. Current metadata standards do not adequately reflect this. The authors of the standards appear to be aware of the difficulties associated with representative fraction as a surrogate for several aspects of data quality, but have not fully adopted the decoupled approach that now seems needed in its place.

2.6.1.2 Uncertainty Early discussions of the quality of digital geographic datasets focused on concepts of accuracy and error, perhaps reflecting the roots of GIS in area measurement (Foresman, 1998) and the earlier work of Maling (1989). The terms *error* and *accuracy* are now generally avoided in the research community, which tends instead to favor *uncertainty* as the umbrella term, along with *imprecision*, *vagueness*, and terms more closely related to the theories of evidence and non-Boolean sets. Yet neither standard shows any evidence of this significant change of thinking. The term *accuracy* occurs 7 times in the ISO standard and 85 times in the FGDC standard, while *uncertainty* occurs in neither.

2.6.1.3 Separability The FGDC and ISO standards distinguish clearly between the accuracies of attributes and positions but it has long been known that in the case of geographic variation conceptualized as a continuous field where the two concepts are not readily separable. It therefore makes little sense to attempt to specify the positional accuracy of iso-lines, or to separate attribute and positional accuracies for area-class maps, as both the FGDC and ISO standards do.

2.6.1.4 Granularity In the earlier world dominated by paper maps, the body of information described by metadata was a single map, and an intimate association existed between a map's contents and its marginalia. In the digital world, however, the concept of a dataset is much more fluid. Existing arrangements for handling metadata attempt to describe the database at the level of the class or collection of features (the *feature dataset* in the terminology of ESRI's Geodatabase). But one could equally well argue that metadata are needed at the level of the entire database, and that the quality of the information about relationships between classes needs to be described.

2.6.1.5 Collection-Level Metadata Metadata that describe the properties of individual datasets are sometimes termed *object-level metadata*, since they focus on a single information object within the larger framework of an entire collection. Many such collections exist in the form of geospatial data warehouses, digital libraries, or geo-libraries, each containing potentially thousands of separate datasets, as described earlier. Users require guidance as to which collections to search based upon their individual needs. Collection-level metadata (CLM; Goodchild and Zhou, 2003) is defined as data about the contents of an entire collection, describing such characteristics as geographic and temporal coverage, the set of themes that dominate the collection, and the general level of data quality. Efforts have been made to develop content standards for CLM (http://www.alexandria.ucsb.edu/~lhill/alex-imp/Metadiversity_narrative.html), but the task of describing collections is far more complex than the task of describing individual datasets.

2.6.1.6 Autocorrelation Tobler's First Law (Tobler, 1970; Sui, 2004) describes the tendency for "nearby things to be more similar than distant things." There is now abundant evidence that this principle applies to errors and uncertainties in geographic datasets. Knowledge of covariance of errors may be of only limited significance to the visualization of geographic data in maps, but it is critical to any analysis of the propagation of uncertainties during manipulation of spatial datasets. Virtually all interesting products of GIS analysis, from simple measures of slope or area to complex analyses, respond directly to the covariances of errors and uncertainties. Thus, appropriately defined parameters should be an essential part of any attempt to describe data quality in metadata. Yet current standards focus entirely on marginal properties such as mean positional error.

2.6.1.7 Cross-Correlation This discussion of autocorrelation leads directly to the final issue, which is in many ways the most problematic. Although the ability to overlay disparate layers is often presented as a major advantage of a GIS approach, many users will have experienced the problems of misfit that almost always occur. The problem of conflation described earlier. If the positional uncertainties in two layers are other than perfectly correlated, and if both layers contain representations of the same features then the result of overlay will be a large number of small slivers, formed by the two versions of each feature. While it is possible to describe the uncertainties of each dataset independently, the results of overlay cannot be obtained from this information—misfit is a *joint* property

of a pair of datasets, rather than a *marginal* property of either of them. Such information seems essential to the entire GIS enterprise, in so far as it is based on the ability to overlay, and to extract layers of data from widely disparate sources. Great effort has been expended over the past decade at making geographic technologies and datasets interoperable. Yet data quality has received very little attention in this drive to open, interoperable GIS (<http://www.opengeospatial.org>), and the approach to metadata reflected in the standards is uniformly unary.

2.6.2 Data Visualization

Data visualization continues to be an emerging issue for the GIS-T community. The growing popularity of commercial visualization tools such as Microsoft's Virtual Earth and Google Earth have put pressure on GIS-T practitioners to build similar tools. The "look and feel" and speed of the commercial tools are sought by executives and the public. Traditional GIS tools, by themselves, do not have these attributes. Particular functions of interest are as follows:

- Data and network security concerns
- Additional and new licensing, maintenance, and development costs
- Executive and management expectations
- Mash-up mapping techniques
- Three-dimensional views

It is an exciting time to be involved in GIS with the advance of geospatial technology into the mainstream. However, expectations for the technology have risen above the resources available to deliver on them. Commercial technology integration with traditional GIS tools will require follow-up discussions concerning development, implementation, and lessons learned.

2.6.3 Enterprise GIS/Asset Management

Although a commonly used theme, this remains poorly defined to the GIS-T community. Without a proper definition, attainment cannot be determined nor measured, and the role of GIS remains obscure (AASHTO, 2009).

This chapter has provided a brief overview of GIS, defined terms, reviewed general principles and analysis technologies, identified issues, and described major application areas within the measurement-based GIS arena. This material can only serve as a superficial introduction to the applications of GIS to engineering measurement problems but it is hoped that the reader will gain valuable insight into how these systems work, where they have been applied and what difficulties GIS users face when implementing such systems. It is further hoped that, with the "veil of ignorance" removed, users will more readily consider this important class of tools for use.

REFERENCES

- AASHTO. Summary Report for 2008 Geographic Information Systems in Transportation Symposium. American Association of State Highway and Transportation Officials 2008, <http://www.gis-t.org/>.

- AASHTO. Summary Report for 2009 Geographic Information Systems in Transportation Symposium. American Association of State Highway and Transportation Officials 2009, <http://www.gis-t.org/>.
- ANSI. American National Standards Institute, Spatial Data Transfer Standard, ANSI NCITS 1998 320-1998.
- Buyong TB, Frank AU, Kuhn W. *A conceptual model of measurement-based multipurpose cadastral systems*. URISA Journal 1991;(3):35–49.
- Buyong TB, Frank AU. “Measurement-Based Multipurpose Cadastre.” Proceedings of the ACSM/ASPRS Annual Convention. Baltimore: Maryland 1989. Vol. 5, p. 55–66.
- BTS U.S. Bureau of Transportation Statistics. Accessed December 2010. <http://www.bts.gov/>.
- Census. 2010. U. S. Census Bureau. Accessed December 2010. <http://www.census.gov/>.
- Comber AJ, Fisher PF, and Wadsworth RA. User-focused metadata for spatial data, geographical information and data quality assessments. Proceedings, Tenth AGILE International Conference on Geographic Information Science, Aalborg, Denmark. 2007. http://www.plan.aau.dk/~enc/AGILE2007/PDF/71_PDF.pdf.
- Davis B. *GIS: A Visual Approach*. New Mexico: OnWord Press; 1996.
- Duckham M. *A user-oriented perspective of error-sensitive GIS development*. Transactions in GIS 2002;6(2):179–194.
- Dueker KJ, Vrana R. Dynamic segmentation revisited: a milepoint linear data model. AASHTO GIS-T Symposium, Portland, OR; 1992.
- Encora. The European Action Plan for Strengthening the Knowledge Base of Sustainable Coastal and Marine Management. Encora Paris Conference 5–7 December. 2007. http://www.coastal-wiki.org/coastalwiki/European_Coastal_Action_Plan_2008.
- ESRI. ArcGIS Survey Analyst Concepts. ESRI Technical Paper, October. 2002.
- FGDC. Data Clearing House. Federal Geographic Data Committee. Accessed December 2010. <http://www.fgdc.gov>.
- Fletcher D, Henderson T, Espinoza J. 1995. Geographic Information Systems–Transportation ISTE Management Systems, Server–Net Prototype Pooled Fund Study Phase B Summary. Sandia National Laboratory, Albuquerque, New Mexico.
- Fletcher MJ, Elflick MH. 2009. U.S. Patent No. US 7,574,302 B2 August.
- Foresman TW, editor. *The History of Geographic Information Systems: Perspectives from the Pioneers*. Upper Saddle River, NJ: Prentice Hall PTR; 1998.
- Goodchild MF, Proctor J. *Scale in a digital geographic world*. Geographical and Environmental Modelling 1997;1(1):5–23.
- Goodchild MF. *Uncertainty: the Achilles heel of GIS?* Geo Info Systems 1998;November:50–52.
- Goodchild MF. Measurement-based GIS. In: Shi W, et al., editors. *Proceedings, International Symposium on Spatial Data Quality*. Hong Kong: Hong Kong Polytechnic University; 1999. p. 1–9.
- Goodchild MF, Zhou J. *Finding geographic information: collection-level metadata*. GeoInformatica 2003;7(2):95–112.
- Goodchild MF. Foreword; imprecision and spatial uncertainty; spatial data analysis. In: Shekhar S, Xiong H, editors. *Encyclopedia of GIS*. New York: Springer; 2008.
- Goodchild MF. Putting research into practice. In: Stein A, Shi W, Bijker W, editors. *Quality Aspects of Spatial Data Mining*. Boca Raton: CRC Press; 2009. p. 345–356.
- Hall JP. Geospatial Information Technologies for Asset Management, Transportation Research Circular E-C108, Transportation Research Board. 2006.
- Hall JP, Kim TJ, Darter MI. *Cost-benefit Analysis of Geographic Information System Implementation: Illinois Department of Transportation*. Transportation Research Record 2000;1719:219–226, Transportation Research Board.

- Heuvelink GBM. *Handling spatial uncertainty in GIS: development of the data uncertainty engine*. Estoril, Portugal: Instituto Geografico Portugues; 2005.
- Hintz RJ, Wahl JL, Wurm K, McKay D. "Geographic Measurement Management: An Operational Measurement-Based Land Information System." Proceedings of the ACSM/ASPRS Annual Convention, Bethesda: Maryland 1996. Vol. 3, p. 141–149.
- ISO. 2003. ISO 19115: Geographic Information-Metadata. International Organization for Standardization. http://www.iso.org/iso/iso_catalogue/catalogue_tc/catalogue_detail.htm?csnumber=26020.
- Joffe B. Survey Analyst: A Dream Come True. ArcNews; 2003. p. 25.
- Kuhn W. A functional ontology of observation and measurement. Proceedings of the Third International Conference on GeoSpatial Semantics, Mexico City 2009;3–4 December, 2009.
- Lewis J, Loftus W. *Java Software Solutions: Foundations of Program Design*. Reading, MA: Addison-Wesley; 1998.
- Maling DH. *Measurement from Maps: Principles and Methods of Cartometry*. New York: Pergamon; 1989.
- Martin J, Odell J. *Object-Oriented Analysis and Design*. Englewood Cliffs, NJ: Prentice-Hall, Inc.; 1992.
- NASA. National Aeronautics and Space Administration. Accessed 2010. <http://www.nasa.gov/>.
- Navratil G, Franz M, Pontikakis E. Measurement-Based GIS Revisited. Proceedings of the 7th AGILE Conference on GIScience. April 29–May 1, 2004. Heraklion, Crete/Greece. http://plone.itc.nl/agile_old/Conference/greece2004/heraklion.html.
- NIST. Spatial Data Transfer Standard, Federal Information Processing Standards Publication 173-1, National Institute of Standards and Technology. 1994.
- OGC. Open GIS Consortium. Accessed 2009. <http://www.opengeospatial.org>.
- Padmanabhan G, Leipnik MR, Yoon J. 1992. A Glossary of GIS Terminology, National Center for Geographic Information and Analysis, Report No. 92-13, ftp://ftp.ncgia.ucsb.edu/pub/Publications/Tech_Reports/92/
- Rumbaugh J, Blaha M, Premerlani W, Eddy F, Lorensen W. *Object-Oriented Modeling and Design*. Englewood Cliffs, NJ: Prentice-Hall, Inc.; 1991.
- Sadiq Z, Duckham M. Integrated storage and querying of spatially varying data quality information in a relational spatial database. Transactions in GIS, International Conference on Geosensor Networks, Oxford, UK; 2009. <http://geosensor.net/2011/>
- Smith RC, Harkey DL, Harris R. "Implementation of GIS-Based Highway Safety Analyses: Bridging the Gap." FHWA-RD-01-039, Federal Highway Administration. Bethesda: Maryland 1996. Vol. 3, p. 141–149.
- Sui DZ. *Tobler's First Law of Geography: a big idea for a small world?* Annals of the Association of American Geographers 2004;94(2):269–277.
- Sutton JC. *Role of geographic information systems in regional transportation planning*. Transportation Research Record. 1996;1518:25–31, Transportation Research Board.
- Tobler WR. *A computer movie simulating urban growth in the Detroit region*. Economic Geography 1970;46(2):234–240.
- USGS. U.S. Geological Survey. Accessed December 2010. <http://www.usgs.gov/>.
- Wikipedia. Global Positioning Systems. Accessed 2009. http://en.wikipedia.org/wiki/Global_Positioning_System.
- Zhang JX, Goodchild MF. *Uncertainty in Geographical Information*. New York: Taylor and Francis; 2002.
- Zhang Z, Smith SG, Hudson RW. *Geographic Information System Implementation Plan for Pavement Management Information System: Texas Department of Transportation*. Transportation Research Record 2001. 1769:46–50, Transportation Research Board.

3

TRAFFIC CONGESTION MANAGEMENT

NAGUI M. ROUPHAIL

- 3.1 Introduction and background
- 3.2 Scope of the chapter
- 3.3 Organization of the chapter
- 3.4 Fundamentals of vehicle emission estimation
 - 3.4.1 VMT
 - 3.4.2 VMT-S
 - 3.4.3 Mode
 - 3.4.4 VSP
 - 3.4.5 Hybrid VSP-S
- 3.5 Inventory of traffic congestion management methods
 - 3.5.1 Travel demand management (TDM)
 - 3.5.2 Traffic operational strategies (TOS)
- 3.6 Assessing emission impacts of traffic congestion management
 - 3.6.1 Direct measurement methods
 - 3.6.2 Mobile emission estimation models
 - 3.6.3 Transportation models
- 3.7 Summary
- Acknowledgments
- References

3.1 INTRODUCTION AND BACKGROUND

In striving toward an environmentally conscious transportation system, the field of traffic congestion management offers a diverse toolbox of solutions to traffic and transportation engineers, all of which are aimed at mitigating the effect of congestion on emissions and air quality. Traffic congestion continues to be on an upward trajectory in the United States. According to the 2005 Urban Mobility Report by the

Texas Transportation Institute, urban daily vehicle miles of travel (VMT) more than doubled in the period of 1982–2003, resulting in a near threefold increase in peak hour delays experienced by motorists in the same time period (from 16 to 47 h of annual delay per peak traveler) (ICF International, 2006). The same study estimated that traffic operational improvements (not including public transportation solutions) could reduce the overall hours of delays by nearly 10% and annual congestion costs by nearly 9%.

The evidence connecting transportation sector use in the United States to prevailing emissions levels for various pollutants is indeed compelling: Direct emissions from surface transportation account for nearly 45% of total national annual emissions of nitrogen oxides (NO_x), 59% of CO, 38% of volatile organic compounds (VOC), and 50% of particulate matter less than 10 μm in diameter (PM_{10}) (Schrang and Lomax, 2005). In addition, electrified surface transportation systems, such as light rail or electric vehicles, generate indirect emissions at the power plants. NO_x , CO, and VOC are precursors to ozone formation in the troposphere layer (the lowest portion of Earth's atmosphere). This layer contains approximately 75% of the mass of the atmosphere and almost all the water vapors and aerosols. Tropospheric ozone is a greenhouse gas that contributes significantly to climate change. Over 158 million Americans live in regions exceeding the National Ambient Air Quality Standard (NAAQS) for ozone (EPA, 2006). Thus, traffic congestion management strategies, whose focus is on improving the operation of the surface transportation system, offer a realistic and substantive means of reducing emissions and improving air quality in the short- and long-term horizons.

3.2 SCOPE OF THE CHAPTER

Congestion management programs that are aimed at reducing overall vehicle emissions on a corridor or network generally fall into two classes: (1) those intended to *restrict the demand or travel intensity* of the transportation system and (2) those aimed at *improving the operational performance* of the transportation system. Some strategies, such as improved transit operations and ridership, could fall into both categories.

In this chapter, the first class will be referred to as *travel demand management (TDM) strategies* and the second as *traffic operational strategies (TOS)*. Absent from this classification are programs aimed at improving vehicle fleet characteristics as they relate to fuel efficiency, the use of alternative fuels, and the associated emission rates, and those targeted at specialized vehicle fleets, such as idling reduction programs for trucks and other heavy vehicles. Thus, the focus here is primarily on the aggregate traffic performance of the average commuter fleet.

Finally, while there are obvious linkages between vehicle emissions and regional air quality measures, these are not explored in this chapter. Air quality indicators generally incorporate multiple sources of emissions, including industrial, residential, and commercial uses, in addition to mobile sources, and are also affected by the regional topography and atmospheric conditions. Therefore, the presentation in this chapter focuses on the emission consequences of traffic management strategies, with the understanding that these strategies relate to only a portion of the total emissions inventory in a region. The relative magnitude of this portion will also vary according to the pollutant considered, as indicated in the national inventory statistics presented in Section 3.1.

3.3 ORGANIZATION OF THE CHAPTER

This chapter is organized around three principal themes. The first theme (Section 3.4) covers the *fundamental measures of emissions* and how they relate to vehicular activity at the individual and traffic stream levels. The second theme (Section 3.5) provides a *survey of traffic congestion management methods* aimed at reducing vehicle emissions, focusing on both demand and supply management. The last theme (Section 3.6) is related to emission-based *assessment methods of the effectiveness of congestion management techniques*, which form the empirical and modeling tools available to engineers and researchers to estimate or predict emission consequences of various traffic control actions or policies.

3.4 FUNDAMENTALS OF VEHICLE EMISSION ESTIMATION

Two fundamental measures underlie the estimation of emissions (E) as they relate to traffic congestion management. These are measures of the vehicle activity (A) and emission factors (EF), where $E = A \times \text{EF}$. In other words, the EF represents the generated emissions per unit vehicle activity.

Vehicle activity can be defined at various levels or scales of aggregation, each with its own purpose, and the associated emission factor. For traffic congestion management purposes, five activity measures are discussed in the following section:

1. Vehicle miles of travel (VMT)
2. Vehicle miles of travel at specified speeds or speed ranges (VMT-S)
3. Travel time distribution by driving mode (MODE)
4. Travel time distribution by vehicle specific power (VSP) mode
5. Travel time distribution using a hybrid VSP-speed (VSP-S) approach

3.4.1 VMT

At a very coarse level, one can define the vehicle activity to be the number of annual vehicle miles traveled by the average gasoline-powered, light-duty vehicle in the United States. According to the 2005 National Transportation Statistics such a vehicle has hydrocarbon (HC), carbon monoxide (CO), and nitrogen oxides (NO_x) emission factors that are equivalent to 1.25, 12.57, and 0.92 g/mile (U.S. Department of Transportation, Bureau of Transportation Statistics, 2005). Therefore, if such a vehicle traveled 10,000 miles per year, its contribution to the annual emission inventory would be 12.5, 125.7, and 9.2 kg/year for HC, CO, and NO_x emissions, respectively.

3.4.2 VMT-S

Although the previous approach could be useful in providing gross estimates of the effect of VMT changes on regional emission levels, it is fundamentally inadequate for characterizing the effectiveness of congestion management strategies. This is because the *nature* of the vehicle activity, and not just its *magnitude*, can have a profound effect on the corresponding emission factor. To illustrate this effect, consider the case where additional

TABLE 3.1 Illustrative Vehicle Speed Effect on CO Emissions

Speed range (mph)	≤10	10–30	30–50	>50
Annual mileage at given speed range	2000	3000	3000	2000
Corresponding EF (g/mile)	18	9	10	11
Annual emissions in speed range (kg)	36	27	30	22

information is known about the mileage distribution of the subject vehicle travel speed throughout the year, as summarized in Table 3.1 (considering CO emissions, and for illustrative purpose only). The implication here is that the emission factor varies with the vehicle speed. Activity in this context is defined as the annual miles traveled in a given speed range.

On the basis of the depicted values, the annual CO emissions in this case would add up to 115 kg/year (contrasted with the 125.7 kg/year value computed earlier). Therefore, a change in the speed distribution—due to changes in the prevailing congestion level—will be automatically reflected in the estimation of total annual emissions produced by the vehicle.

3.4.3 Mode

It is possible to further pursue this line of increased detail by refining the definition of vehicle activity to include the effects of short-term (also known as *microscale*) events that are associated with changes in the emission factor. A *microscale* event represents a different load requirement on the vehicle engine and, therefore, generates different fuel and emission rates. One way to represent these events is to categorize the trip time (or trip distance) into driving modes. A simplified modal emission model by Frey et al. (2002) considers four such modes: acceleration, deceleration, idle, and cruise. Each mode is then associated with a unique emission factor (per unit time or distance), and the activity is defined as the trip travel time spent in each of the four modes.

An example of modal emission factors computed from on-board vehicle measurements in the Frey study is illustrated in Figure 3.1, which depicts the average and 95th percentile

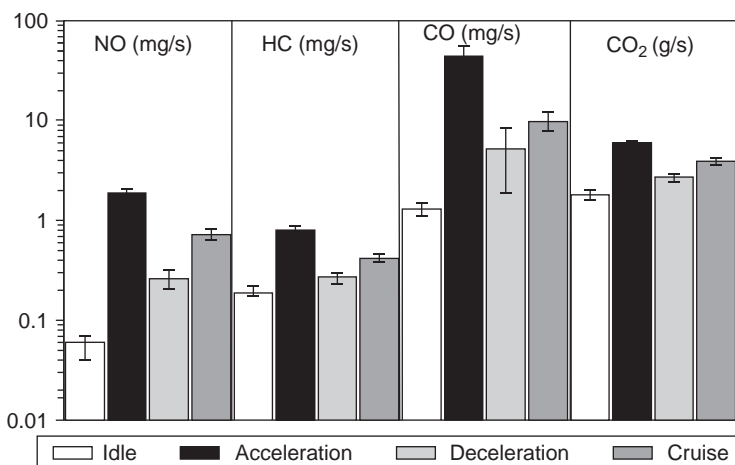
**FIGURE 3.1** Illustration of emission factors by driving mode.

TABLE 3.2 Driving Mode Effect on CO Emissions

Driving Mode	Acceleration	Deceleration	Cruise	Idle
Travel time by mode (h/year)	60	40	150	100
Corresponding EF (mg/s)	40	7	10	2
Annual emissions by mode (kg)	86.4	10.1	54	7.2

confidence intervals for four different pollutants. It is obvious that travel during the acceleration mode has a disproportionately higher emission factor for all pollutants considered, while the idle mode has the lowest emission factor. On an average, the emission factors in Figure 3.1 for CO are 48, 11, 6, and 1.2 g/s for the acceleration, cruise, deceleration, and idle modes, respectively. Therefore, a commuter trip that includes many acceleration and deceleration cycles is likely to generate higher emissions than one in which speed does not vary substantially.

An application of the driving mode approach is shown in Table 3.2, which represents driving conditions in a congested environment for an urban commuter. In this case, travel time by mode in lieu of distance is used to describe vehicle activity. The results indicate that while acceleration constituted only about 17% of the vehicle modal activity, it accounted for more than 55% of the overall annual CO emissions.

3.4.4 VSP

Three obvious deficiencies in the MODE vehicle activity descriptor involve (1) the lack of accounting for the various levels of accelerations (the mode where the highest emission rates occur), (2) the fact that the cruise mode has a fixed emission factor independent of speed, and (3) the fact that this approach does not account for the roadway infrastructure effects on power demand. These limitations can be partially overcome through the use of a *vehicle-specific power term*, or VSP. This parameter is highly correlated with emissions, and appears to form the basis for the development of the next generation emission models at the U.S. Environmental Protection Agency (Koupal et al., 2004; Frey et al., 2002). Instantaneous, second-by-second VSP values for a generic light duty vehicle is expressed as follows:

$$\text{VSP} = v \times [1.1a + 9.81 \times (\sin(a \tan(\varphi))) + 0.132] + 0.000302v^3 \quad (3.1)$$

where VSP = vehicle specific power (kW/ton), v = instantaneous vehicle speed (in m/s), a = instantaneous acceleration or deceleration rate (in m/s^2), \tan = inverse tangent function, equivalent to $\tan^{-1}(\varphi)$, φ = road grade (dimensionless fraction). It has been shown by Frey et al. (2002) that it is possible to discretize the VSP values computed from Equation (1) into a limited number of “bins,” each having a unique average emission factor, which is also significantly different from all other bins. The bin assignments for a light duty vehicle are depicted in Table 3.3.

An example of CO emission factors for a generic light-duty vehicle associated with the VSP bins defined in Table 3.3 is shown in Figure 3.2. Note that modes 1 and 2 represent deceleration models ($\text{VSP} < 0$), while mode 3 includes the idle mode ($\text{VSP} = 0$). All the remaining modes represent combinations of speed and (positive) accelerations.

One obvious difficulty in applying a modal activity model (whether the four-mode or VSP) to assess the impact of congestion management strategies on emissions is the

TABLE 3.3 VSP Modal Definitions

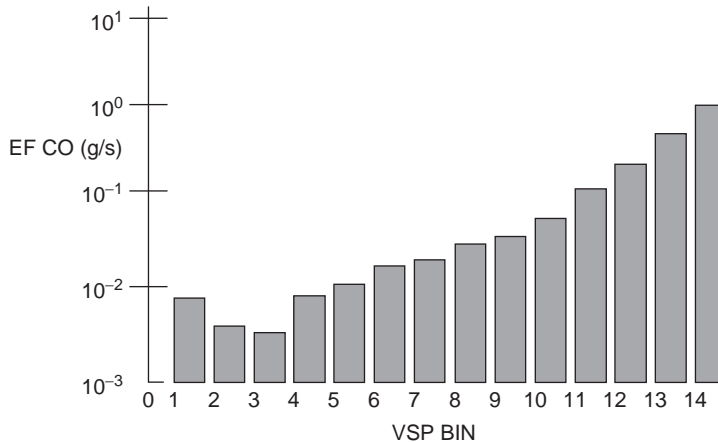
VSP Mode	Bin Range	VSP Mode	Bin Range
1	$VSP < -2$	2	$-2 \leq VSP < 0$
3	$0 \leq VSP < 1$	4	$1 \leq VSP < 4$
5	$4 \leq VSP < 7$	6	$7 \leq VSP < 10$
7	$10 \leq VSP < 13$	8	$13 \leq VSP < 16$
9	$16 \leq VSP < 19$	10	$19 \leq VSP < 23$
11	$23 \leq VSP < 28$	12	$28 \leq VSP < 33$
13	$33 \leq VSP < 39$	14	$VSP \geq 39$

Source: Frey et al. (2002).

requirement for high-resolution vehicle travel data (e.g., second-by-second), which are not typically available to transportation engineers and transit planners. On the rare occasions where a passenger car or transit vehicle fleet is equipped with an automatic vehicle location (AVL) system, speed profiles can be captured and geo-referenced to a point on the roadway, which enables the implementation of modal or VSP models to estimate instantaneous emissions in real time. In practice, however, an instrumented roadway infrastructure can at most produce measures of traffic flow and average speeds on those links that happen to be equipped with vehicle detectors at discrete points on the road. In this case, a hybrid approach that relates VSP to average speed on a roadway link offers a good compromise.

3.4.5 Hybrid VSP-S

This approach is derived from observational studies that are able to relate micro-scale events that occur on a traffic link, such as second-by-second VSP, to aggregate link performance measures, namely average traffic speed. It is based on taking concurrent measures of both parameters using portable emission measurement systems or PEMS (Frey et al., 2003). The concept is straightforward. For each range of average link speed, there appears to be a corresponding, and relatively stable, distribution of VSP values that can

**FIGURE 3.2** CO emission factors versus VSP bins.

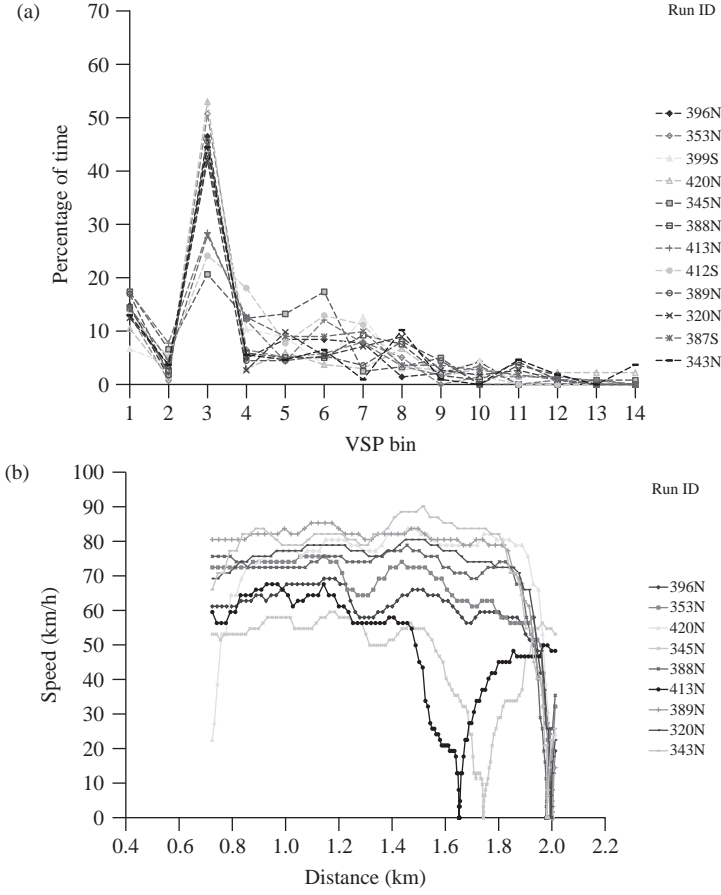


FIGURE 3.3 Speed profiles and VSP distribution (speed range 8–12 mph).

then be used to estimate emission factors (Frey et al., 2006). An example of this approach is shown in Figure 3.3. Figure 3.3a shows multiple speed profiles observed on a single roadway link. These profiles have one thing in common: the *average link speed* from all profiles varies from 8 to 12 mph. Figure 3.3b shows the corresponding travel time distribution of VSPs associated with that speed range. Remarkably, these distributions appear to be fairly uniform in shape, with a mode around VSP bin #3 (not surprising, given the low average speed on the link of 10 mph). The hybrid approach thus entails several steps, namely average link speed prediction, selection of the corresponding distribution of VSP, and application of the appropriate emission factor for each VSP bin (from Figure 2), to the link travel time. Mathematically, the emission estimation method is expressed as follows:

$$E_j = \sum_{t=1}^{T_j} a_{ijt} \times EF_i, \quad i = 1, \dots, 14 \text{ and } j = 1, 2, \dots, J \quad (3.2)$$

where E_j = emissions *per vehicle* traveling on link (j), in g/vehicle, a_{ijt} = 1 if VSP mode (i) occurs in time step (t) on link (j), zero otherwise, EF_i = emission factor associated with VSP mode (i) (e.g., Figure 3.2 for CO), T_j = average travel time on link j (computed as

link (j) length over average link travel speed). By multiplying the per-vehicle emission rate with the link traffic flow (typically on an hourly basis), the above equation produced the overall link emission over the hour for the subject link. Aggregating over all the network links will generate network-wide emissions. Further disaggregation of Equation (3.2) by vehicle type can be carried out to distinguish between emission factors for various vehicle classes (cars, buses, etc.).

3.5 INVENTORY OF TRAFFIC CONGESTION MANAGEMENT METHODS

3.5.1 Travel Demand Management (TDM)

Travel demand management involves a family of strategies that are aimed at altering the demand pattern of travelers' activities. This requires the reader to distinguish between travel demand that is expressed in *person-miles*, and *motorized vehicle-miles* of travel. An implicit assumption in the following discussion is that the demand for person-miles of travel remains unchanged in order to satisfy the individual's objectives to have access to employment and to exercise other desirable economic and social activities. Demand management therefore focuses on altering the motorized vehicle-miles of travel and associated emissions, in order to enable the accomplishment of one or more of the following objectives:

- An overall reduction in demand for vehicular travel activities (VMT)
- A spatial shift in the current vehicular travel demand to less congested facilities
- A temporal shift in the current vehicular travel demand to less congested periods of the day
- A modal shift in the current vehicular travel demand to less-polluting motorized or nonmotorized alternatives

A sampling of some commonly used TDM strategies is described next (ICF International, 2006). It should be noted that several of the strategies can affect more than one of the stated objectives.

- *Congestion and Parking Pricing*: These strategies are aimed at altering traveler behavior by increasing the out-of-pocket cost for users of the transportation system during peak demand periods. Some of the common technologies include toll roads, high occupancy toll or HOT lanes (these are physically separated lanes that provide reduced cost of access to vehicles meeting a pre-specified passenger occupancy and may provide more costly access to other vehicles not meeting the occupancy threshold), cordon pricing as exemplified by the Central London congestion pricing scheme (Litman, 2007), peak-period parking costs, or restrictions. Their intended primary effect is to produce a temporal reduction in peak period demand (Objective 3) for the high pollutant travel modes. When combined with preferential pricing policies toward the use of public transportation, these strategies could also affect a modal shift (Objective 4).
- *Flexible Work Hours*: These employer-based programs are aimed at reducing peak travel (Objective 3) by allowing employees some flexibility in the start and end hours of their workday, and in some cases compress the weekly work schedule into four working days.

- *Park-and-Ride Facilities*: Park-and-ride facilities are targeted toward inducing a modal shift (Objective 4) from private to public transportation modes (mostly bus, light rail, or heavy rail modes). In addition to providing direct access to those modes at the trip origin point, free or low-cost parking is often provided for commuters at those facilities.
- *High Occupancy Vehicle (HOV) Facilities*: HOV facilities apply to both car-pool and (normally) bus transit vehicles. The facilities are exclusive to the users of these modes and, therefore, offer an improved level of service as an incentive for the modal shift (Objective 4). It is important to note that providing easy access to and egress from those facilities is critical; otherwise, their use becomes of limited value to the traveler. This could be in the form of separate entry and exit ramps on free-ways, or by providing preferential access at the entry ramps to HOV facilities on freeways.
- *Ridesharing and Vanpool Programs*: These programs are aimed at producing a direct reduction in VMT (Objective 1) by promoting and facilitating ride sharing and (normally employer-based) vanpool programs. These may include incentives such as free parking, free fuel, and the use of HOV facilities among others. Membership in such programs would be on a subscription basis, although recent advances in computer-aided dispatching could enable real-time ride matching and vanpooling.
- *Nonmotorized Alternatives*: These programs promote nonpolluting walking and bicycling modes (Objective 4) through a variety of infrastructure improvements. Examples include dedicated bicycle lanes, wide sidewalks, grade-separated pedestrian crossings at busy intersections, and a variety of priority treatments at signalized intersections. To accommodate longer trips the provision of bicycle storage space on transit buses and rail cars, and exclusive bicycle park-and-ride facilities at train stations are often considered.
- *Transit Alternatives*: These programs are aimed primarily at shifting single occupancy vehicle (SOV) drivers to public transportation modes, which are highly effective in reducing pollutants per person-mile of travel (Objective 4). Improvements are generally implemented in four areas: (1) service, including added transit lines, shorter headways, and higher reliability; (2) amenities at bus and transit stops, including covered, and climatized shelters; (3) pricing, including reduced or fare-free rides for the general or special populations; and (4) travel information systems disseminated through a variety of media, including the internet, 511 call lines, cable television, and strategically located traveler information kiosks.
- *Land-Use Strategies*: These programs represent a significant departure from the predominant suburban land-use development patterns with their reliance on the auto mode for virtually all trip purposes, to one that is transit, pedestrian, and bicycle friendly. Known also as *smart growth*, or *neotraditional* patterns, these strategies promote mixed-use development that can negate or at least curtail the need for auto travel. By definition, such strategies are not expected to have a short- or even medium-term impact on trip making behavior, but in conjunction with improved vehicle emission technology, have the most significant potential for reducing emissions on a large scale through major reductions in auto VMT (Objective 1). The work by Johnston et al. and Rodier et al. quantifies the effect of land-use changes on vehicle emissions (Johnston and Ceerla, 1995; Rodier et al., 2002).

TABLE 3.4 Mapping TDM Strategies to VMT Changes and Time Horizon

TDM Strategy	Primary Objective 1 = Overall VMT Reduction; 2 = Spatial Shift; 3 = Temporal Shift; 4 = Modal Shift				Implementation Time Horizon 1 = Short; 2 = Med.; 3 = Long
	1	2	3	4	
Congestion and parking pricing	✓				2
Flexible work hours			✓		1
Park-and-ride facilities				✓	2
High occupancy vehicle facilities				✓	1
Ride sharing and vanpool programs				✓	1
Nonmotorized alternatives				✓	2
Transit alternatives				✓	2
Land-use strategies	✓				3
Highway information services	✓	✓	✓	✓	1

- *Highway Information Services:* These strategies fall under the umbrella of *advanced traveler information systems* (ATIS). Their impact is ubiquitous in that they can affect the spatial, temporal, modal, and number of trips that are taken by the traveler (Objectives 1–4). Yet, the effect of ATIS on VMT of travel, much less emissions and air quality, remains somewhat elusive, due to the difficulty of acquiring the appropriate empirical data. Travel information can be imparted pre-trip (at home), which could alter the traveler's departure time, route, and mode taken or forgo the trip altogether; it could be given en route via changeable message signs (CMS) or from in-vehicle navigation and routing systems, which will tend to alter the spatial pattern of the trips.

Table 3.4 provides a summary of the general impacts of the stated travel demand management strategies on VMT-based activities.

3.5.2 Traffic Operational Strategies (TOS)

A common characteristic across TDM strategies is that they all tend to target trip-making behavior *before* the trip is actually made. Traffic operational strategies, by contrast, are geared towards reducing emissions under prevailing traffic conditions. In most cases, congestion mitigation measures will also have a positive impact on emission reductions. This section, therefore, focuses on strategies that have an effect on the individual and traffic speed profiles of vehicles as they travel through the network. It has been established in Section 3.3 that significant speed changes over a trip will result in increased emissions. A simple example by Unal illustrates this point vividly (Unal, 2002). Figure 3.4 shows speed profiles for two trips that have identical average travel speeds (19 mph), using the same vehicle on the same route. Trip 2, however, appears to have much larger speed variance than trip 1. Table 3.5 depicts the pollutant emissions measured for both trips using a portable emission measurement system (PEMS). It is evident that a speed profile that is not *smooth* will generate much higher emissions. In the case presented, CO emissions for

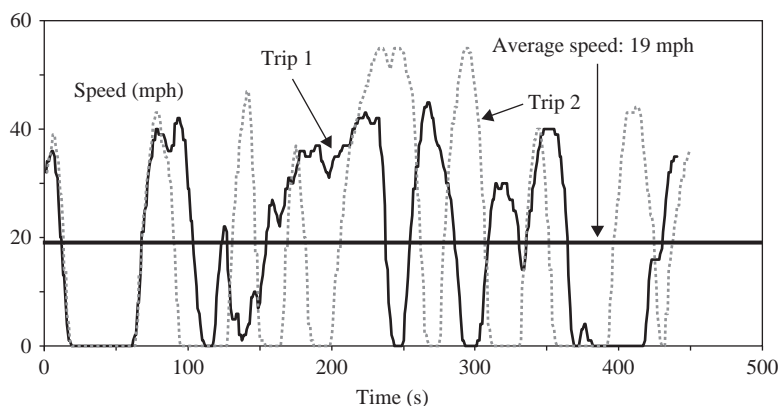


FIGURE 3.4 Speed profiles for two trips at an average speed of 19 mph.

trip 2 are three times as high as for trip 1; NO emissions for trip 2 are close to five times as high as those for trip 1. The same pattern applies to the other pollutant, although less dramatically. In the following presentation, while the focus may be on solutions that increase travel speed, it is critical to note that speed variance is the stronger indicator of emission impacts.

Traffic operational strategies are typically identified by the facility type where they are applied, namely freeways or surface streets (which includes arterials, collectors, and local roads). An important consideration of the effect of TOS is the possibility of induced traffic demand as an outcome of congestion mitigation. In other words, while TOS are intended to reduce the unit emissions by smoothing traffic flow, as evident from the preceding charts, they could have the unintended consequence of increasing the vehicle miles of travel, which could partially offset their benefits. Unfortunately, there are no standard methods to estimate the elasticity (or sensitivity) of travel demand to short TOS-type improvements, but the engineer should, nevertheless, be cognizant of such unintended consequences. Two comprehensive sources of information on the emission effects of traffic flow improvements include the USDOT Intelligent Transportation Systems Benefits Page and a National Cooperative Highway Research Project report by Dowling et al. (U.S. Department of Transportation, 2007; Dowling et al., 2005).

3.5.2.1 Freeway Operational Strategies

- **Ramp Metering:** Ramp metering has the beneficial effect of reducing traffic turbulence in the ramp entrance zone, particularly when the freeway mainline is operating close to capacity. Evidence shows that with ramp metering, average speeds increase, and delays are reduced when proper metering rates are applied (Piotrowicz and

TABLE 3.5 Measured Trip Emission Rates for Trips Shown in Figure 3.4

Trip	Pollutant Emission Rate			
	CO (mg/s)	NO (mg/s)	HC (mg/s)	CO ₂ (g/s)
1	6.1	0.41	0.91	2.3
2	19	1.9	1.2	3.1

Robinson, 1995). Metering can be implemented in an isolated fashion at selected ramps, or via a system-wide coordinated metering plan that covers entire corridors. Possible drawbacks of metering include the possibility of long queues at ramps that may spill back onto the surface street, and inducing large acceleration rates for entering traffic (see Figure 3.1).

- *Incident Management*: Incident management systems mitigate the effects of nonrecurring congestion due to accidents, stalled vehicles, unanticipated road work, and so on. Evidence shows that between 52 and 58% of all congestion delay may be due to nonrecurring events, or incidents (ICF International, 2006). By reducing the response and clearance times for such incidents, delays and travel times can be markedly reduced. There are also safety benefits that can accrue from a reduction in secondary (mostly rear-end) accidents that occur as a result of the large speed differentials between approaching traffic and traffic within the vicinity of the incident.
- *Bottleneck Mitigation*: Freeway bottlenecks are leading causes of recurring traffic congestion on urban freeways, resulting in local stop-and-go operations in their vicinity. They are typically activated when traffic demand surges over capacity downstream of a congested on-ramp or when geometric restrictions reduce the normal freeway capacity, as in the case of a lane drop or a weaving section. Besides major (and costly) capacity additions, there are low-cost solutions that may mitigate their effect, including short auxiliary lanes, the use of hard shoulders during the peak hours, and implementation of truck restrictions. A unique strategy that enables the use of a wide shoulder as a narrow lane for queue storage during peak flow periods only (called the plus-lane) has been applied in Utrecht, The Netherlands, and is illustrated in Figure 3.5 (Roupail and Chae, 2003).
- *Lane Control*: Lane control strategies are more commonly used in European cities to control speed and restrict lane use to specific time periods or vehicle classes. From an emissions control perspective, lane control promotes gradual speed reduction ahead of a congested region through the use of a series of advisory or regulatory speed limit signals erected on overhead gantries. This control will tend to smooth the traffic speed profile and minimize high acceleration and deceleration events. An example application of speed control ahead of a lane drop in Stockholm, Sweden, is illustrated in Figure 3.6.

In addition, many urban areas in the United States use a combination of reversible and HOV lanes to provide added capacity in the peak flow direction to meet the prevailing directional demand. More recently, entire directional freeway facilities have operated in the reverse direction to evacuate traffic from hurricane-threatened coastal areas (Williams et al., 2007).

- *High-Speed Tolling*: On toll facilities operating with manual toll collection, approaching vehicles will experience significant speed change cycles both upstream (deceleration) and downstream (acceleration) of the toll booth. The zone just downstream of the toll plaza, where large accelerations are prevalent, is considered to be a high CO emission hot spot (Coelho et al., 2005). The use of electronic, high-speed tolling negates the requirement for a vehicle to come to a full stop and then accelerate to freeway speeds, and, therefore, reduces the number and severity of speed change cycles in the vicinity of toll facilities. Such systems use electronic tag readers on the moving vehicle to record its arrival and to charge the user's account.



FIGURE 3.5 Bottleneck mitigation using plus-lane concept.

3.5.2.2 Surface Street Operational Strategies

- Signal Timing and Coordination:** Improved traffic signal timing and coordination is considered to be one of the most cost-effective measures for congestion mitigation and emission reductions on surface streets. According to the Institute for Transportation Engineers, there are over 300,000 traffic signals in the United States, with the majority needing upgrading of equipment or adjusting of the timing or coordination, while servicing nearly two-thirds of the vehicle-miles traveled (Institute for Transportation Engineers, 2007). Comprehensive signal retiming programs have documented benefits of a 7–13% reduction in overall travel time, a 15–37% reduction in delays, and a 6–9% fuel savings. Signal coordination, in particular, has a significant impact on reducing vehicular stop-and-go cycles. Section 6.1 discusses the results of a study on the effect of improved timing and coordination on real-world measured vehicle emissions.



FIGURE 3.6 Lane and speed control system (Courtesy Rouphail, 2001).

- *Tidal or Reversible Flow Lanes*: Similar to reversible lanes on freeways, tidal flow lanes on arterials are used to provide flexible capacity in one direction to accommodate peak flow traffic. These are also used to cater for special event traffic that requires the provision of capacity in different directions before, during, and after the event (sporting events, concerts, major evacuation, etc.).
- *Roundabouts*: This form of intersection traffic control is becoming very popular in the United States, with over 700 modern roundabouts installed (Kittelson and Associates Roundabout Database, 2007). From a vehicle emission perspective, roundabouts, especially those that are not highly congested, promote a smoother speed profile in the vicinity of the intersection, by not requiring traffic to stop at the circle, and by reducing overall speeds on the approach and in the circle. Evidence, from both empirical and modeling studies, indicates positive environmental effects. As an example, a recent study in Sweden shows that at a small roundabout that replaced a signalized intersection, CO emissions decreased by 29%, NO_x emissions by 21%, and fuel consumption by 28% (Varhelyi, 2002). At roundabouts, replacing yield-controlled intersections, CO emissions increased on average by 4%, NO_x emissions by 6%, and fuel consumption by 3%. An area that requires further investigation is the effect of exiting traffic on emissions, where drivers tend to accelerate from the circulating lane design speed to that of the open road ahead.
- *Continuous Flow Intersections*: Continuous flow intersections (CFIs) represents an innovative intersection design scheme that physically separates and services left-turn movements upstream of the main intersection, which are then rerouted to a lane to the left of the opposing flow. This design enables the left turns to merge without stopping into the cross-street traffic. According to recent studies, there are only two such installations in the United States, in New York and Maryland, with the majority of the installations in Mexico. Similar to other TOS strategies, CFIs have been shown to reduce delays and speed change cycles at intersections and, thus, would result in lower fuel and emission use. An illustration of an operational CFI in Mexico City is shown in Figure 3.7.
- *Bus Signal Priority*: This strategy entails the provision of traffic signal control priority for buses on arterials. This can be accomplished in several ways, depending on the signal indication when the bus arrives at the intersection: an early green start, an extension of the current green (if needed), or an off-sequence special green phase. The purpose of these strategies is to reduce transit travel time, improve its reliability, and, in the long term, help increase transit ridership (essentially a TDM strategy). The emissions consequences of implementing bus priority schemes are unclear and very much depend on the site conditions. Areas of concerns are the effect of the bus signal priority or preemption on disrupting the existing signal coordination plan, and on side street traffic that may be penalized as a result of the additional main street green times. Advanced algorithms that account for schedule delay (using automated vehicle location systems) and real-time passenger counts in deciding whether preemption is desirable may alleviate some of these concerns.

A comparison of TDM and TOS strategies indicates that the first has a diverse set of objectives, ranging from reducing travel demand, to shifting it in time, in space, or to other environmentally friendly modes. Traffic operational strategies, by contrast, tend to



FIGURE 3.7 Continuous flow intersection operation (Courtesy Francisco Miero).

hone on the single objective of smoothing traffic flow by reducing stop-and-go cycles, maintaining near-uniform speed profiles, and minimizing the episodes of high emissions associated with large acceleration and deceleration events.

3.6 ASSESSING EMISSION IMPACTS OF TRAFFIC CONGESTION MANAGEMENT

There are numerous methods available to transportation and environmental analysts to evaluate the effectiveness of traffic congestion management methods. These can be classified into two general categories: on-road direct measurements and emission estimation models. The latter can be further subdivided into emission factors and traffic models. Some methods combine elements of both types of models.

3.6.1 Direct Measurement Methods

This section covers two direct measurement methods for vehicle emissions.

3.6.1.1 Portable Emission Measurement Systems (PEMS)

Vehicle emissions can be directly measured under various traffic conditions using PEMS (Frey et al., 2003). This method is rapidly gaining wide acceptance by the user community and regulatory agencies such as the U.S. EPA. These devices are installed in a vehicle in a matter of minutes and enable the analyst to measure tailpipe emissions of several pollutants, including CO, NO, HC, CO₂, and O₂, in addition to fuel consumption, on a second-by-second basis under real-world traffic conditions. PEMS is connected and synchronized with the vehicle engine data scanner, which includes, among other variables, a second-by-second speed profile. It draws AC power from the cigarette lighter outlet. Figure 3.8 depicts a PEMS device, the engine data scanner, its positioning in the vehicle, and the connection to the tailpipe sampling probe.

An example assessment study using PEMS was reported by Unal et al. (2003). In this study, the effectiveness of a traffic signal timing and coordination scheme for an arterial corridor was carried out. The study involved multiple runs of two PEMS-instrumented

vehicles using the same drivers and data collection periods before and after the signal system was coordinated. Figure 3.9 shows speed and CO traces for a representative single run in the before and after case.

It is evident that this strategy was effective in reducing the corridor travel time, as well as smoothing the speed profile by reducing the number of stops and CO emissions. Summary statistics of traffic and emission changes after the signals were retimed are reported in Table 3.6 for both directions of traffic, two peak periods, and two vehicles. The results confirm the cost-effectiveness of signal control strategies as an emission mitigation measure.

The same referenced study also compared emissions for uncongested and congested traffic conditions on another corridor, by contrasting trip emissions in the peak and off-peak directions (Unal et al., 2003). The results were even more dramatic, with a decrease in trip emissions in the range of 40–60%, depending on the time period and the pollutant considered.



FIGURE 3.8 Example of a PEMS and its installation in a vehicle.



FIGURE 3.8 (Continued).

3.6.1.2 Remote Sensing Device (RSD)

Remote sensing technology utilizes infrared spectroscopy to flag high-emitting vehicles as they pass a point on the highway. It is motivated by statistics that show that only about 10% of the on-road vehicles are responsible for over 50% of the CO, NO_x, and HC emissions. The device consists of an infrared source that emits a light beam across the highway just before and during the vehicle crossing the measurement point, as shown in the schematic Figure 3.10 (EPA Fact Sheet OMS-15, 1993). The system determines the ratio of CO/CO₂ in the vehicle plume, and if the level of carbon monoxide (CO) is above a certain threshold, it is tagged as a high-emitting vehicle, typically using video detection technology. The device could also be used to assess the emission consequences of traffic control strategies. However, it has many limitations with regards to lane (one lane at a time) and spatial coverage (single point on the road), and as such could miss many of the high-

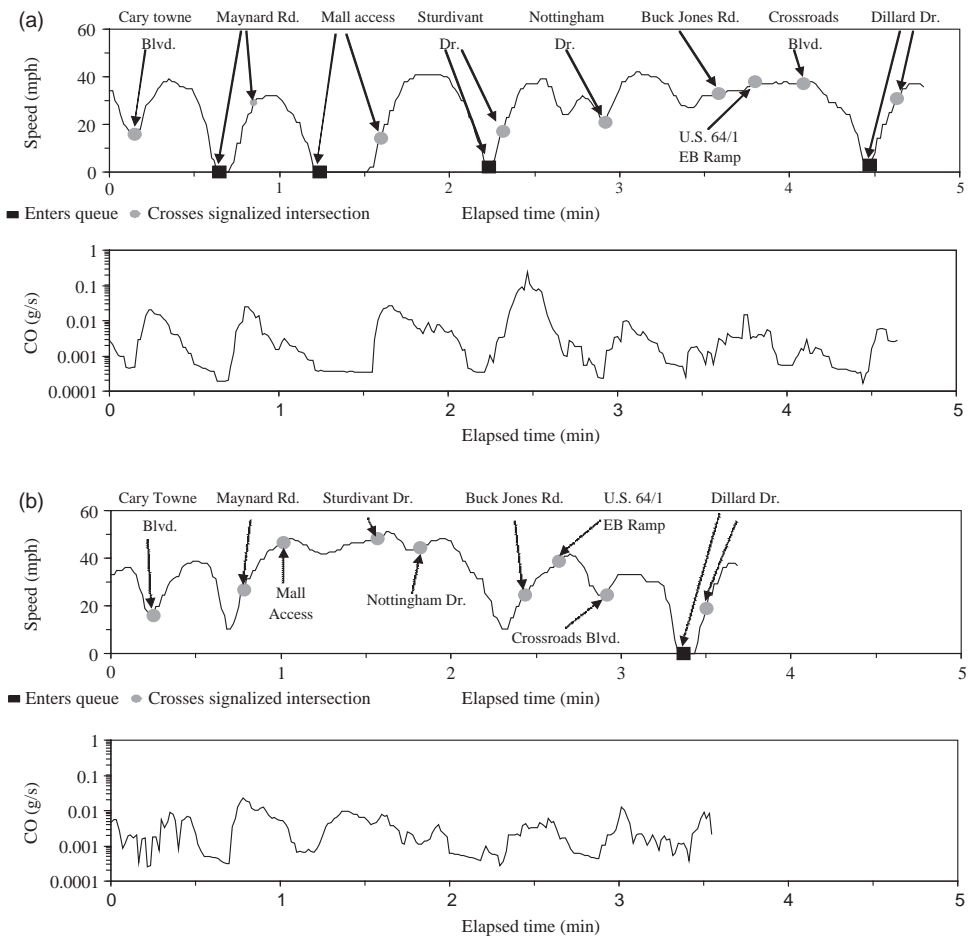


FIGURE 3.9 Representative speed profiles before and after signal coordination. *Source:* Unal et al. (2003).

TABLE 3.6 Summary Traffic Performance and Emission Changes After Implementing a Signal System Coordination Project (Using PEMS)

Time Period	Ford Taurus				Oldsmobile Cutlass			
	Morning		Afternoon		Morning		Afternoon	
Direction	N	S	N	S	N	S	N	S
Trip duration (%)	-14	-24	-23	+0.3	-16	-17	-21	-0.9
Ave. speed (%)	+14	+32	+29	-1.8	+18	+20	+29	-1.8
Control delay (%)	-40	-63	-55	-4.6	-38	-50	-56	+8.5
Total stops (%)	-30	-60	-29	-2.3	-29	-46	-29	-11
HC emissions (%)	-12	-18	-11	+1	-12	-13	-12	-1
NO emissions (%)	-8	-12	-1	+1	-13	-14	-19	-1
CO emissions (%)	-12	-19	-5	+1	-4	-9	-1	-1

Source: EPA Fact Sheet OMS-15 (1993).

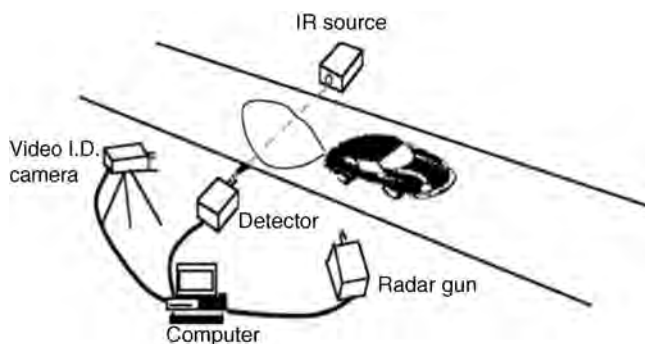


FIGURE 3.10 Schematic of operation of an emission RSD.

acceleration events that are concomitant with episodes of high emissions. Unal et al. (1999) reported on the concurrent use of RSD and area-wide vehicle detection for the simultaneous analysis of traffic performance and emissions.

3.6.2 Mobile Emission Estimation Models

This section covers three models for estimating mobile emissions.

3.6.2.1 Mobile

By far, the most widely used family of mobile emission factor models is EPA's MOBILE series, of which MOBILE6.2 is the latest release (EPA User's Guide to MOBILE6.1 and MOBILE6.2, 2003). These models represent the state of the practice for developing emission inventories and control strategies at the regional or state levels, as part of the required state implementation plans (SIP) under the 1990 Federal Clean Air Act. MOBILE emission factor values represent *national average* conditions, in terms of vehicle fleet (classified into 28 vehicle categories), VMT distribution among roadway types, standardized driving cycles on these roadways, and speed distributions among roadway types. These averages can be modified to fit local conditions through a series of adjustments to the base emission factors on the basis of modified user input to better reflect local conditions. From a congestion management perspective, MOBILE inputs that are relevant include the types of roadways (freeways, ramps, arterials, and local streets), the VMT distribution across these facilities (national defaults are 34, 3, 50, and 13% for the four facilities), and the average speed on those facilities (national default *daily* speeds are 36.5, 34.6, 31.2, and 12.9 mph, respectively). The pollutants of interest include HC, CO, NO_x, and CO₂. MOBILE can represent any vehicle fleet from 1952 to 2050, making assumptions on the distribution of vehicle age in a given calendar year (CY). Sample results of the sensitivity of MOBILE6 freeway and arterial emission factors to average speed are shown in Figure 3.11.

The data shown are for the projected fleet for CY2008, and include light-duty vehicles *only*. Several patterns emerge from those data. First, MOBILE emission factors for freeways and arterials are not significantly different from each other when controlling for the pollutant type. This may be a result that the national default speed distributions between these two facilities are very similar as indicated earlier. Second, CO emissions factors

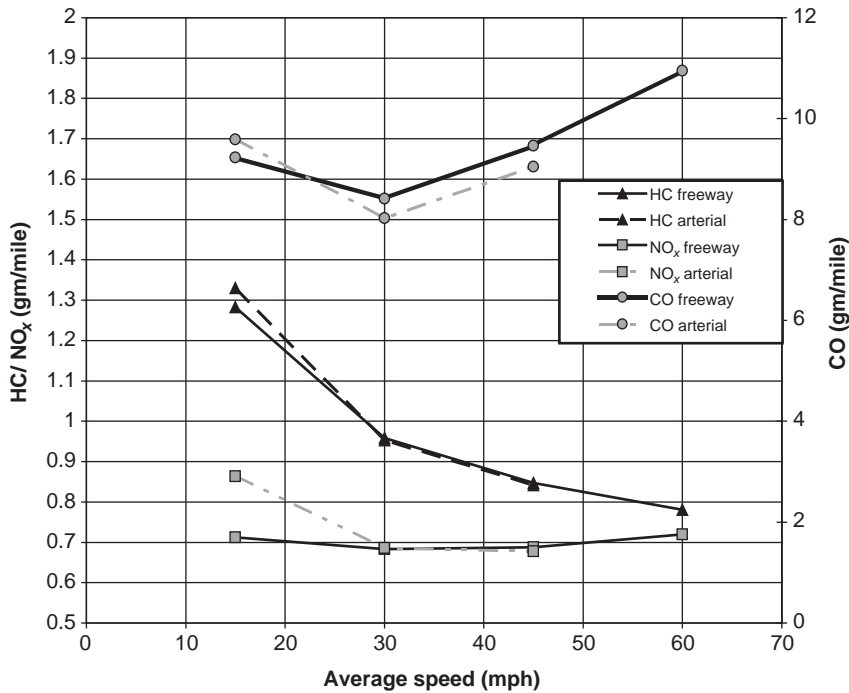


FIGURE 3.11 MOBILE6 emission factors' sensitivity to average speed on freeway and major arterials; in g/mile of travel.

exhibit a nonlinear relationship with speed. They tend to be lowest in a speed range of 25–35 mph, and increase below or above that range. However, HC and NO_x emissions factors decrease monotonically with speed for the low average speed range (15–30 mph) and then remain essentially unchanged for NO_x while HC emission rates continue to drop as speed increases. This indicates that stop-and-go traffic patterns are likely to be associated with high emissions of these two pollutants.

An important caution in the application of MOBILE6 factors for the assessment of traffic control measures is the need to acknowledge the sensitivity of travel demand (i.e., VMT) to level of service (LOS). The user must account for the possibility that improved traffic performance could generate additional (new) trips that could partially offset the benefits gained from reducing the emissions factors.

There are many more applications of MOBILE than can be covered in a comprehensive manner in this chapter. The reader is strongly encouraged to consult the MOBILE6 user guide and associated documentation on the EPA Office of Transportation and Air Quality, OTAQ (2007). It is noteworthy to mention that EPA is currently developing a new emission estimation framework called MOVES (MOTOR Vehicles Emission Simulator) that will rely on more refined data, based partly on the VSP approach explained earlier. MOVES will eventually replace MOBILE6. The current release MOVES2004 provides estimates for fuel consumption and greenhouse gas emissions for on-road vehicles only (EPA, 2005). Future releases will include the estimation of HC, CO, and NO_x, and for on-road vehicles (Koupal et al., 2004).

3.6.2.2 MODAL, VSP, and VSP-S Emission Factor Models

These models have been discussed in Section 3.4 and are mostly derived from microscale speed profile data gathered in the lab using engine dynamometers, or in the field using PEMS.

3.6.2.3 COMMUTER2.0 Model

This model, developed for EPA, estimates the travel and emission impacts of many of the transportation demand management strategies that were described in Section 5.1 (EPA, 2005). It is designed to focus on small programs and transportation control measures that do not have large regional travel impacts, in which case traditional regional travel demand models should be used. In COMMUTER2.0 travel impacts are modeled using a logit model formulation that produces modal and temporal shifts based on the value of a utility function of modal travel times, transfer times, and costs. By specifying a baseline scenario of mode choice, trip lengths, and current VMT, COMMUTER2.0 will estimate *changes* in VMT's and associated emissions, the latter mostly derived from the MOBILE6 model. Figure 3.12 shows a sample scenario output of this model, based on the input and traffic measures provided by the users. Additional details can be found in the model user guide.

3.6.3 Transportation Models

This section discusses the use of transportation models to assess the emission impacts of transportation control measures. For the most part, these models focus on estimating the travel or traffic impacts of regional and local transportation improvements, and, as such, the model selection depends on the nature and geographic impacts of the application. Models can be classified as travel demand models, macroscopic traffic operations models, and microscopic simulation models.

3.6.3.1 Travel Demand Models

This class of models is typically used for estimating long-term transportation system needs (15–20 years). Their focus is on predicting required link-level characteristics (type, capacity, etc.) that are needed to accommodate future traffic demands. Typically (with some exceptions), this class of models does have the capability of predicting emissions. Instead, a post-processing step is executed to use the travel model output (e.g., VMT by roadway class, vehicle type, and speed) as input into MOBILE6. Because MOBILE represents national average emission factors, it is important for the user to *not* rely on or use link-level emission estimates, since these are highly simplified. The intent is to obtain a fairly coarse estimate of the emission inventory on a regional scale. Commonly used travel demand models include TRANPLAN, EMME2, TransCad, and TRANUS (TRANPLAN homepage, 2007; TransCad homepage, 2005; Modelistica, 2007).¹

3.6.3.2 Macroscopic Traffic Operations Models

This class of models considers the effect of local or corridor-wide improvements (such as ramp metering, signalization, HOV lanes, and so on) on traffic performance with the implicit assumption that vehicle activities (i.e., VMT) remain fixed before and after the

¹ EMME2 Homepage, INRO Solutions, <http://www.inro.ca/en/products/emme2/index.php>.

COMMUTER MODEL RESULTS

SCENARIO INFORMATION

Description	Example Scenario v2.0
Scenario Filename	Example Scenario v20.vme
Emission Factor File	
Performing Agency	Cambridge Systematics, Inc.
Analyst	C. Porter
Metropolitan Area	Boston
Area Size	1- Large (over 2 million)
Analysis Scope	1 -Area-Wide (e.g., MSA, county)
Analysis Area/Site	Cambridge
Total Employment	100,000

PROGRAMS EVALUATED

Site Walk Access Improvements

Transit Service Improvements

X

Financial Incentives

Employer Support Programs

X

Alternative Work Schedules

User-Supplied Final Mode Shares

MODE SHARE IMPACTS

Mode	Baseline	Final	%Change
Drive Alone	78.2%	75.4%	-2.9%
Carpool	12.1%	11.7%	-0.4%
Vanpool	0.5%	0.5%	-0.0%
Transit	4.9%	6.3%	+1.4%
Bicycle	0.4%	0.4%	-0.0%
Pedestrian	3.0%	3.0%	-0.0%
Other	0.8%	0.8%	-0.0%
No Trip	-	2.0%	+2.0%
Total	100.0%	100.0%	-

Shifted from Peak to Off-Peak

0.0%

TRAVEL IMPACTS (relative to affected employment)

Quantity	Peak	Off-Peak	Total
Baseline VMT	1,289,745	810,817	2,100,561
Final VMT	1,243,026	781,447	2,024,473
VMT Reduction	46,718	29,370	76,089
% VMT Reduction	3.6%	3.6%	3.6%
Baseline Trips	102,780	64,614	167,394
Final Trips	98,987	62,230	161,217
Trip Reduction	3,793	2,384	6,177
%Trip Reduction	3.7%	3.7%	3.7%

EMISSION REDUCTIONS (positive values are decreases)

lbs/day:

Pollutant	Peak	Off-Peak	Total
HC	21.60	12.94	34.53
CO	337.14	217.22	554.36
NO _x	22.20	13.92	36.12
PM2.5	1.16	0.73	1.90
Toxics			
Acetaldehyde	0.056	0.034	0.091
Acrolein	0.007	0.005	0.012
Benzene	0.605	0.369	0.974
1, 3-Butadiene	0.081	0.050	0.131
Formaldehyde	0.185	0.113	0.299
MTBE	1.195	0.713	1.908
CO ₂	47,257	29,709	76,966

tons/day:

Pollutant	Peak	Off-Peak	Total
HC	0.011	0.006	0.017
CO	0.169	0.109	0.277
NO _x	0.011	0.007	0.018
CO ₂ (metric tons)	21.4	13.5	34.9

GASOLINE CONSUMPTION AND COST SAVINGS

Reduction in gasoline consumption (gallons/day)	3,934
Gasoline cost savings (\$/day)	\$8,852

FIGURE 3.12 Example of travel and emission impacts derived from a COMMUTER2.0 model run.

improvements. Many of these models have also incorporated emission prediction algorithms within the models, although it is often unclear what the assumed vehicle fleet is and how the emission rates were calibrated. Some models, such as Synchro, use a fuel-based approach that relates HC, CO, and NO emissions to fuel consumption (per hour or per mile) (Husch, 1998). Others, most notably aaSIDRA and SIGNAL200, incorporate four-mode (acceleration, deceleration, cruise, and idle) and three-mode (acceleration, idle, and deceleration) modal emission models in their software, respectively (Akcelik and Beskley, 2003; Strong, 2007). The U.S. Highway Capacity Manual, the most widely

used macroscopic traffic operational model in the United States (and possibly worldwide), has no provisions for emission estimation (Transportation Research Board, 2000). This has led users of the HCM to rely on other macroscopic or microscopic models to meet those needs.

3.6.3.3 Microscopic Traffic Simulation Models

This class of models traces individual vehicle movements over a traffic network, with vehicle position and speeds updated at one second or subsecond resolutions. The advance logic is based on driver behavioral models, such as car following, lane changing, and gap acceptance. Driver level of aggressiveness, which could have a significant impact on acceleration rates and emissions, is also considered, although rarely calibrated from empirical observations. Because these models can produce individual speed profiles their results can be easily integrated with microscale emission rate models such the MODE or VSP models described earlier to produce second-by-second emissions. Often, this integration is built in the simulation, as in the CORSIM model that uses a speed-acceleration look-up table (by vehicle type) to calculate emissions based on unpublished dynamometer testing at the Oak Ridge National Laboratory (Federal Highway Administration, 1997). A similar approach is adopted in the INTEGRATION model, the VISSIM model, and the PARAMICS model (M. Van Aerde and Transportation Systems Group, 1995; Paramics, 2007).²

It is often advisable to decouple the simulated speed profiles from the default emission estimation methods embedded in the simulators. This is because the emissions calibration data could be either out of date or not representative of the U.S. vehicle fleet. An example of representative speed profiles in the VISSIM model based on a study that compared traffic operations on an urban arterial using a series of signals (baseline) and roundabouts (proposed) is shown in Figure 3.13 (Rouphail and Chae, 2003). These vehicles were selected as their travel time coincided with the average travel time. While both profiles show considerable speed variations, there were more stops in the signalized arterial case. By applying the VSP approach described in Section 3.4, VSP values, VSP bins, and associated emission rates were estimated. Table 3.7 summarizes the emission effects of replacing the signalized intersection system with a series of roundabouts. Multiplying

TABLE 3.7 Unit Emissions for Traffic Signal and Roundabout Control

Direction and Type of Control		NO (g/veh.)	HC (g/veh.)	CO ₂ (g/veh.)	CO (g/veh.)
EB	Signal	0.65	0.29	733.06	6.86
	Roundabout	0.59 (−9.4%)	0.23 (−19.5%)	693.38 (−5.4%)	4.71 (−31.3%)
WB	Signal	0.63	0.24	630.22	7.65
	Roundabout	0.61 (−2.3%)	0.28 (+15.4%)	816.45 (+29.6%)	4.77 (−37.7%)
Overall	Signal	1.28	0.53	1363.29	14.51
	Roundabout	1.20 (−0.9%)	0.51 (−3.5%)	1509.84 (10.7%)	9.48 (−34.7%)

Source: Rouphail and Chae (2003).

²PTV, *VISSIM User's Manual*, Release 4.1.0, PTV-AG, Karlsruhe, Germany.

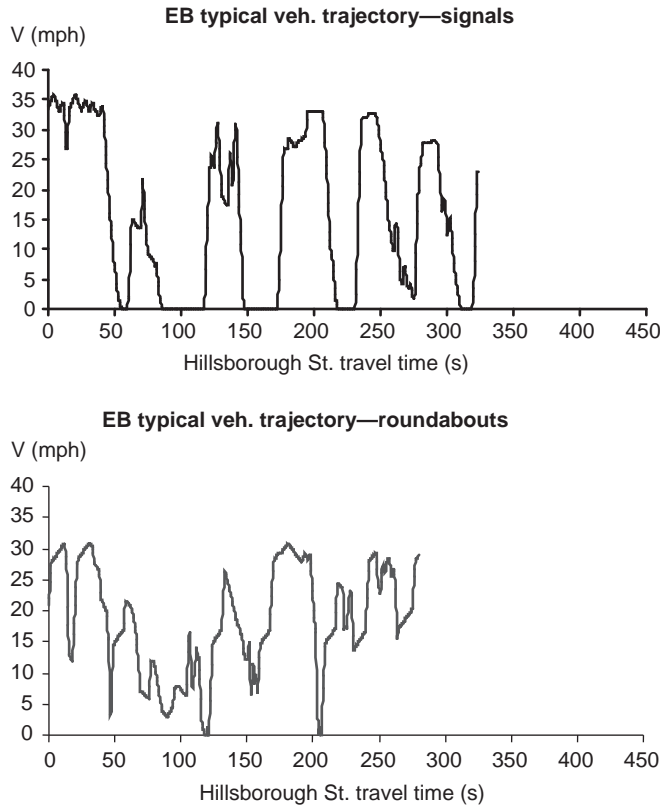


FIGURE 3.13 Representative VISSIM-generated speed profiles. *Source:* Rouphail and Chae (2003).

these values by the corresponding VMT in each case gives the overall emission impacts (Rouphail and Chae, 2003).

3.7 SUMMARY

This chapter presents a survey of the impacts of traffic congestion management strategies on vehicular emissions and, by extension, on air quality. It begins with the premise that mobile source emissions are a significant contributor to the national emission inventory for some pollutants and that congestion mitigation strategies that focus on both transportation demand and supply sides can significantly reduce the harmful effects of airborne pollutants. The chapter also discusses tools that are available to the engineer and analyst to measure, estimate, and predict emissions in response to changes in supply or demand-oriented transportation strategies. Although the literature contains a wealth of information on such effects, there is no general consensus in the transportation and environmental community on which tools are appropriate for what types of analysis, especially those at the corridor and local intersection levels. The author hopes that the material presented herein, although by necessity much abbreviated, will be of assistance to the users in identifying and evaluating the available tools

and in understanding the tools' strengths and limitations. Regardless of the method selected, the analyst must be cognizant of and account for the *long-term effects* of traffic control measures on overall emissions, particularly the impact of improved traffic flow conditions on latent demand and modal choice effects. Emerging EPA emission estimation paradigms exemplified by the MOVES model will, in the future, enable the analyst to carry out multi-scale level analyses that are sensitive to regional as well as local transportation improvements. Finally, whenever possible it is critical that local travel behavior and emission rates be used, especially if the analysis is in response to regulatory requirements such as state implementation plans (SIPs) mandated under the 1990 Clean Air Act.

ACKNOWLEDGMENTS

The author is grateful to many colleagues and graduate students whose work has been cited in this document. In particular, I am indebted to my colleague Chris Frey, professor of civil and environmental engineering at NC State University, who first raised my interest in the topic of transportation and air quality and whose continuous contributions in this field have been noted nationally and internationally. I would also like to thank Dr. Billy Williams, assistant professor of civil engineering, for his insightful comments and suggestions on a previous draft, and Ms. Katie McDermott, communications manager at ITRE, for being exceedingly thorough as my first-round editor. Many thanks to several former and current NC State students whose research was cited in this document: Alper Unal, Kosok Chae, James Colyar, and Haibo Zhai. Last but not least, I am grateful to my wife, Dr. Maria Roupail, who persevered while this manuscript was being written, for her incisive editing and advice, without which this chapter could not have been completed.

REFERENCES

- Akcelik A, Beskley M. "Operating Costs, Fuel Consumption and Emissions in aaSIDRA and aaMotion." Presentation at the 25th Conference of Australian Institutes for Transportation Research (CAITR), U. of South Australia, December 3–5, 2003. Available in pdf format, [http://www.akcelik.com.au/documents/AKCELIK_COSTModels/CAITR%202003\)v2.pdf](http://www.akcelik.com.au/documents/AKCELIK_COSTModels/CAITR%202003)v2.pdf).
- Coelho M, Farias T, Roupail N. Measuring and modeling emission effects for toll facilities, Transportation Research Record 1941. *Journal of the Transportation Research Board* 2005;136–144.
- Dowling R, Ireson R, Skabardonis A, Gillen D, Stopher P. Predicting Air Quality Effects of Traffic Flow Improvements, Final Report and User Guide, National Cooperative Highway Research Program Report 535, TRB Washington, DC. 241 p., 2005.
- EMME2 Homepage, INRO Solutions, <http://www.inro.ca/en/products/emme2/index.php>. 2012a
- EPA Fact Sheet OMS-15, "Remote Sensing: A Supplemental Tool for Vehicle Emission Control," Report No. EPA-420-F-92-017, August 1993.
- EPA Office of Transportation and Air Quality (OTAQ), "Mobile Model (on road vehicle)," <http://www.epa.gov/otaq/mobile.htm>. Accessed on February 20, 2007.
- EPA User's Guide to MOBILE6.1 and MOBILE6.2, "Mobile Source Emission Factor Model," Report No. EPA-420-R-030-010, August 2003, 262 p.

- EPA, "A Roadmap to MOVES2004," Report No. EPA 420-S-05-002, Ann Arbor Michigan, March 2005.
- EPA, "COMMUTER Model v2.0 User Manual," Report No. EPA420-B-05017, October 2005, 59 p.
- EPA, "8-hour Ozone Data Summary," <http://www.epa.gov/oar/oaqps/greenbk/gnsum.html>. Accessed December 5, 2006.
- Frey HC, Unal A, Roupail N, Colyar JD, On-road measurements of vehicle tailpipe emissions using a portable instrument. *Journal of Air & Waste Management Association* 2003;53: 992–1002.
- Federal Highway Administration, *CORSIM User Manual*. Mc Lean, Virginia: U.S. DOT Office of Traffic Safety and Operations; 1997.
- Frey H, Roupail N, Zhai H. Speed and facility-specific emission estimates for on-road light-duty vehicles on the basis of real-world speed profiles, Transportation Research Record No. 1987. *Journal of the Transportation Research Board* 2006; 128–137.
- Frey HC, Roupail NM, Unal A, Colyar JD. "Emission Reductions through Better Traffic Management," Report No. FHWA/NC/2002-001, December 2001, 368 p.
- Frey HC, Unal A, Chen J, Li S, Xuan X. "Methodology for Developing Modal Emission Rates for EPA's Multi-scale Motor Vehicle & Equipment Emission System," Report No. EPA420-R-02-027, August, 2002.
- Husch D. *Synchro 3.2 User Guide*. Berkeley, CA: Trafficware; 1998.
- ICF International, "Multi-Pollutant Emissions Benefits of Transportation Strategies," Final Report Prepared for the USDOT-Federal Highway Administration, Report No. FHWA-HEP-07-004, November 14, 2006.
- Institute for Transportation Engineers, "ITE Mega Issues, Signal Timing," <http://www.ite.org/signal/index.asp>. Accessed on February 19, 2007.
- Johnston R, Ceerla R. Land use and transportation alternatives, In: Sperling D, Shaheen S, editors. *Transportation and Energy: Strategies for a Sustainable Transportation System*. Washington, DC: ACEEE; 1995.
- Kittelson and Associates Roundabout Database, <http://roundabouts.kittelson.com/InvRoundabout.asp>. Accessed February 19, 2007.
- Koupal J, Nam E, Giannelli B, Bailey C. "The MOVES Approach to Modal Emission Modeling," Presented at the CRC On-Road Vehicle Emissions Workshop, San Diego, CA, March 29, 2004.
- Litman T. "London Congestion Pricing: Implications for other Cities", Victoria Transport Policy Institute, <http://www.vtpi.org/london.pdf>. Accessed February 19, 2007.
- M. Van Aerde and Transportation Systems Group, *INTEGRATION User's Guide. Vol. 1. Fundamental Model Features*. Ontario, Canada: Queen's University; 1995.
- Modelistica, "Complete Mathematical Description of the TRANUS Model," <http://www.modelistica.com>. Accessed on February 20, 2007.
- Paramics, www.paramics-online.com. Accessed February 19, 2007.
- Piotrowicz G, Robinson JR. "Ramp Metering Status in North America," Report No. DOT-T-95-17, FHWA, U.S. Department of Transportation, Washington, DC, 1995.
- PTV, *VISSIM User's Manual*, Release 4.1.0, PTV-AG, Karlsruhe, Germany. 2012b.
- Roupail N, Chae K. Comparison of Traffic Operations with Traffic Signals vs. Roundabouts on Hillsborough Street, ITRE Technical Report submitted to The North Carolina Department of Transportation, April 2003, 18 pp.
- Rodier C, Johnston R, Abraham J. Heuristic policy analysis of regional land use, transit, and travel pricing scenarios using two urban models. *Transportation Research Part D Transport and the Environment* 2002; 7(4):243–254.

- Schrank D, Lomax, T. *The 2005 Urban Mobility Report.*, May Texas: Texas Transportation Institute, Texas A&M University; 2005.
- Strong D. *SIGNAL2000 Tutorial/Reference Manual*, Strong Concepts; 2007.
- TranPlan homepage, <http://www.citilabs.com/tranplan>. Accessed February 19, 2007.
- TransCad homepage, Caliper Corporation, Newton, MA, <http://www.caliper.com/tcovu.htm>. Accessed July 13, 2005.
- Transportation Research Board, *Highway Capacity Manual*. Washington, DC: U.S. Government Printing Office; 2000.
- U.S., Department of Transportation, ITS Benefits Page, <http://www.itsbenefits.its.dot.gov/its/benecost.nsf/ByLink/BenefitsHome>. Accessed on February 19, 2007.
- Unal A, Dalton RH, Frey HC, Roupail NM. "Simultaneous Measurement of On-Road Vehicle Emissions and Traffic Flow Using Remote Sensing and an Area-Wide Detector," Paper No. 99-712, Proceedings of the 92nd Annual Meeting, St. Louis, June 20-24, 1999.
- Unal A. On-board Measurement and Analysis of on Road Vehicle Emissions, Ph.D. Dissertation, North Carolina State University, Raleigh, NC, July 23, 2002.
- U.S. Department of Transportation, Bureau of Transportation Statistics, *National Transportation Statistics 2005*. Washington, DC: Government Printing Office; 2005.
- U.S. Department of Transportation, Bureau of Transportation Statistics. *National Transportation Statistics 2004*. Washington, DC: Government Printing Office; 2005.
- Unal A, Roupail N, Frey HC. Effect of arterial signalization and level of service on vehicle emissions, Transportation Research Record 1842. *Journal of the Transportation Research Board* 2003; 47-56.
- Varhelyi A. The effects of small roundabouts on emissions and fuel consumption: A case study. *Transportation Research D: Transport and Environment* 2002;7(1):65-71.
- Williams BM, Tagliaferri Anthony P, Meinhold Stephen S, Hummer Joseph E, Roupail Nagui M. Simulation and analysis of freeway lane reversal for coastal hurricane evacuation. *ASCE Journal of Urban Planning and Development* 2007;33(1):61-72.

4

SEISMIC TESTING OF HIGHWAY BRIDGES

ERIC V. MONZON, AHMAD M. ITANI, AND GOKHAN PEKCAN

- 4.1 Introduction
- 4.2 Similitude requirements
 - 4.2.1 Dimensional analysis
 - 4.2.2 Seismic modeling
 - 4.2.3 Example
- 4.3 Specimen fabrication
 - 4.3.1 Substructure
 - 4.3.2 Superstructure
- 4.4 Input motion
- 4.5 Instrumentation
 - 4.5.1 Strain measurement
 - 4.5.2 Displacement measurement
 - 4.5.3 Force measurement
 - 4.5.4 Acceleration measurement
- 4.6 Data acquisition and processing
- 4.7 Results
- References

4.1 INTRODUCTION

Seismic behavior of structures is quite complex. This is due to the nonlinearity effect of the structure itself coupled with the erratic movement of the ground. The seismic response and the damage pattern have yet to be discovered because of the rapid evolution of the construction specifications and codes. Computational models for analyses of structural systems are simpler than the actual behavior and mechanisms

of the structures. The complexity in the structure response lies in the repetition of large cycles of deformation coupled with the geometric nonlinearity of the structure. Therefore, laboratory experiments are conducted where loading can be applied in controlled environment and where the response of the structure can be observed and measured. Laboratory and field studies are fundamental part of earthquake engineering research. Several examples can be shown of how limited amount of testing have already improved the seismic response of highway bridges. The dramatically improved responses of new versus older reinforced concrete bridges during the Northridge Earthquake are true statements about the efficacy of seismic experimental investigation. The poor performance of older reinforced concrete bridges has demonstrated the need, yet again, to the importance of experimental testing of structures in order to improve their response during large ground motions. The significant economic loss and the social disruption after each earthquake proved the need to enhance the U.S. seismic standards for construction.

Out of the laboratory studies, using earthquake simulators (shake tables) is considered one of the main tools that can be used to determine the seismic response of structures. Shake-table testing provides the most realistic means of simulating seismic effect in the laboratory. This type of testing subjects the structure to seismic motions that have controlled intensity, frequency of content, and duration. One of the drawbacks of the shake-table testing is the payload limitation of the simulator itself since the specimen has to be scaled down in order to fail the specimen under realistic ground motion. In general a structure model is said to have similarity with the real structure (prototype) if the two share geometric, kinematic, and dynamic similarity. The most effective type of similarity is the kinematic similarity where the path and the velocity of the moving particles are same.

4.2 SIMILITUDE REQUIREMENTS

Civil engineering structures such as highway bridges are large such that experimental testing at full scale is economically and technologically difficult. Scaled model testing is favored over full-scale testing because it can be performed inside a laboratory with a more controlled environment. The similitude requirements, when followed properly, provide scaled models that capture the behavior of the full-scale structure. The seismic response of structures is a dynamic mechanical problem where the fundamental physical quantities measured are length (L), force (F), and time (T).

4.2.1 Dimensional Analysis

Buckingham's Pi Theorem (Buckingham, 1914; Harris and Sabnis, 1999) states that any dimensionally homogeneous equation involving certain physical quantities can be reduced to an equivalent equation involving a complete set of dimensionless products. For n physical variables, it can be shown that any equation of the form

$$F(X_1, X_2, \dots, X_n) = 0 \quad (4.1)$$

can be equivalently expressed in the form

$$G(\pi_1, \pi_2, \dots, \pi_m) = 0 \quad (4.2)$$

The π terms are dimensionless products of the n physical variables. The number of π terms that can be defined by the variables X_1, X_2, \dots, X_n is equal to

$$m = n - r \quad (4.3)$$

where r is the number of fundamental measures like length (L), force (F), and time (T). To illustrate, consider the axial deformation of a steel bar given by the equation

$$u = \frac{PL}{EA} \quad \text{or} \quad uEA - PL = 0 \quad (4.4)$$

where u is the axial deformation, P is the axial force, E is the modulus of elasticity of material, A is the cross-sectional area, and L is the length. Equation (4.4) has five variables and has the form of Equation (4.1) thus,

$$F(u, P, L, A, E) \quad (4.5)$$

Writing the displacement as a function of the other variables,

$$u = F'(P, L, A, E) \quad (4.6)$$

Any equation of this form can be represented as a product of powers

$$u = KP^a L^b A^c E^d \quad (4.7)$$

where K is a dimensionless constant. Equation (4.7) expressed in dimensional form using the fundamental measures of length (L) and force (F) becomes

$$L \doteq F^a L^b (L^2)^c (FL^{-2})^d \quad (4.8)$$

where \doteq means dimensional equivalence. Writing the exponential equalities gives the following

$$\text{for } F : \quad 0 = a + d \quad (4.9)$$

$$\text{for } L : \quad 1 = b + 2c - 2d \quad (4.10)$$

Equations (4.9) and (4.10) have four unknowns and thus there is twofold infinity of solutions. Solving the equations in terms of a and b ,

$$a = -d \quad (4.11)$$

$$b = 1 - 2c + 2d \quad (4.12)$$

It should be noted that the exponents a and b were chosen because it takes in the fundamental measures F and L in Equation (4.7). Then, substituting Equations (4.11) and (4.12) to Equation (4.7) gives

$$u = KP^{-d}L^{1-2c+2d}A^cE^d \quad (4.13)$$

$$\left(\frac{u}{L}\right) = K\left(\frac{L^2E}{P}\right)^d\left(\frac{A}{L^2}\right)^c \quad (4.14)$$

Thus,

$$G\left(\frac{u}{L}, \frac{L^2E}{P}, \frac{A}{L^2}\right) = 0 \quad (4.15)$$

The terms inside the parenthesis of Equation (4.15) are called the π terms. As shown, the five variables were reduced to three dimensionless products as determined from Equation (4.3)

$$m = n - r = 5 - 2 = 3$$

where the two fundamental measures (r) are L and F . Therefore, the dimensionless π terms are

$$\pi_1 = \frac{u}{L}, \quad \pi_2 = \frac{L^2E}{P}, \quad \pi_3 = \frac{A}{L^2} \quad (4.16)$$

For complete similarity,

$$\pi_{1p} = \pi_{1m}, \quad \pi_{2p} = \pi_{2m}, \quad \pi_{3p} = \pi_{3m} \quad (4.17)$$

where π_{ip} refers to the prototype and π_{im} refers to the model. Then, using these relations to determine the scale factors,

$$\frac{\pi_{1p}}{\pi_{1m}} = 1 = \frac{(u/L)_p}{(u/L)_m} = \frac{(u_p/u_m)}{(L_p/L_m)} = \frac{S_u}{S_L} \quad \text{or} \quad S_u = S_L \quad (4.18)$$

$$\frac{\pi_{2p}}{\pi_{2m}} = 1 = \frac{(L^2E/P)_p}{(L^2E/P)_m} = \frac{S_L^2 S_E}{S_P} \quad \text{or} \quad S_P = S_L^2 S_E \quad (4.19)$$

$$\frac{\pi_{3p}}{\pi_{3m}} = 1 = \frac{(A/L^2)_p}{(A/L^2)_m} = \frac{S_A}{S_L^2} \quad \text{or} \quad S_A = S_L^2 \quad (4.20)$$

If the same material is used in the prototype and the model, Equation (4.19) becomes

$$S_P = S_L^2 \quad (4.21)$$

It was shown above that the formulation of scaling relations can be established by translation of the π terms into required scale factors. The procedure to determine the scaling relations can be summarized as

1. Select the set of quantities
2. Derive the set of π terms
3. Establish the similitude relations.

Lack of similarity between the prototype and the model will lead to erroneous results. Most often, however, the selection of model geometric scale and material properties are limited by the available equipments and materials. Thus, certain degree of lack of similarity is allowed provided that the behavioral difference between the prototype and the model is negligible.

4.2.2 Seismic Modeling

The basic quantities in earthquake loading are length (L), force (F), time (T), modulus of elasticity (E), mass (M), and acceleration (a). Following the procedure described above, some of the π terms that can be derived from these quantities are

$$G\left(a, \frac{L}{T^2}, \frac{L^2 E}{F}, \frac{ML}{FT^2}\right) = 0 \quad (4.22)$$

It follows that

$$S_a = 1 \quad (4.23)$$

$$S_L = S_T^2 \quad \text{or} \quad S_T = \sqrt{S_L} \quad (4.24)$$

$$S_F = S_L^2 S_E \quad (4.25)$$

$$S_M = \frac{S_F S_T^2}{S_L} \quad (4.26)$$

Substituting Equations (4.24) and (4.25) to Equation (4.26), it can be shown that

$$S_M = \frac{(S_L^2 S_E) S_L}{S_L} = S_L^2 S_E \quad (4.27)$$

If the prototype and the model is made of the same material,

$$S_M = S_L^2 \quad (4.28)$$

Thus, the required mass of the model M_m^r is given by

$$M_m^r = \frac{M_p}{S_L^2} \quad (4.29)$$

where M_p is the mass of the prototype. However, by scaling the dimensions of the prototype to get the geometry of the model and using the same material, the actual mass of the model is less than what is required by Equation (4.29), as explained below. From the relationship mass M equals density ρ times volume V , it can be shown that

$$S_M = S_\rho S_V^3 = S_\rho S_L^3 \quad (4.30)$$

which follows that the actual mass of the model is

$$M_m^a = \frac{M_p}{S_\rho S_L^3} \quad (4.31)$$

Subtracting Equation (4.31) from Equation (4.29) to determine the difference between the required and actual mass gives

$$\Delta M = M_m^r - M_m^a = M_p \left(\frac{1}{S_L^2} - \frac{1}{S_\rho S_L^3} \right) \quad (4.32)$$

ΔM can also be expressed in terms of the actual model mass by substituting Equation (4.31) to Equation (4.32)

$$\Delta M = M_m^a S_\rho S_L^3 \left(\frac{1}{S_L^2} - \frac{1}{S_\rho S_L^3} \right) = M_m^a (S_\rho S_L - 1) \quad (4.33)$$

Since the prototype and the model is of the same material as assumed in Equation (4.28),

$$\Delta M = M_m^a (S_L - 1) \quad (4.34)$$

This indicates that additional mass is needed in the model for correct dynamic simulation. For example, if $S_L = 3$ and the model weight is 50 kN, an additional weight of 100 kN is needed. The added weight must be dispersed throughout the model volume for complete similarity. However, this would be difficult to achieve so an artificial mass simulation where the added mass is concentrated at certain locations is usually used (see Section 4.2.3).

Other scaling relations typically used for seismic testing is shown in Table 4.1. Note that when the materials used in the prototype and the model are the same, the scale factors become a function of length L only.

4.2.3 Example

This example shows the three-span, steel I-girders, horizontally curved bridge used for seismic testing at University of Nevada, Reno (Buckle et al., 2011). The bridge has a high curvature (total subtended angle of 104°) representing a highway on-ramp or off-ramp. A 2.5 scale (S_L) model of this bridge is tested using the Network for Earthquake Engineering Simulation (NEES) multiple shake tables at the University of Nevada, Reno.

TABLE 4.1 Typical Scale Factors for Complete Similarity Seismic Modeling

Parameter	Dimension	Scale Factors and Relations
Length, L	L	S_L
Area, A	L^2	S_L^2
Volume, V	L^3	S_L^3
Force, F	F	$S_E S_L^2$
Pressure, q	FL^{-2}	S_E
Acceleration, a	LT^{-2}	1.00
Velocity, v	LT^{-1}	$S_L^{1/2}$
Displacement, u	L	S_L
Time, t	T	$S_L^{1/2}$
Frequency, f	T^{-1}	$S_L^{-1/2}$
Modulus of elasticity, E	FL^{-2}	S_E
Stress, σ	FL^{-2}	S_E
Strain, ε	—	1.00
Poisson's ratio, ν	—	1.00
Mass density, ρ	$FL^{-4}T^2$	S_E/S_L
Energy, E	FL	$S_E S_L^3$

The model is used to experimentally study the seismic performance of a highly curved bridge in variety of configurations including:

- conventional superstructure, columns, and bearings without live load (benchmark bridge) and with live load (see Figure 4.1);
- fully base isolated superstructure on conventional columns;
- partially isolated superstructure with ductile cross-frames on conventional columns (hybrid protective system);
- conventional superstructure on rocking foundation;
- conventional superstructure and columns with abutment interaction and simulated backfill.

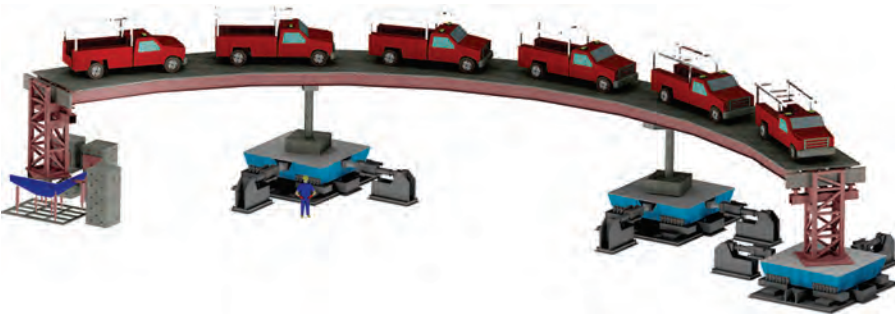
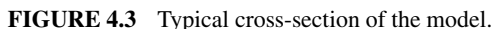


FIGURE 4.1 3D rendering of the curved bridge model on shake tables with trucks on deck representing the live loads.



The total weight of the model superstructure (girder, cross-frames, and deck) is 476 kN. As explained in Section 4.2.2, additional mass is required for simulation because the prototype and the model materials have the same densities. Due to the difficulty in distributing the additional mass throughout the volume or surface of the model, they are added externally and are distributed on discrete locations on the deck using steel and lead plates (see also Section 4.3.2). The steel and lead plates are evenly distributed on the deck such that the bridge center-of-mass is not changed. Using the scaling relations described above, the additional mass weighs 714 kN.

Due to the technological improvements, seismic testing of bridges nowadays usually involves large-scale models. The University of Nevada, Reno (UNR) has been in the forefront of seismic testing of large-scale bridges, which includes 1-span steel I-girder bridge, 4-span concrete bridge, and 3-span curved steel I-girder bridge. The NEES facility at UNR is unique because of the four relocatable shake tables that can be configured for bridge seismic testing at large-scale.

After the dimension of the model has been finalized, the construction drawings are produced. Because of the size and complexity of the bridge components, contractors are usually hired to fabricate the bridge components. Depending on the type of the project, different contractors may be hired to fabricate the superstructure and the substructure. The components are then assembled inside the laboratory by the contractor or by the researchers and laboratory staff. The researchers are responsible for the coordination and quality control. In the case of curved steel bridge, an AISC-certified steel bridge erector was hired to fabricate the deck while a concrete contractor was hired to fabricate the columns. The typical process in specimen fabrication is discussed below with the curved bridge fabrication used as an example.

Usually, the main concern in column fabrication is the size of reinforcement. For large-scale models, which is usually the case in bridge seismic testing nowadays, regular deformed bars are used for longitudinal (main) reinforcement and either deformed bars or

plain wires are used for transverse (lateral) reinforcement. In the United States, the smallest available deformed bar is #2 (No. 6) but the most readily available is #3 (No. 10). Whenever the scale factor does not permit the use of an exact scaled size of reinforcement, the number and size of reinforcement is determined by keeping the steel ratio (total steel area over total concrete area) the same in the prototype and the model.

Figure 4.4 shows the rebar cage of the 24-in. (610-mm) diameter column used in the curved bridge model, deformed bars are used in the main and lateral reinforcements because of large scale. The main reinforcement is #5 (No. 16) deformed bars and the lateral reinforcement is #3 (No. 10) deformed bars. The researchers provided the contractor the bar cut-off schedule for accurate bar arrangement and placement. After the rebar cage was assembled, locations where the strain gages will be attached are marked. The strain gages are usually placed at the expected location of plastic deformation in the column, which is at the bottom of the column above the footing and/or at the top of the column below the cap beam. After grinding and cleaning, the strain gages are glued to the main and lateral reinforcements and wrapped with an electrical tape to protect from damage during the concrete pour. To protect the strain gage wires from damage during concrete pour, they are inserted into plastic tubes as shown in Figure 4.5.

The footing is cast first. Then, after the footing concrete has hardened, the forms for the column and the cap beam are placed. Threaded rods, used for mounting of displacement transducers that will monitor the column rotation, are inserted into the sides of the



FIGURE 4.4 Complete column rebar cage.

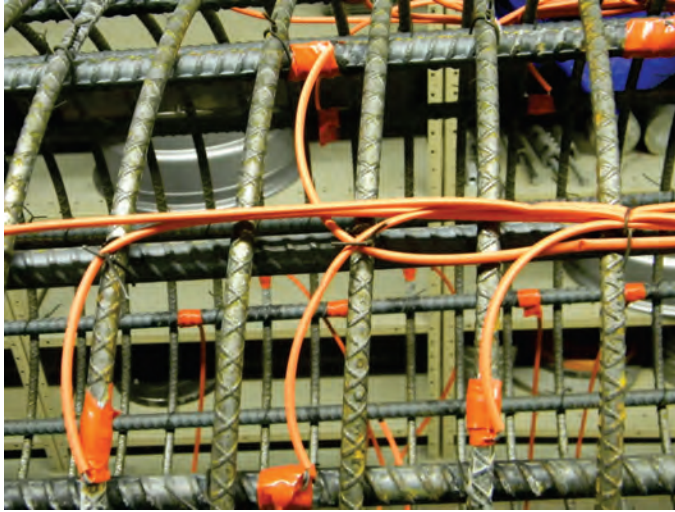


FIGURE 4.5 Close-up of the rebar cage showing the locations of the strain gages.

column (see Figure 4.6). The column and cap beam may be cast at once or at different stages. Concrete cylinder samples are obtained during each concrete pour. These will be used to determine the concrete strength at 7, 14, and 28 days. Usually, extra concrete cylinders are also cast to determine the concrete strength at the day of bridge test.

Figure 4.7 shows the completed columns with the footing and cap beam. The footing is usually bigger than needed because it will be rigidly attached to the shake tables by post-tensioning. The footing is designed based on the column capacity and the clamping forces between the footing and the shake table. Also shown in Figures 4.6 and 4.7 are the lifting



FIGURE 4.6 Picture showing the threaded rods inserted into the column sides and the lifting eye on the footing.



FIGURE 4.7 Completed columns.

eyes in the footing that will be used to transport the column from the fabrication yard to inside the laboratory.

4.3.2 Superstructure

Superstructures of highway bridges can be classified as either stiff or flexible. Examples of stiff superstructures are concrete box-girder and steel box-girder that are not only stiff laterally but also torsionally. Examples of laterally and torsionally flexible superstructures are steel I-girder bridges.

For structures that are laterally and torsionally stiff, equivalent superstructure can be used. Figure 4.8 shows the bridge model of a 4-span bridge on shake tables (Nelson et al.,



FIGURE 4.8 4-Span bridge on shake tables (Nelson et al., 2007).

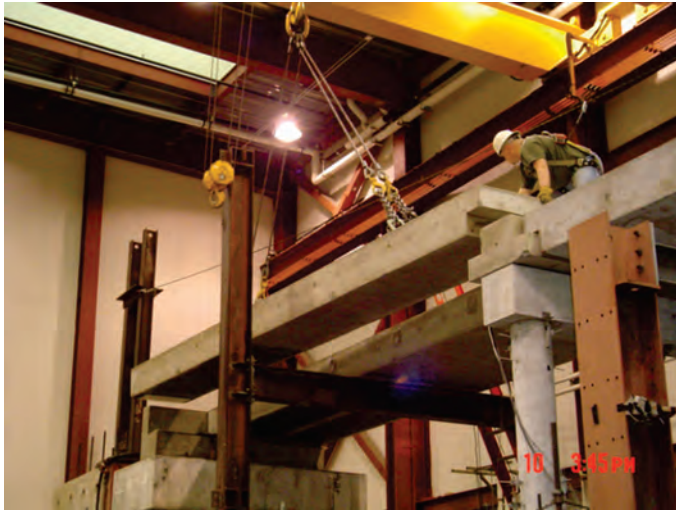


FIGURE 4.9 Placement of the superstructure of the 4-span bridge (Nelson et al., 2007).

2007). The superstructure of the prototype was post-tensioned concrete box-girder, which is relatively high in strength and stiffness compared to the substructure. Since the superstructure is essentially rigid, damage in the bridge during seismic testing is concentrated in the columns only. Because of this, instead of fabricating the scaled concrete box-girder, a solid rectangular section was used to represent the superstructure. The dimensions of the solid rectangular section superstructure were chosen such that the bending properties are the same as that of scaled box-girder. The model superstructure was designed to carry its weight, the weight of the added mass (shown as concrete blocks on top of the deck in Figure 4.8), and the seismic forces. The superstructure was constructed as three separate post-tensioned rectangular beams running in the longitudinal direction (see Figure 4.9). The three beams were then tied together to act as one by post-tensioning in the transverse direction.

For structures that are laterally and torsionally flexible, like the curved steel I-girder bridge discussed in Section 4.2.3, it is important to model the actual geometry and components of the superstructure. Figure 4.10 shows the as-built superstructure of the curved steel I-girder bridge. Due to the bridge model size and complexity, an AISC-certified steel bridge erector was hired to fabricate the superstructure. The girders are built-up steel plates, the cross-frames are angle sections, stiffener plates are provided at each cross-frame locations, and shear connectors are provided (see Figure 4.11) just as it was designed in the prototype bridge. The shear connectors provided the composite action between the girders and the deck. Their size, length, and spacing are the exact scale of those in the prototype which is, again, possible because of the large scale. In case this is not possible, shear connector size, length, and spacing in the model must be designed to provide composite action. Care must be taken when selecting the shear connector size in the model because when the shear connector diameter is larger than the thickness of the plate it is welded to, there is tendency for the shear connector to fail by fracture at the connection. The girders were cambered and the amount of camber was determined from



FIGURE 4.10 As-built superstructure of the curved bridge.

the model weight and the added weight. Figure 4.12 shows the steel reinforcement in the deck. Plywood was used for the forms in this deck. For larger deck thickness, permanent corrugated steel deck forms may be used.

The bridge model length is 145 ft (44.21 m). It is not possible to construct the entire superstructure as one continuous segment so the superstructure was divided into three segments. The abutments and piers were placed on the shake tables first then the superstructure end segments were placed on top of them. Figure 4.13 shows one of the end



FIGURE 4.11 Picture showing the assembled girders and cross-frames, and shear connectors.



FIGURE 4.12 Deck steel reinforcement.

segments being flown onto position. The middle segment was added last and connected to the two end segments as shown in Figure 4.14. The splice connections are at the location of inflection point of the center span.

Finally, the added masses using lead and steel plates were bolted into the deck. They were placed evenly at discrete locations on the deck, as shown in Figure 4.15. It is important to note that the added masses should be positioned such that the bridge center-of-mass is not changed.



FIGURE 4.13 One of the curved bridge segments being flown into position.

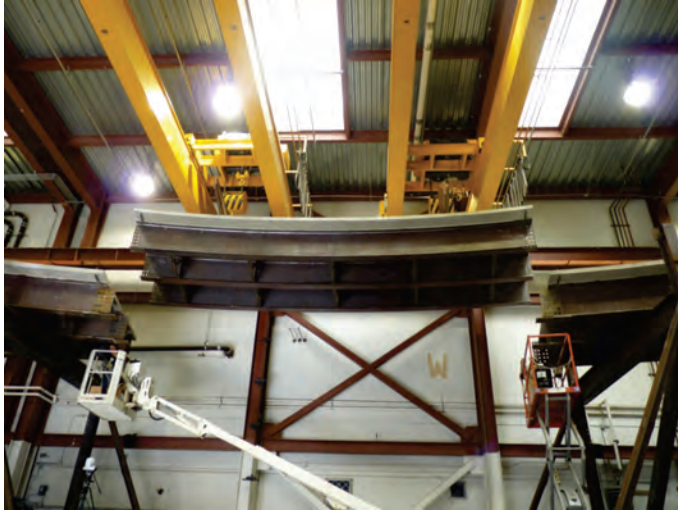


FIGURE 4.14 The middle segment of the curved bridge is connected to the end segments.



FIGURE 4.15 Picture showing the added mass on the deck.

4.4 INPUT MOTION

In shake-table studies that involve spatially long structural systems such as bridges, preliminary sensitivity analyses on reliable numerical/computational models must be conducted to ensure that wide range of ground motions with characteristics that are critical to the seismic response evaluation of the prototype/model structure is determined. When multiple shake tables are used, the possible account of the spatial variability of seismic

ground motions can also be evaluated. It is important to note that the selection of ground motions with high amplitudes of spectral acceleration in the expected frequency range of the shake table and the model structure must be considered such that the peak response parameters (displacement, velocity, and acceleration) do not exceed the physical limits of the shake table. These verifications can be done based on the numerical simulation of the scaled model structure where the time axes of the acceleration and displacement records are condensed by applying the time scale factor. In other words, the ground motion records that will be used in the numerical simulation as well as the shake table signal must be compressed in time by $S_T = \sqrt{S_L}$ to satisfy the similitude requirements. In this way the relationship between the frequency content of the compressed record and the natural period of the undamaged scale model is maintained with respect to that of the original record and the prototype structure (Harris and Sabnis, 1999). In most cases, the ground motion signals will also need to be filtered to eliminate the high-frequency content outside the performance range of the shake table and the low-frequency content that correspond to large displacements exceeding the displacement limits of the shake table. The effect of filtering should also be verified based on the numerical simulation studies of the prototype and the scaled model structures.

In general, the shake-table input signals that can be used are (1) simulated (artificial) spectrum compatible records; (2) synthetic acceleration records; and (3) recorded (historical) ground motions.

Spectrum-compatible signals can be obtained by generating a frequency domain function, for example, power spectral density function from a code-specified response spectrum, and deriving signals compatible to that spectrum. However, in most of the cases, this approach could lead to accelerograms not reflecting the real phasing of seismic waves and cycles of motion, and therefore frequency and energy content. Regardless, the most common approach in generating spatially variable seismic ground motions is by means of simulations; the motions are modeled as stochastic vector processes, described by their power spectral density, coherency, and wave propagation (Deodatis, 1996).

Synthetic acceleration records are obtained via modeling of the seismological source and may account for path and site effects. These methods range from stochastic simulation (Boore, 2003) of point or finite sources to dynamic models of rupture. However, they often require setting of physical parameters associated with the ground motion characteristics and mechanisms (e.g., rupture parameters, rise-time, and so on). Some state-of-the-art simulation methods seem to overcome these shortcomings, but they are not readily available to engineers. Also, it is noted that the synthetic acceleration records provide time histories that do not necessarily incorporate spatial coherency models that may be critical in multiple the shake-table studies of long structural systems.

Historical ground motions are recorded from real events, which become more and more readily available. The availability of online, user-friendly, databases of strong motion recordings, and the rapid development of digital seismic networks worldwide, has increased the accessibility to recorded accelerograms, which, therefore, have become the most promising candidates for the use in shake-table studies (e.g., PEER Ground Motion Database, http://peer.berkeley.edu/peer_ground_motion_database/). However, there is large variability in the recorded ground motions in terms of ground motion parameters, seismic source, soil conditions, and so on, and these may not exactly represent the specific scenario and/or seismic hazard demand desired for the bridge. Because of the highly complex nature of and uncertainties involved in the earthquake ground motions, earthquake engineering research has focused for the past two decades on the selection of recorded

ground-motions for the seismic response assessment and evaluation (both analytically and experimentally) of structural systems.

4.5 INSTRUMENTATION

Instruments are placed on the model to understand and characterize its behavior. For seismic testing of highway bridges, the typical quantities measured are strain, displacement, force, and acceleration. Strains are measured to determine the local component inelasticity. Displacements measure the global and local response of the system as well as its components. Forces measure the base shear and the distribution of forces among the boundary supports. Accelerations are measured to determine the model's dynamic response.

It is always tempting to put as many instruments in the model but these may not be needed and could significantly increase the cost of the experiment. Thus, analytical investigations are conducted prior to the experiment to determine the critical locations where the quantities mentioned above must be measured. Several factors that should be taken into account in planning the instrumentation include available number of instruments, cost if additional instruments are needed, available channels in the data acquisition hardware, installation of the instruments (absolute or relative measurements), and applicability and sensitivity of the instruments that will be used.

4.5.1 Strain Measurement

Strain gages are used to measure local strain of the components. The most widely used type of strain gage is the electrical resistance strain gage due to its small size, lightweight, highly sensitive to strain, low cost, and ease of attachment. The increase and decrease in the gage resistance is used to determine the amount of elongation and contraction in the component where the strain gage is attached. The strain is measured by relating the change in electrical resistance (ΔR) to the strain using gage factor (K_g)

$$K_g = \frac{\Delta R/R}{\Delta L/L} = \frac{\Delta R/R}{\epsilon} \quad (4.35)$$

where R is the original electrical resistance and ϵ is the strain. Typical value for K_g is about 2.0.

Figure 4.16 shows the typical electrical resistance, foil strain gage used to measure uniaxial strain. End A of the strain gage is glued to the surface of the component while end B is plugged into the strain gage block (see Figure 4.17). Installation of the strain gage is relatively easy. For example, the typical strain gage installation procedure in a column rebar is

1. surface is grinded, cleaned, and coated with metal neutralizer;
2. adhesive is applied to end A of the strain gage;
3. strain gage is attached to the rebar surface by applying thumb pressure;
4. rubber mastic tape is placed around the strain gage;
5. moisture sealing electrical tape is placed on the area;

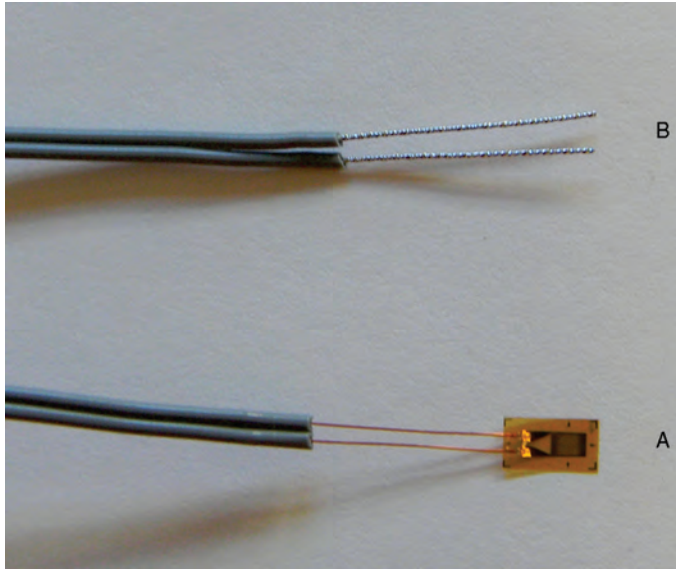


FIGURE 4.16 Typical uniaxial foil strain gage.

6. strain gage wire is inserted into heat shrink plastic tube (see Figure 4.18) to protect it against damage during concrete pour.

In columns, the strain gages are placed on the main and lateral reinforcements at the expected plastic hinge location. For a column behaving in single curvature, the plastic hinge location is at the bottom of the column above the footing. For a column behaving in double curvature, the plastic hinge locations are at the top (below the cap beam) and the

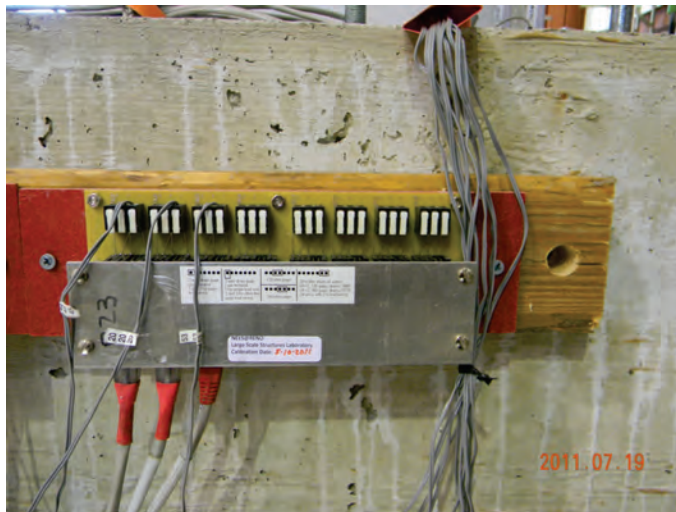


FIGURE 4.17 Strain gage block where end B of strain gage is plugged into.



FIGURE 4.18 Strain gage wires are inserted into plastic tubes to protect from damage during concrete pour.

bottom (above the footing) of the column. In the girders, the stresses can be measured by attaching strain gages at the flanges and webs. Strain gages placed on the top and bottom flanges can be used to determine the vertical and lateral girder moments. In the cross-frames, the axial forces are determined by placing strain gages in the legs of the angle section used as cross-frame member. Cross-frame forces are usually monitored at the support locations where the deck seismic forces are transferred to the bearings through the cross-frames.

4.5.2 Displacement Measurement

Displacement transducers are used to measure the bridge global displacements as well as displacements of local components. The bridge global displacements are measured by attaching displacement transducers in the longitudinal, transverse, and vertical directions. This provides a three-dimensional response of the bridge at any particular time during the test. For large displacements, such as deck displacements, cable extension transducer or string pot (UniMeasure, Celesco Transducer Products, Inc.) is one of the ideal instruments. This instrument utilizes a flexible cable and a potentiometer to detect and measure linear position. Figure 4.19 shows some of the string pots attached to the deck of the curved bridge. The recorded displacement can be either absolute or relative displacement depending on the location where the other end of the instrument is attached. Those instruments that are connected to fixed points in the laboratory records absolute displacements while those connected to rigid structures moving with the shake-table records relative displacements. In the end, only the relative values are reported so the absolute displacements are converted into relative displacements by subtracting the shake-table displacements.

Relative displacement between the girders can be measured using linear potentiometric transducers as shown in Figure 4.20. This instrument records the diagonal displacement between the girders that are then used to determine the girder rotation and drift.



FIGURE 4.19 String pots attached to the deck of the curved bridge.

The column curvature can also be measured using displacement transducers. One common transducer used for this purpose is the position sensors with restoring springs (NovoTechnik U.S., Inc.) which are attached to the opposite sides of the column as shown in Figure 4.21. With the distance between the opposite transducers known, the column rotation is measured using the recorded displacement in the opposite transducers. Similar to strain gages, these instruments are placed at the expected plastic hinge locations.



FIGURE 4.20 Picture showing the displacement transducer connected diagonally between the girders.



FIGURE 4.21 Displacement transducers attached to the column sides to measure column curvature.

4.5.3 Force Measurement

Forces in the boundary supports are measured using load cells. The typical load cells used in bridge seismic testing are strain gage load cells because of their relatively low cost and they can be easily fabricated using the equipments available in a laboratory. The load cell principles are basically the same as those for the strain gages. The strain gages are attached to the sides of the structural section (e.g., steel tube, pipe, or beam) used as load cell. The electrical signals are converted into strains that in turn are converted into force. The strain gages can be configured to measure axial force, shear, and bending moments. Figure 4.22 shows the load cells placed between the bearings and the cap beam of the



FIGURE 4.22 Load cells (colored black) between the bearings and cap beam.



FIGURE 4.23 Triaxial micro electro-mechanical accelerometer on the curved bridge deck.

curved bridge to measure the girder support reactions and the force transmitted to the column.

4.5.4 Acceleration Measurement

Accelerations in the structure are measured for system identification, mode shapes, natural frequency, comparison between the input motion and bridge response, and for estimation of inertia loads. Accelerometers are placed on strategic locations on the deck like those at the midspan of each span and at the supports. In the curved bridge test, accelerometers are also placed on the cap beam to calculate the inertia loads in the cap beam. Figure 4.23 shows the triaxial micro electro-mechanical accelerometer to measure the deck acceleration in the longitudinal, transverse, and vertical directions. Although the input motions are in the longitudinal and transverse directions only, vertical acceleration is measured because vertical vibration is one of the dominant vibration modes of the bridge.

4.6 DATA ACQUISITION AND PROCESSING

Data acquisition (DAQ) systems convert the signals recorded by sensors (e.g., the electrical signals recorded by the strain gages) into digital numeric values, known as data. A typical data production chain is comprised of the sensor attached to the model, the cable connecting the sensor to the DAQ hardware, the DAQ hardware processing the signal, and the PC recording the data. Figure 4.24 shows the DAQ hardware used at the Large-Scale Structures Laboratory (LSSL) at UNR.

Data is the most important aspect of any test. It must be recorded, saved, and be accurate relative to the test. In order for the data to be meaningful and processed properly, metadata such as instrument location and polarity must be known, the instrument was not damaged during the test, and instrument was not altered between the tests. The researcher



FIGURE 4.24 DAQ system used at UNR's Large Scale Structures Laboratory.

must be aware of the errors introduced in the instrument, such as kinematic error, error due to instrument calibration, and error due to DAQ calibration. Kinematic error is due to instrument location and structure movement. For example, consider the two string pots attached to the deck of a straight bridge to measure longitudinal and transverse displacements, as shown in Figure 4.25. Because the bridge moves diagonally, the recorded displacement is larger than the actual longitudinal or transverse displacement. This error must be recognized and corrected during data processing. It can be minimized during instrument installation by increasing the distance between the measured point and the reference point. The error due to instrument calibration and DAQ calibration is usually small compared to that due to instrument attachment.

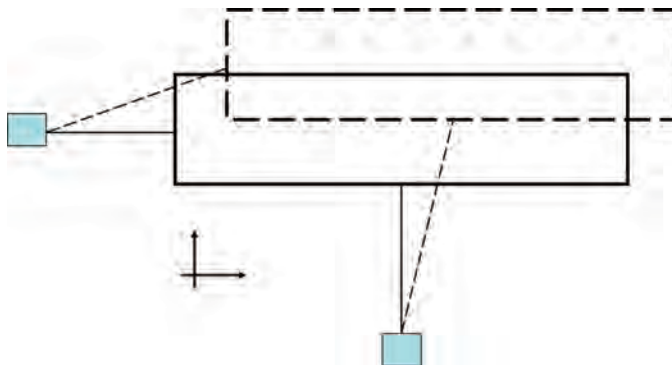


FIGURE 4.25 Diagram (plan view) illustrating the kinematic error in measuring longitudinal and transverse displacement.

4.7 RESULTS

Several experimental results can be reported but the most important ones include the following:

- force–displacement hysteresis plots
- energy dissipated
- system capacity
- ductility
- damage progression
- natural frequency
- damping

The force–displacement hysteresis is plotted from the load cell readings and the corresponding displacement transducer readings. Total base shears can be determined from the shake-table outputs that include the shake-table lateral forces. Energy dissipated due to hysteretic damping is calculated by integrating the area inside the hysteresis loops. Also, from the hysteresis plots, the system capacity (force versus displacement) is established by plotting the envelope (maximum and minimum) force–displacement relationship obtained from each run. The resulting “backbone curve” of system capacity is used to determine the yield capacity and ultimate capacity. Then the ductility, ratio of ultimate deformation to yield deformation, is determined.

The damage progression is evaluated locally (at the component level) and globally (at the system level). Local damage progression is reported by plotting the strain history for each test and also by visual inspection of each component after each test.

At the system level, damage progression is established by exciting the bridge with randomly generated low level motion (called as white noise). The frequency response spectrum is plotted and frequency shift of the dominant modes is observed by comparing it with those from the previous tests. The natural frequency corresponds to the peak value of the frequency response curve. Also, from the frequency response curve, damping can be evaluated using the half-power bandwidth (Chopra, 2007). For example, consider the frequency response curve shown in Figure 4.26, the natural frequency f_n is the frequency

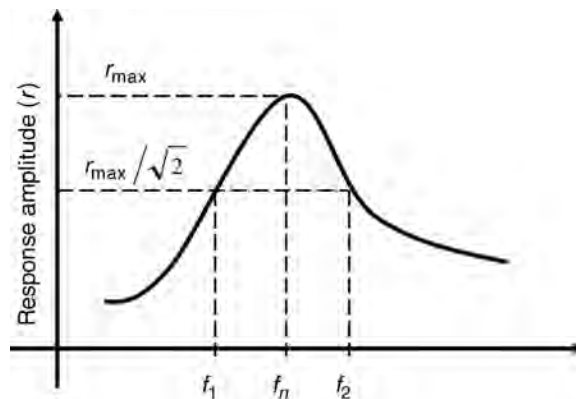


FIGURE 4.26 Frequency response curve.

corresponding to the maximum response amplitude r_{\max} . The damping is calculated using the equation

$$\zeta = \frac{f_2 - f_1}{2f_n} \quad (4.36)$$

where ζ is the damping ratio and f_1 and f_2 are the frequencies corresponding to $r_{\max}/\sqrt{2}$.

In addition, the measured quantities are used to calibrate the analytical models such that the measured value is almost equal to those obtained from the analysis.

REFERENCES

- AASHTO. *AASHTO LRFD Bridge Design Specifications*. 5th ed. Washington, DC: American Association of State Highway and Transportation Officials; 2010.
- Boore DM. Simulation of ground-motion using the stochastic method. *Pure and Applied Geophysics* 2003;160:635–676.
- Buckingham E. On physically similar systems, *Physical Review*, London. 1914.
- Buckle IG, Monzon EV, Itani AM. Seismic analysis of curved steel-girder highway bridges. 8th International Conference on Urban Earthquake; Tokyo, Japan: Engineering; 2011.
- Chopra AK. *Dynamics of Structures: Theory and Applications to Earthquake Engineering*. 3rd ed., Upper Saddle River, NJ: Pearson Prentice Hall; 2007.
- Deodatis G. Non-stationary stochastic vector processes: Seismic ground motion applications. *Probabilistic Engineering Mechanics* 1996;11:149–167.
- Harris HG, Sabnis GM. *Structural Modeling and Experimental Methods*. 2nd ed. Boca Raton, FL: CRC Press; 1999.
- Nelson R, Saiidi M, Zadeh M. 2007. Experimental Evaluation of Performance of Conventional Bridge Systems, Report No. CCEER 07-04, University of Nevada, Reno, NV.

5

MEASUREMENTS IN ENVIRONMENTAL ENGINEERING

DANIEL A. VALLERO

- 5.1 Introduction
 - 5.1.1 Data quality objectives
 - 5.1.2 Monitoring plan example
 - 5.1.3 Selection of a monitoring site
- 5.2 Environmental sampling approaches
- 5.3 Laboratory analysis
 - 5.3.1 Extraction
 - 5.3.2 Chromatography
 - 5.3.3 Limits of detection
 - 5.3.4 Aerosol limits of detection
 - 5.3.5 Microbial limits of detection
- 5.4 Measurement uncertainty
- 5.5 Measurement decision making
- 5.6 Environmental indicators
 - 5.6.1 Oxygen indicators
 - 5.6.2 Indices
- 5.7 Extending measurement data using models
- 5.8 Summary
- Nomenclature
- References

5.1 INTRODUCTION

Environmental engineers apply the sciences to improve the public health and ecosystems. As such, they must have reliable data and information about the condition of air, water, soil, sediment, and biota (see Figure 5.1).

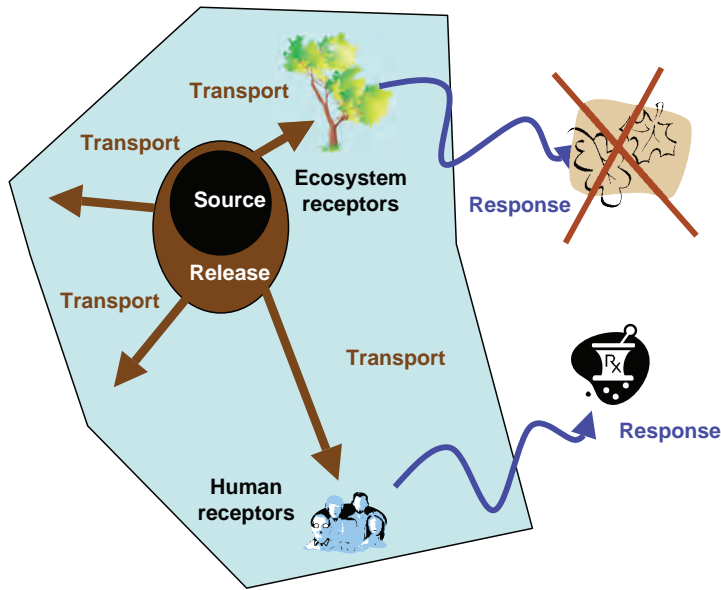


FIGURE 5.1 Sites of environmental measurements. *Source:* Letcher and Vallero (2011).

Environmental monitoring is dependent on the quality of sample collection, preparation, and analysis. Sampling is a statistical term, and usually a geostatistical term.

An environmental sample is a fraction of air, water, soil, biota, or other environmental media (e.g., paint chips, food, and so on for indoor monitoring) that represents a larger population or body. For example, a sample of air may consist of a canister or bag that holds a defined quantity of air that will be subsequently analyzed. The sample is representative of a portion of an air mass. The number of samples must be collected and their results aggregated to ascertain with defined certainty the quality of an air mass. More samples will be needed for a large urban air shed than for that of a small town. Intensive sampling is often needed for highly toxic contaminants and for sites that may be particularly critical, for example, near a hazardous waste site or in an “at risk” neighborhood (such as one near a manufacturing facility that uses large quantities of potentially toxic materials). Similar to other statistical measures, environmental samples allow for statistical inference. In case, inferences are made regarding the condition of an ecosystem and the extent and severity of exposure of a human population.

For example, to estimate the amount of a chemical compound in a lake near a chemical plant, an engineer gathers a 500 mL sample in the middle of the lake that contains 1 million liters of water. Thus, the sample represents only 5×10^{-7} of the lake’s water. This is known as a “grab” sample, that is, a single sample taken to represent an entire system. Such a sample is limited in location vertically and horizontally, so there is much uncertainty. However, if 10 samples are taken at 10 spatially distributed sites, the inferences are improved. Furthermore, if the samples were taken in each season, then there would be some improvement to understanding of intra-annual variability. If the sampling is continued for several years, the inter-annual variability is better characterized. Indeed, this approach can be used in media other than water, for example, soil, sediment, and air.

5.1.1 Data Quality Objectives

A monitoring plan must be in place before samples are collected and arrive at the laboratory. The plan includes quality assurance provisions and describes the procedures to be employed. These procedures must be strictly followed to investigate environmental conditions. The plan describes in detail the sampling apparatus (e.g., real-time probes, sample bags, bottles, and soil cores), the number of samples needed, and sample handling and transportation. The quality and quantity of samples are determined by data quality objectives (DQOs), which are defined by the objectives of the overall contaminant assessment plan. DQOs are qualitative and quantitative statements that translate nontechnical project goals into scientific and engineering outputs need to answer technical questions (U.S. Environmental Protection Agency, 2006).

Quantitative DQOs specify a required level of scientific and data certainty, whereas qualitative DQOs express decisions goals without specifying those goals in a quantitative manner. Even when expressed in technical terms, DQOs must specify the decision that the data will ultimately support, but not the manner that the data will be collected. DQOs guide the determination of the data quality that is needed in both the sampling and the analytical efforts. U.S. Environmental Protection Agency (2001) has listed three examples of the range of detail of quantitative and qualitative DQOs:

1. *Example of a Less Detailed, Quantitative DQO:* Determine with greater than 95% confidence that contaminated surface soil will not pose a human exposure hazard.
2. *Example of a More Detailed, Quantitative DQO:* Determine to a 90% degree of statistical certainty whether the concentration of mercury in each bin of soil is less than 96 ppm.
3. *Example of a Detailed, Qualitative DQO:* Determine the proper disposition of each bin of soil in real time using a dynamic work plan and a field method able to turn-around lead (Pb) results on the soil samples within 2 h of sample collection.

If the condition in question is tightly defined, for example, the seasonal change in pH near a fish hatchery, a small number of samples using simple pH probes would be defined as the DQO. Conversely, if the environmental assessment is more complex and larger in scale, for example, the characterization of year-round water quality for trout in the stream, the sampling plan's DQO may dictate that numerous samples at various points be continuously sampled for inorganic and organic contaminants, turbidity, nutrients, and ionic strength. This is even more complicated for biotic systems, which may also require microbiological monitoring.

The sampling plan must include all environmental media, for example, soil, air, water, and biota, which are needed to characterize the exposure and risk of any biotechnological operation. The sampling and analysis plan should explicitly point out which methods will be used. For example, if toxic chemicals are being monitored, the U.S. EPA specifies specific sampling and analysis methods (U.S. EPA, 1994, 2002, 2011c).

The geographic area where data are to be collected is defined by distinctive physical features such as volume or area, for example, metropolitan city limits, the soil within the property boundaries down to a depth of 6 cm, a specific water body, length along a shoreline, or the natural habitat range of a particular animal species. Care should be taken to define boundaries. For example, Figure 5.2 shows a sampling grid, with a sample taken from each cell in the grid (U.S. EPA, 2002). The target population may be divided into

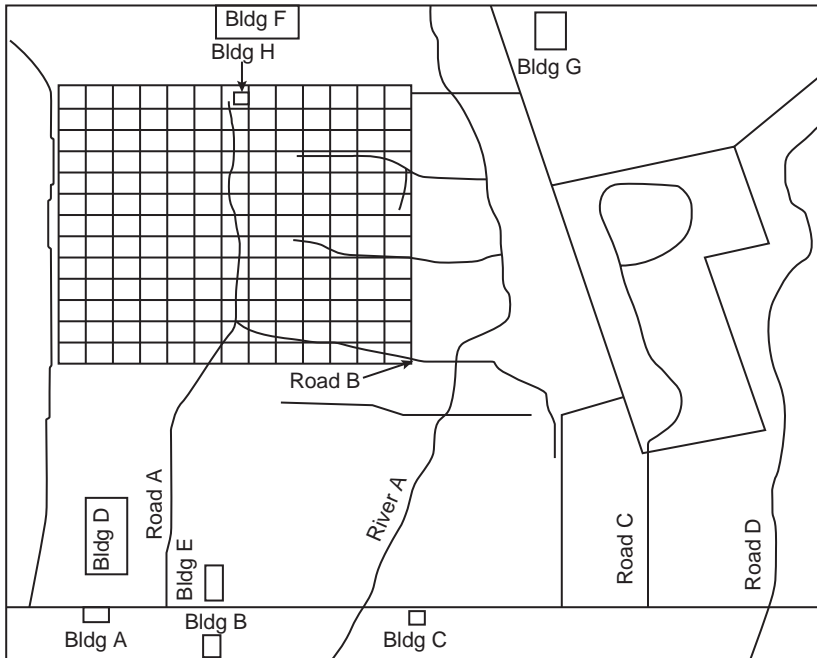


FIGURE 5.2 Environmental assessment area delineated by map boundaries. *Source:* U.S. Environmental Protection Agency (2002).

relatively homogeneous subpopulations within each area or subunit. This can reduce the number of samples needed to meet the tolerable limits on decision errors and to improve efficiency.

Time is another essential parameter that determines the type and extent of monitoring needed. Conditions vary over the course of a study due to changes in weather conditions, seasons, operation of equipment, and human activities. These include seasonal changes in groundwater levels, seasonal differences in farming practices, daily or hourly changes in airborne contaminant levels, and intermittent pollutant discharges from industrial sources. Such variations must be considered during data collection and in the interpretation of results. Some examples of environmental time-sensitivity are as follows:

- concentrations of lead in dust on windowsills that may show higher concentrations during the summer when windows are raised and paint/dust accumulates on the windowsill;
- terrestrial background radiation levels that may change due to shielding effects related to soil dampness;
- amount of pesticides on surfaces that may show greater variations in the summer because of higher temperatures and volatilization;
- instruments that may not give accurate measurements when temperatures are colder; and
- airborne particulate matter measurements that may not be accurate if the sampling is conducted in the wetter winter months rather than the drier summer months.

Feasibility should also be considered. This includes gaining legal and physical access to the properties, equipment acquisition and operation, and environmental conditions, times and conditions when sampling is prohibited (e.g., freezing temperatures, high humidity, and noise).

5.1.2 Monitoring Plan Example

Consider a plan to measure mobile source air toxics (MSAT) concentrations and variations in concentrations as a function of distance from the highway and to establish relationships between MSAT concentrations as related to highway traffic flows including traffic count, vehicle types and speeds, and meteorological conditions such as wind speed and wind direction. Specifically, the monitoring plan has the following goals (Kimbrough et al., 2008):

1. Identify the existence and extent of elevated air pollutants near roads.
2. Determine how vehicle operations and local meteorology influence near road air quality for criteria and toxic air pollutants.
3. Collect data that will be useful in evaluating and refining, if necessary, models used to determine the emissions and dispersion of motor vehicle related pollutants near roadways.

A complex monitoring effort requires management and technical staff with a diversity of skills that can be brought to bear on the implementation of this project. This diverse skill set includes program management, contracts administration, field monitoring experience, laboratory expertise, and quality assurance oversight.

The purpose of any site selection process is to gather and analyze sufficient data that would lead one to draw informed conclusions regarding the selection of the most appropriate site for the monitoring at a specific location. Moreover, the site selection process needs to include programmatic issues to ensure an informed decision is reached.

5.1.3 Selection of a Monitoring Site

Selecting a monitoring site must be based on scientific and feasibility factors, as shown in Table 5.1 and Figure 5.3. Each step has varying degrees of complexity due to “real-world” issues. The first step was to determine site selection criteria (Table 5.2). The follow-on steps include (2) develop list of candidate sites and supporting information; (3) apply site selection filter (“coarse” and “fine”), (4) site visit; (5) select candidate site(s) via team discussion; (6) obtain site access permission(s); and (7) implement site logistics.

A list of candidate sites based on these criteria can then be developed. Geographic information system (GIS) data, tools and techniques, and on-site visits would be used to compare various sites that meet these criteria.

Quite commonly, even a well-designed environmental monitoring plan will need to be adjusted during the implementation phase. For example, investigators may discover barriers or differing conditions from what was observed in the planning phase (e.g., different daily traffic counts or new road construction).

After applying site selection criteria as a set of “filters,” candidate sites are incrementally eliminated. For example, the first filter would be sites with low traffic counts; the next filter, the presence of extensive sound barriers, eliminates additional sites; and other

TABLE 5.1 Example of Steps in Selecting an Air Quality Monitoring Site

Step	Site Selection Steps	Method	Comment
1	Determine site selection criteria	Monitoring protocol	
2	Develop list of candidate sites	Geographic information system data; on-site visit(s)	Additional sites added as information is developed
3	Apply coarse site selection filter	Team discussions, management input	Eliminate sites below acceptable minimums
4	Site visit	Field trip	Application of fine site selection filter
5	Select candidate site(s)	Team discussions, management input	
6	Obtain site access permissions	Contact property owners	If property owners do not grant permission, then the site is dropped from further consideration
7	Site logistics (i.e., physical access, utilities—electrical and communications)	Site visit(s), contact utility companies	

Source: Kimbrough et al. (2008).

filters, for example, complex geometric design or lack of available traffic volume data, eliminates additional sites. Next, feasibility considerations would eliminate additional candidate sites.

An important component of “groundtruthing” or site visit is to obtain information from local sources. Local businesses and residents can provide important information needed in

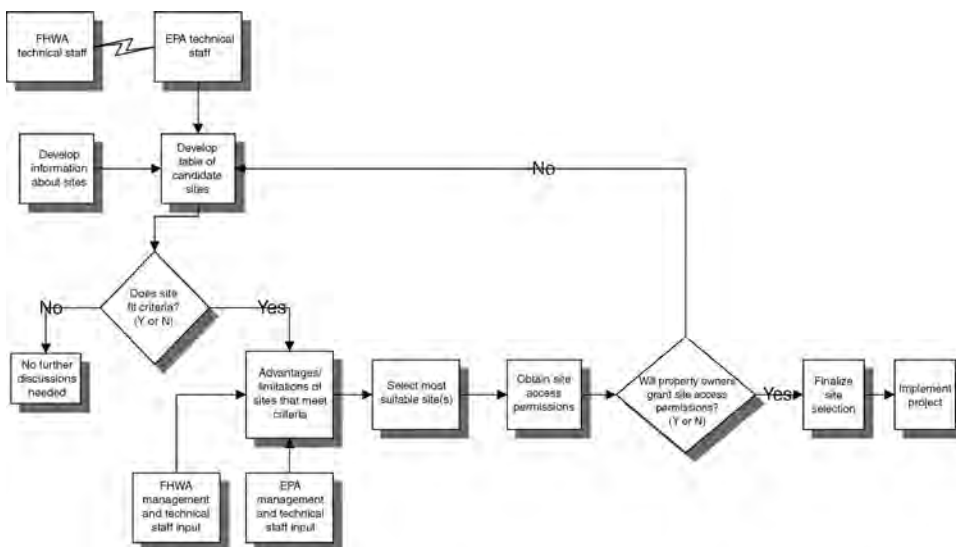


FIGURE 5.3 Monitoring location selection decision flow chart. Source: Kimbrough et al. (2008).

TABLE 5.2 Example Selection Considerations and Criteria

Selection Considerations	Monitoring Protocol Criteria
<i>Essential criteria for this monitoring study</i>	
AADT (>150,000)	Only sites with more than 150,000 annual average daily traffic (AADT) are considered as candidates
Geometric design	The geometric design of the facility, including the layout of ramps, interchanges and similar facilities, will be taken into account. Where geometric design impedes effective data collection on MSATs and PM _{2.5} , those sites will be excluded from further consideration. All sites have a “clean” geometric design
Topology (i.e., sound barriers, road elevation)	Sites located in terrain making measurement of MSAT concentrations difficult or that raise questions of interpretation of any results will not be considered. For example, sharply sloping terrain away from a roadway could result in under representation of MSAT and PM _{2.5} concentration levels on monitors in close proximity to the roadway simply because the plume misses the monitor as it disperses
Geographic location	Criteria applicable to representing geographic diversity within the United States as opposed to within any given city
Availability of data (traffic volume data)	Any location where data, including automated traffic monitoring data, meteorological or MSAT concentration data, is not readily available or instrumentation cannot be brought in to collect such data will not be considered for inclusion in the study
Meteorology	Sites will be selected based on their local climates to assess the impact of climate on dispersion of emissions and atmospheric processes that affect chemical reactions and phase changes in the ambient air
<i>Desirable, but not essential criteria</i>	
Downwind sampling	Any location where proper siting of downwind sampling sites is restricted due to topology, existing structures, meteorology, and so on may exclude otherwise suitable sites for consideration and inclusion in this study
Potentially confounding air pollutant sources	The presence of confounding emission sources may exclude otherwise suitable sites for consideration and inclusion in this study
Site access (admin/physical)	Any location where site access, is restricted or prohibited either due to administrative or physical issues, will not be considered for inclusion in the study

Source: Kimbrough et al. (2008).

a decision process, such as types of chemicals stored previously at a site, changes in vegetation or even ownership histories.

The use of spatial tools in decision processes is increasing (Malczewski, 2009). Until recently, the use of GIS and other spatial tools in decision processes has required the acquisition of large amounts of the data. In addition, the software has not been user-friendly. GIS data have now become more readily available in both quantity and quality, and GIS exists in common operating system environments.

Typical data layers that are required are the location of suitable soils, wells, surface water sources, residential areas, schools, airports, roads, and so on. From these data, layers queries are formulated to provide the most suitable sites (e.g., depth to water table may help identify sources of pollution). Typically, quantitative weighting criteria are associated with the siting criteria as well as elements of the data layers (e.g., certain types of soils would be more suitable than others and thus would have applicable quantitative values; ESRI, 1995).

5.2 ENVIRONMENTAL SAMPLING APPROACHES

Engineers use various methods of collecting environmental samples. As mentioned, the grab sample is simply a measurement at a site at a single point in time. Grab sampling generally consists of steps to ensure representativeness, including consistent collection, to prevent postsampling contamination, and proper storage and transport. For regulated samples, this also includes good record keeping (labeling, chain-of-custody, accountability). For example, grab sampling of effluents under the Clean Water Act requires the following steps (U.S. EPA, 1992):

- Sample containers must be labeled before sampling event.
- Cooler with ice must be available at time of sampling at the sampling point (to slow chemical and microbial reactions within the sample).
- Grab is made according to flow (e.g., from the horizontal and vertical center of a water channel).
- Disturbances must be avoided (e.g., not stirring up bottom sediments if interested in the water column).
- Container is held so the opening faces upstream.
- Gloves must be worn and the inside of the container must not be touched to prevent contamination.
- Water samples must be free from uncharacteristic floating debris.
- Samples are transferred into proper containers (e.g., from bucket to sample container), with notable exceptions (e.g., fecal coliform, fecal streptococcus, phenols, and oil and gas samples should remain in original containers).
- If taking numerous grabs, samples must be kept separate and clearly labeled.
- Safety and other sampling factors may vary by the media and site specifics (e.g., water versus air; surface water versus ground water; indoor versus outdoor air measurements; manufacturing versus disposal sites).

Composite sampling physically combines and mixes multiple grab samples (from different locations or times) to allow for physical, instead of mathematical, averaging. The acceptable composite provides a single value of contaminant concentration measurement that can be used in statistical calculations. Multiple composite samples can provide improved sampling precision and reduce the total number of analyses required compared to noncomposite sampling (U.S. EPA, 2011a) (e.g., “grab” or integrated soil sample of x mass or y volume), the number of samples needed (e.g., for statistical significance) the minimum acceptable quality as defined by the quality assurance (QA) plan and sampling standard operating procedures (SOPs), and sample handling after collection.

A weakness of composite sampling is the false negative effect. Composite sampling must be avoided whenever the integrity of the individual sample values changes, that is, due to physical mixing of samples. Integrity is diminished as a result of chemical precipitation, exsolution, or volatilization during sample mixing. For example, constituents with relatively high vapor pressures may be lost when mixed or reactive constituents may form new compounds within the mixture. This often requires extraction of individual samples within a laboratory environment instead of mixing individual samples, as they are collected.

The U.S. EPA notes that

... compliance with the land disposal regulation (LDR) numeric treatment standards is to be determined using “grab” samples rather than composite samples. Grab samples processed, analyzed, and evaluated individually normally reflect maximum process variability, and thus reasonably characterize the range of treatment system performance. In those cases in which only composite data were available to develop a treatment standard, the EPA used these data. For wastes for which the standards are based on composite data, enforcement of the standard is to be based on composite data. (U.S. EPA, 2011b)

Consider, for example, samples collected from an evenly distributed grid of homes to represent a neighbor exposure to a contaminant, as shown in Figure 5.4. The assessment found the values of 3, 1, 2, 12, and 2 mg/L, and the mean contamination concentration is only 4 mg/L. If cleanup is needed above the threshold of 5 mg/L, the mean concentration would indicate that the area does not need remediation and would be reported below the threshold level. However, the fourth home is well above the safety level. This could also have a false positive effect. For example, if the mean concentration were 6 mg/L in the example, the whole neighborhood may not need cleanup if the source is isolated to a confined area in the yard of home 5.

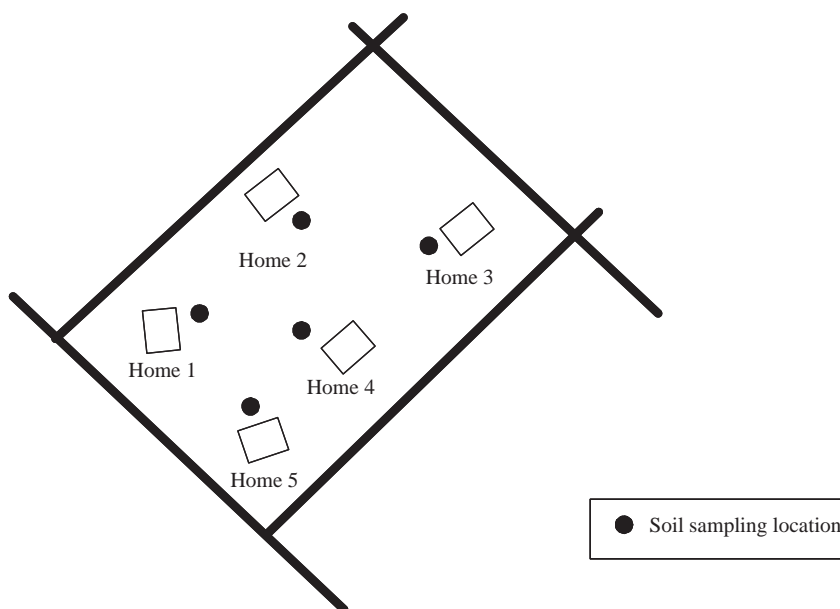


FIGURE 5.4 Composite sampling grid for a neighborhood.

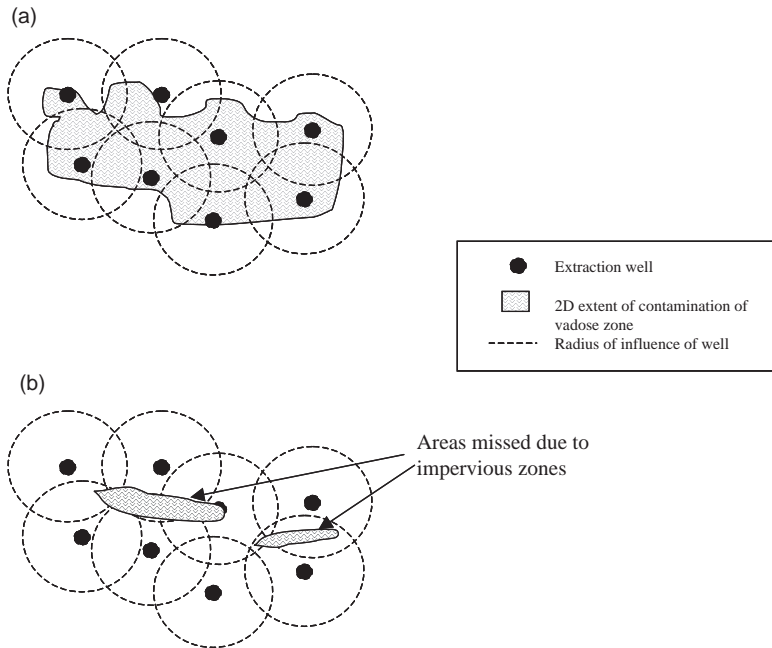


FIGURE 5.5 Extraction well locations on a geometric grid, showing hypothetical cleanup after 6 months. (a) Before treatment. (b) After treatment.

Another example of where geographic composites may not be representative is in cleaning up and monitoring the success of cleanup actions. For example, if a grid is laid out over a contaminated groundwater plume (Figure 5.5), it may not take into account horizontal and vertical impervious layers, unknown sources (e.g., tanks), and flow differences among strata, so that some of the plume is eliminated but pockets are left (as shown in Figure 5.5b).

It is often good practice to assume that a contaminated site will have a heterogeneous distribution of contamination. The amount of heterogeneity will influence not only the number of samples needed, but also the manner in which the samples are interpreted. Some of the most common sampling approaches include (Australian Department of Health and Ageing, 2002) as follows:

Random Sampling: While it has the value of statistical representativeness, with a sufficient number of samples for the defined confidence levels, for example, x samples needed for 95% confidence, random sampling may lead to large areas of the site being missed for sampling due to chance distribution of results. It also neglects prior knowledge of the site. For example, if maps show an old tank that may have stored contaminants, a purely random sample will not give any preference to samples near the tank.

Stratified Random Sampling: By dividing the site into areas and randomly sampling within each area, avoiding the omission problems of random sampling alone.

Stratified Sampling: Contaminants or other parameters are targeted. The site is subdivided and sampling patterns and densities varied in different areas. Stratified sampling can be used for complex and large sites, such as mining.

Grid or Systematic Sampling: The whole site is covered. Sampling locations are readily identifiable, which is valuable for follow-on sampling, if necessary. The grid does not have to be rectilinear. In fact, rectangles are not the best polygon to use in the value is to be representative of a cell. Circles provide equidistant representation, but overlap. Hexagons are sometimes used as a close approximation to the circle. The U.S. Environmental Monitoring and Assessment Program (EMAP) has used a hexagonal grid pattern, for example.

Judgmental Sampling: Samples are collected based on knowledge of the site. This overcomes the problem of ignoring sources or sensitive areas, but is vulnerable to bias of both inclusion and exclusion. Obviously, this would not be used for spatial representation, but for pollutant transport, plume characterization, or monitoring near a sensitive site (e.g., a day care center).

At every stage of monitoring from sample collection through analysis and archiving, only qualified and authorized persons should be in possession of the samples. This is usually ensured by requiring chain-of-custody manifests. Sample handling includes specifications on the temperature range needed to preserve the sample, the maximum amount of time the sample can be held before analysis, special storage provisions (e.g., some samples need to be stored in certain solvents), and chain-of-custody provisions (only certain, authorized persons should be in possession of samples after collection).

Each person in possession of the samples must require that recipient sign and date the chain-of-custody form before transferring the samples. This is because samples have evidentiary and forensic content, so any compromising of the sample integrity must be avoided.

5.3 LABORATORY ANALYSIS

Although real-time analysis of air and other media is becoming more commonplace, most environmental samples must be brought to a laboratory analyzed after collection. The steps that must be taken to interpret the concentration of a chemical in the sample are known as “wet chemistry.”

5.3.1 Extraction

When an environmental sample arrives at the laboratory, the next step may be “extraction.” Extraction is needed for two reasons. First, the environmental sample may be in sediment or soil, where the chemicals of concern are sorbed to particles and must be freed for analysis to take place. Second, the actual collection may have been by trapping the chemicals onto sorbents; meaning that the chemicals must first be freed from the sorbent matrix. Numerous toxic chemicals have low vapor pressures and may not be readily dissolved in water. Thus, they may be found in various media, for example, sorbed to particles, in the gas phase, or in the water column suspended to colloids (and very small amounts dissolved in the water itself). To collect such chemicals in the gas phase, a common method calls for trapping it on polyurethane foam (PUF). Thus, to analyze dioxins in the air, the PUF and particle matter must first be extracted, and to analyze dioxins in soil and sediment, those particles must also be extracted.

Extraction makes use of physics and chemistry. For example, many compounds can be simply extracted with solvents, usually at elevated temperatures. A common solvent

extraction is the Soxhlet extractor, named after the German food chemist, Franz Soxhlet (1848–1913). The Soxhlet extractor (the U.S. EPA Method 3540) removes sorbed chemicals by passing a boiling solvent through the media. Cooling water condenses the heated solvent, and the extract is collected over an extended period, usually several hours. Other automated techniques apply some of the same principles as solvent extraction, but allow for more precise and consistent extraction, especially when large volumes of samples are involved. For example, supercritical fluid extraction (SFE) brings a solvent, usually carbon dioxide to the pressure and temperature near its critical point of the solvent, where the solvent's properties are rapidly altered with very slight variations of pressure (Ekhtera et al., 1997). Solid phase extraction (SPE) uses a solid and a liquid phase to isolate a chemical from a solution, and is often used to clean up a sample before analysis. Combinations of various extraction methods can enhance the extraction efficiencies, depending on the chemical and the media in which it is found. Ultrasonic and microwave extractions may be used alone or in combination with solvent extraction. For example, the U.S. EPA Method 3546 provides a procedure for extracting hydrophobic (i.e., not soluble in water) or slightly water soluble organic compounds from particles such as soils, sediments, sludges, and solid wastes. In this method, microwave energy elevates the temperature and pressure conditions (i.e., 100–115°C and 50–175 psi) in a closed extraction vessel containing the sample and solvent(s). This combination can improve recoveries of chemical analytes and can reduce the time needed compared than the Soxhlet procedure alone.

5.3.2 Chromatography

Not every sample needs to be extracted. For example, air monitoring using canisters and bags allows the air to flow directly into the analyzer. Water samples may also be directly injected. Surface methods, such as fluorescence, sputtering, and atomic absorption, require only that the sample be mounted on specific media (e.g., filters). Also, continuous monitors such as the chemiluminescent system mentioned in the next section provide ongoing measurements.

Chromatography consists of separation and detection. Separation makes use of the chemicals' different affinities for certain surfaces under various temperature and pressure conditions. The first step, injection, introduces the extract to a "column." The term column is derived from the time when columns were packed with sorbents of varying characteristics, sometimes meters in length, and the extract was poured down the packed column to separate the various analytes. Today, columns are of two major types, gas and liquid. Gas chromatography (GC) makes use of hollow tubes (columns) coated inside with compounds that hold organic chemicals. The columns are in an oven, so that after the extract is injected into the column, the temperature is increased, as well as the pressure, and the various organic compounds in the extract are released from the column surface differentially, whereupon they are collected by a carrier gas (e.g., helium) and transported to the detector. Generally, the more volatile compounds are released first (they have the shortest retention times), followed by the semivolatile organic compounds. So, boiling point is often a very useful indicator as to when a compound will come off a column. This is not always the case, since other characteristics such as polarity can greatly influence a compound's resistance to be freed from the column surface. For this reason, numerous GC columns are available to the chromatographer (different coatings, interior diameters, and lengths). Rather than coated columns, liquid chromatography (LC) makes use of columns packed with different sorbing materials with differing affinities for compounds.

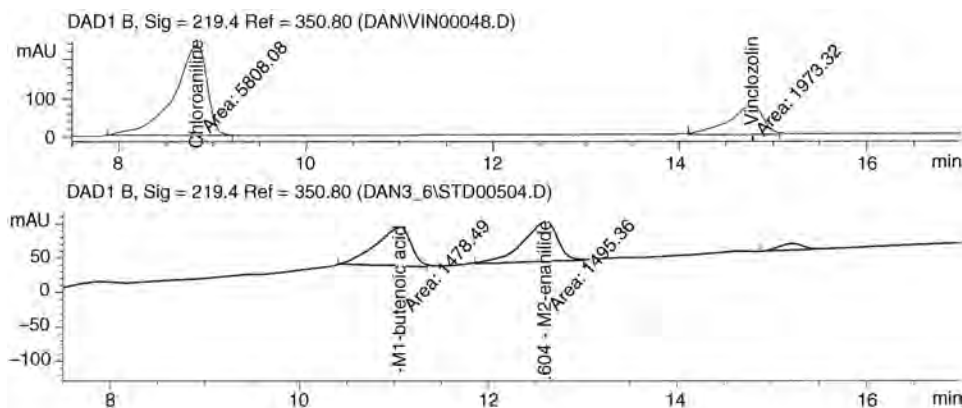


FIGURE 5.6 Chromatogram. Source: Vallero (2003.).

Also, instead of a carrier gas, LC uses a solvent or blend of solvents to carry the compounds to the detector. In the high-performance LC (HPLC), pressures are also varied.

Detection is the final step for quantifying the chemicals in a sample. The type of detector needed depends on the kinds of pollutants of interest. Detection gives the “peaks” that are used to identify compounds (Figure 5.6). For example, if hydrocarbons are of concern, GC with flame ionization detection (FID) may be used. GC-FID gives a count of the number of carbon atoms, so for example, long chains can be distinguished from short chains. The short chains come off the column first and have peaks that appear before the long-chain peaks. However, if pesticides or other halogenated compounds are of concern, electron capture detection (ECD) is a better choice.

Numerous detection approaches are also available for LC. Probably, the most common is absorption. Chemical compounds absorb energy at various levels, depending on their size, shape, bonds, and other structural characteristics. Chemicals also vary in whether they will absorb light or how much light they can absorb depending on wavelength. Some absorb very well in the ultraviolet (UV) range, while others do not. Diode arrays help to identify compounds by giving a number of absorption ranges in the same scan. Some molecules can be excited and will fluoresce. The *Beer–Lambert Law* states that energy absorption is proportional to chemical concentration:

$$A = eb [C] \quad (5.1)$$

where A is the absorbency of the molecule, e is the molar absorptivity (proportionality constant for the molecule), b is the light’s path length, and $[C]$ is the chemical concentration of the molecule. Thus, the concentration of the chemical can be ascertained by measuring the light absorbed.

One of the most popular detection methods for environmental pollutants is mass spectrometry (MS), which can be used with either GC or LC separation. The MS detection is highly sensitive for organic compounds and works by using a stream of electrons to consistently break apart compounds into fragments. The positive ions resulting from the fragmentation are separated according to their masses. This is referred to as the “mass to charge ratio” or m/z . No matter which detection device is used, software is used to decipher the peaks and to perform the quantitation of the amount of each contaminant in the sample.

For inorganic substances and metals, the additional extraction step may not be necessary. The actual measured media (e.g., collected airborne particles) may be measured by surface techniques such as atomic absorption (AA), X-ray fluorescence (XRF), inductively coupled plasma (ICP), or sputtering. As for organic compounds, the detection approaches can vary. For example, ICP may be used with absorption or MS. If all one needs to know is elemental information, for example, to determine total lead or nickel in a sample, AA or XRF may be sufficient. However, if speciation (i.e., knowing the various compounds of a metal), then significant sample preparation is needed, including a process known as derivatization. Derivatizing a sample is performed by adding a chemical agent that transforms the compound in question into one that can be recognized by the detector. This is done for both organic and inorganic compounds, for example, if the compound in question contains too many polar functional groups (e.g., $-OH$) it may not be detected with MS.

5.3.3 Limits of Detection

The physical and chemical characteristics of the compounds being analyzed must be considered before visiting the field and throughout all the steps in the laboratory. The quality of results generated about contamination depends on the sensitivity and selectivity of the analytical equipment. Table 5.3 defines some of the most important analytical chemistry threshold values.

TABLE 5.3 Expressions of Chemical Analytical Limits

Type of Limit	Description
Limit of detection (LoD)	Lowest concentration or mass that can be differentiated from a blank with statistical confidence. This is a function of sample handling and preparation, sample extraction efficiencies, chemical separation efficiencies, and capacity and specifications of all analytical equipment being used (see IDL below).
Instrument detection limit (IDL)	The minimum signal greater than noise detectable by an instrument. The IDL is an expression of the piece of equipment, not the chemical of concern. It is expressed as a signal to noise (S:N) ratio. This is mainly important to the analytical chemists, but the engineer should be aware of the different IDLs for various instruments measuring the same compounds, so as to provide professional judgment in contracting or selecting laboratories and deciding on procuring for appropriate instrumentation for all phases of remediation.
Limit of quantitation (LoQ)	The concentration or mass above which the amount can be quantified with statistical confidence. This is an important limit because it goes beyond the “presence–absence” of the LoD and allows for calculating chemical concentration or mass gradients in the environmental media (air, water, soil, sediment, and biota).
Practical quantitation limit (PQL)	The combination of LoQ and the precision and accuracy limits of a specific laboratory, as expressed in the laboratory’s quality assurance/quality control (QA/QC) plans and standard operating procedures for routine runs. The PQL is the concentration or mass that the engineer can consistently expect to have reported reliably.

Source: Vallero (2003.).

TABLE 5.4 Limits of Detection for Mercury and Monomethylmercury

Measurement	Method	Method Detection Limits
Total mercury in water	EPA 1631	0.121 ng/L (ppt)
Monomethylmercury in water	EPA 1630 with distillation	0.0192 ng/L (ppt)
Total mercury in sediment/soil	EPA 1631	0.302 ng/g (ppb)
Monomethylmercury in sediment/soil	EPA 1630 with extraction	0.0124 ng/g (ppb)
Total mercury in tissue	EPA 1631	0.378 ng/g (ppb)
Monomethylmercury in tissue	EPA 1630 with digestion	1.29 ng/g (ppb)

Source: Pacific Northwest National Laboratory and Mercury Analytical Laboratory (2009).

The limit of detection is the lowest concentration or mass that can be differentiated from a blank with statistical confidence, which is a function of sample handling and preparation, sample extraction efficiencies, chemical separation efficiencies, and capacity and specifications of all analytical equipment being used. For example, sampling aerosols around a bioreactor is limited to about 0.002–100 μm aerodynamic diameter, since this is the lower limit of detection using a condensation nucleus counter, and that assumes proper sampling prior to the analysis.

The method detection limit (MDL) is the minimum concentration of a substance that can be measured and reported with 99% confidence that the analyte concentration is greater than zero and is determined from analysis of a sample in a given matrix containing the analyte.¹ The MDL is expected to apply to a wide range of environmental sample including water and wastewater effluent. As such, the MDL for an analytical procedure can vary according to the kind of sample taken and requires a complete, specific, and well-defined analytical method. The minimum limit (ML) is the lowest concentration at which an entire analytical system must give a recognizable signal and acceptable calibration point for the analyte. It is equivalent to the concentration of the lowest calibration standard, so long as the system employs all method-specified sample weights, volumes, and cleanup procedures. The ML is calculated as the product of the MDL by 3.18 and rounding the result to the number nearest to $(1, 2, \text{ or } 5) \times 10n$, where n is an integer.

Many states have adopted water quality criteria for the protection of aquatic life and human health that require chemical measurements at very low concentrations. For example, they may limit mercury (Hg) in the range of 1–50 ppt, although current federal methods in the United States do not detect or quantify Hg in this range. A nondetect result using the U.S. Environmental Protection Agency Method 1631, for example, would show only that Hg concentrations are below 200 ppt but would not establish that they are at or below the applicable water quality criterion (Hanlon, 2007). The current limits of detection for Hg are shown in Table 5.4. Note that these are all below the lowest ambient water quality criterion for Hg of 12 ng/L² (nationwide)² and 1.3 ng/L² (Great Lakes)³.

¹United States Code of Federal Regulations: 40 CFR 131.6, Appendix B—Definition and procedure for determination of the method detection limit—Revision 1.11.

²United States Code of Federal Regulations: 40 CFR 132.6, Table 4.

³United States Code of Federal Regulations: 40 CFR 131.6. The method detection limit, according to 40 CFR 136, Appendix B, is 0.2 ng L⁻¹ and the minimum level of quantitation is 0.5 ng L⁻¹.

Following are the steps needed to find the limit of detection:⁴

1. Select an estimated detection limit using one of the following
 - (a) The concentration value corresponding to an instrument signal/noise ranging from 2.5 to 5.
 - (b) The concentration equivalent of three times the standard deviation of replicate instrumental measurements of the analyte in reagent water.
 - (c) That region of the standard curve where there is a significant change in sensitivity, that is, a break in the slope of the standard curve.
 - (d) Instrumental limitations.
 2. Prepare reagent (blank) water that is as free of analyte as possible. Reagent or interference free water is defined as a water sample in which analyte and interferent concentrations are not detected at the method detection limit of each analyte of interest. Analytes are the compounds that are being investigated, for example, if a pesticide is being analyzed. The analytes would include the active ingredient and the so-called inert ingredients (so-called, because the term “active” applies only to the particular biocidal mode of action that makes the pesticide efficacious against a particular pest, not to whether a chemical is chemically or biologically active). The interferents are the substances, when present may interfere with the particular analytical method and add to incorrect findings. Interferences are defined as systematic errors in the measured analytical signal of an established procedure caused by the presence of interfering chemical species. Thus, the interferent concentration is pre-supposed to be normally distributed in representative samples of a given matrix.
 3. (a) If the MDL is to be determined in reagent (blank) water, prepare a laboratory standard (analyte in reagent water) at a concentration that is at least equal to or in the same concentration range as the estimated method detection limit. (Recommend between one and five times the estimated method detection limit.) Proceed to Step 4.
 - (b) If the MDL is to be determined in another sample matrix, analyze the sample. If the measured level of the analyte is in the recommended range of one to five times the estimated detection limit, proceed to Step 4.
 - (c) If the measured level of analyte is less than the estimated detection limit, add a known amount of analyte to bring the level of analyte between one and five times the estimated detection limit.
- If the measured level of analyte is greater than five times the estimated detection limit, and there are two options.
- (i) Obtain another sample with a lower level of analyte in the same matrix if possible.
 - (ii) The sample may be used as is for determining the method detection limit if the analyte level does not exceed 10 times the MDL of the analyte in reagent water. The variance of the analytical method changes as the analyte concentration increases from the MDL; hence, the MDL determined under these circumstances may not truly reflect method variance at lower analyte concentrations.

⁴This procedure is taken directly from United States Code of Federal Regulations: 40 CFR 131.6, Appendix B—Definition and procedure for determination of the method detection limit—Revision 1.11.

4. (a) Take a minimum of seven aliquots of the sample to be used to calculate the method detection limit and process each through the entire analytical method. Make all computations according to the defined method with final results in the method reporting units. If a blank measurement is required to calculate the measured level of analyte, obtain a separate blank measurement for each sample aliquot analyzed. The average blank measurement is subtracted from the respective sample measurements.
- (b) It may be economically and technically desirable to evaluate the estimated method detection limit before proceeding with 4a. This will (1) prevent repeating this entire procedure when the costs of analyses are high and (2) ensure that the procedure is being conducted at the correct concentration. It is quite possible that an inflated MDL will be calculated from data obtained at many times the real MDL even though the level of analyte is less than five times the calculated method detection limit. To ensure that the estimate of the method detection limit is a good estimate, it is necessary to determine that a lower concentration of analyte will not result in a significantly lower method detection limit. Take two aliquots of the sample to be used to calculate the method detection limit and process each through the entire method, including blank measurements as described above in 4a. Evaluate these data
 - (i) If these measurements indicate the sample is in desirable range for determination of the MDL, take five additional aliquots and proceed. Use all seven measurements for calculation of the MDL.
 - (ii) If these measurements indicate the sample is not in correct range, re-estimate the MDL, obtain new sample as in 3 and repeat either 4a or 4b.
5. Calculate the variance (S^2) and standard deviation (S) of the replicate measurements as follows:

$$S^2 = \frac{1}{n-1} \left[\sum_{i=1}^n x_i^2 - \frac{(\sum_{i=1}^n X_i)^2}{n} \right] S = (S^2)^{1/2} \quad (5.2)$$

where X_i ; $i = 1$ to n , are the analytical results in the final method reporting units obtained from the n sample aliquots; and Σ refers to the sum of the X values from $i = 1$ to n .

6. (a) Compute the MDL:

$$\text{MDL} = (n-1, 1-\alpha=0.99)(S) \quad (5.3)$$

where MDL = the method detection limit; $t(n-1, 1-\alpha=0.99)$ = the student's t value appropriate for a 99% confidence level and a standard deviation estimate with $n-1$ degrees of freedom (see Table 5.5).

- (b) The 95% confidence interval estimates for the MDL derived in 6a are computed according to the following equations derived from percentiles of the χ^2 over degrees of freedom distribution (χ^2/df).

$$\text{LCL} = 0.64 \text{ MDL}$$

$$\text{UCL} = 2.20 \text{ MDL}$$

where LCL and UCL are the lower and upper 95% confidence limits, respectively based on seven aliquots.

TABLE 5.5 Students' t Values at the 99% Confidence Level

Number of Replicates	Degrees of Freedom ($n - 1$)	$t_{(n-1, 0.99)}$
7	6	3.143
8	7	2.998
9	8	2.896
10	9	2.821
11	10	2.764
16	15	2.602
21	20	2.528
26	25	2.485
31	30	2.457
61	60	2.390
00	00	2.326

Source: United States Code of Federal Regulations: 40 CFR 131.6, 49 FR 43430, FR 694, 696, FR 23703.

7. Optional iterative procedure to verify the reasonableness of the estimate of the MDL and subsequent MDL determinations.

- If this is the initial attempt to compute MDL based on the estimate of MDL formulated in Step 1, take the MDL as calculated in Step 6, spike the matrix at this calculated MDL and proceed through the procedure starting with Step 4.
- If this is the second or later iteration of the MDL calculation, use S^2 from the current MDL calculation and S^2 from the previous MDL calculation to compute the F-ratio. The F-ratio is calculated by substituting the larger S^2 into the numerator S^2A and the other into the denominator S^2B . The computed F-ratio is then compared with the F-ratio found in the table which is 3.05 as follows: if $S^2A/S^2B < 3.05$, then compute the pooled standard deviation by the following equation:

$$S_{\text{pooled}} = \left[\frac{6S^2A + 6S^2B}{12} \right]^{1/2} \quad (5.4)$$

If $S^2A/S^2B > 3.05$, respoke at the most recent calculated MDL and process the samples through the procedure starting with Step 4. If the most recent calculated MDL does not permit qualitative identification when samples are spiked at that level, report the MDL as a concentration between the current and the previous MDL that permits qualitative identification.

- Use the S_{pooled} as calculated in 7b to compute the final MDL according to the following equation

$$\text{MDL} = 2.681 (S_{\text{pooled}}) \quad (5.5)$$

where 2.681 is equal to $t(12, 1 - \alpha = 0.99)$.

- The 95% confidence limits for MDL derived in 7c are computed according to the following equations derived from percentiles of the χ^2 over degrees of

freedom distribution.

$$\text{LCL} = 0.72 \text{ MDL}$$

$$\text{UCL} = 1.65 \text{ MD}$$

where LCL and UCL are the lower and upper 95% confidence limits respectively based on 14 aliquots.

8. An example calculation of the MDL is

A laboratory conducts seven runs of a sample containing 0.1 ng/mL/Hg, with a SD = 0.007.

The standard deviation of 7 runs of 0.1 ng/mL = $0.007 \times \text{Student's } t \text{ value at the } 99\% \text{ confidence level}$. From Table 5.5., this is found to be 3.143.

Therefore, the MDL = $0.007 \text{ ng/mL} \times 3.143 = 0.022 \text{ ng/mL}$.

9. Report the results. The analytical method used must be specifically identified by number or title and the MDL for each analyte expressed in the appropriate method reporting units. If the analytical method permits options that affect the method detection limit, these conditions must be specified with the MDL value. The sample matrix used to determine the MDL must also be identified with MDL value. Report the mean analyte level with the MDL and indicate if the MDL procedure was iterated. If a laboratory standard or a sample that contained a known amount analyte was used for this determination, also report the mean recovery.

If the level of analyte in the sample was below the determined MDL or exceeds 10 times the MDL of the analyte in reagent water, do not report a value for the MDL.

The limit of detection is both an analytical and a sampling threshold. If an instrument can only detect down to 1 ppb, this is an analytical limitation. However, in reality, if the sample has been held for some time, or the sample must be extracted from the soil or trapping device in the field, this is a limit, even if the laboratory can detect down to 1 ppb. Statistical methods for dealing with nondetects are used, but a nondetect should never be reported as 0, since one can only say with confidence that it was not seen. It may not be present, but the scientist or engineer can only report what is known, and that is dictated by the limits of detection.

The performance is expressed in terms of precision, accuracy, specificity, false positives, and false negatives. Precision describes how refined and repeatable an operation can be performed, such as the exactness in the instruments and methods used to obtain a result. It is an indication of the uniformity or reproducibility of a result. This can be likened to shooting arrows⁵ with each arrow representing a data point. The spread of arrows is equally precise in targets A and B in Figure 5.7. Assuming that the center of the target, that is, the bull's eye, is the "true value," data set B is more accurate than A. If the archer is consistently missing the bull's eye in the same direction at the same distance, this is an example of bias (systematic error). This consistent deviation from the true value can be corrected by calibration and adjustments to equipment (e.g., by running known standards in our analytical equipment). To stay with the archery analogy, the archer would move her sight up and to the right.

⁵The target is a widely used way to describe precision and accuracy, so originator is unknown to author.

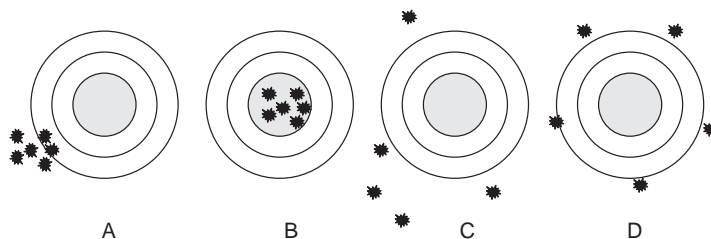


FIGURE 5.7 Precision and accuracy. The bull's eye represents the true value. Targets A and B demonstrate data sets that are precise; targets B and D data sets that are accurate, and targets C and D data sets that are imprecise. Target B is the ideal data set, which is precise and accurate.

5.3.4 Aerosol Limits of Detection

In the earlier Hg example, total mercury in water consists of the various chemical species of Hg (e.g., elemental— Hg^0 , alkylated—mono- or dimethylmercury) and whether they are in solution or part of the solids (suspended or sediment). However, in atmospheric sciences; the most important phase distribution is that of vapor versus particulate phase. The distinction between gases and vapors has to do with the physical phase that a substance would be under environmental conditions, for example, at standard temperature and pressure. Particulate matter (PM) is an expression of all particles, whether liquid or solid. An aerosol is a liquid or solid particle that is suspended in a gas; in environmental sciences this gas is usually air, but in reactors, stacks and other nonambient conditions, this can be various flue gases. Standard atmospheric conditions can be defined as 1 atm pressure (760 mmHg) and 25° (29°K) (U.S. Environmental Protection Agency, 1998).

The Clean Air Act established the national ambient air quality standards (NAAQS) for particulate matter in 1971, requiring measurements of total suspended particulates (TSP) as measured by a high volume sampler, that is, a device that collected a large range of sizes of particles (aerodynamic diameters up to $50\text{ }\mu\text{m}$). Smaller particles are more likely to be inhaled than larger particles, so in 1987 the U.S. EPA changed the standard for PM from TSP to PM_{10} , that is, particle matter $\leq 10\text{ }\mu\text{m}$ diameters. The diameter most often used for airborne particle measurements is the “aerodynamic diameter.” The aerodynamic diameter (D_{pa}) for all particles greater than $0.5\text{ }\mu\text{m}$ can be approximated as the product of the Stokes particle diameter (D_{ps}) and the square root of the particle density (ρ_p):

$$D_{\text{pa}} = D_{\text{ps}} \sqrt{\rho_p} \quad (5.6)$$

If the units of the diameters are in μm , the units of density are g/cm^3 .

The Stokes diameter D_{ps} is the diameter of a sphere with the same density and settling velocity as the particle. The Stokes diameter is derived from the aerodynamic drag force caused by the difference in velocity of the particle and the surrounding fluid. Thus, for smooth, spherical particles, the Stokes diameter is identical to the physical or actual diameter.

Aerosol textbooks provide methods to determine the aerodynamic diameter of particles less than $0.5\text{ }\mu\text{m}$. For larger particles, gravitational settling is more important and the aerodynamic diameter is often used.

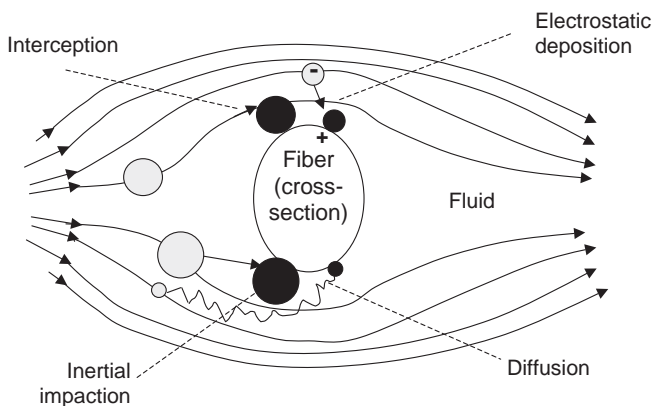


FIGURE 5.8 Mechanical processes important to filtration. *Source:* Vallero (2008); adapted from Rubow (2004).

The NAAQS for PM_{10} became a 24-h average of $150 \mu\text{g}/\text{m}^3$ (not to exceed this level more than once per year), and an annual average of $50 \mu\text{g}/\text{m}^3$ arithmetic mean. However, subsequent research showed the need to protect people breathing even smaller PM in air, since most of the particles that penetrate deeply into the air–blood exchange regions of the lung are quite small. Thus, in 1997, the U.S. EPA added a new fine particle (diameters ≤ 2.5), known as $\text{PM}_{2.5}$.⁶

Aerosols are collected using equipment that separates out the size fraction of concern. Filtration is an important technology in every aspect of environmental engineering, that is, air pollution, wastewater treatment, drinking water, and even hazardous waste and sediment cleanup. Basically, filtration consists of four mechanical processes: (1) diffusion; (2) interception; (3) inertial impaction; and (4) electrostatics (Figure 5.8).

Diffusion is important only for very small particles ($\leq 0.1 \mu\text{m}$ diameter) because the Brownian motion allows them to move away in a “random walk” away from the air stream. Interception works mainly for particles with diameters between 0.1 and $1 \mu\text{m}$. The particle does not leave the air stream but comes into contact with the filter medium (e.g., a strand of fiberglass). Inertial impaction collects particles that are sufficiently large to leave the air stream by inertia (diameters $\geq 1 \mu\text{m}$). Electrostatics consists of electrical interactions between the atoms in the filter and those in the particle at the point of contact (Van der Waal’s force), as well as electrostatic attraction (charge differences between particle and filter medium). Other important factors affecting filtration efficiencies include the thickness and pore diameter of the filter, the uniformity of particle diameters and pore sizes, the solid volume fraction, the rate of particle loading onto the filter (e.g., affecting particle “bounce”), the particle phase (liquid or solid), capillarity and surface tension (if either the particle or the filter media are coated with a liquid), and characteristics of air or other carrier gases, such as velocity, temperature, pressure, and viscosity.

Thus, aerosol measurement is an expression of the mass of a particle size of particle. Figure 5.9 shows an inlet of the $\text{PM}_{2.5}$ sampler that is designed to extract ambient aerosols

⁶For information regarding particle matter (PM) health effects and inhalable, thoracic, and respirable PM mass fractions (see U.S. Environmental Protection Agency, 1996), Air Quality Criteria for Particulate Matter. Technical Report No. EPA/600/P-95/001aF, Washington, DC.

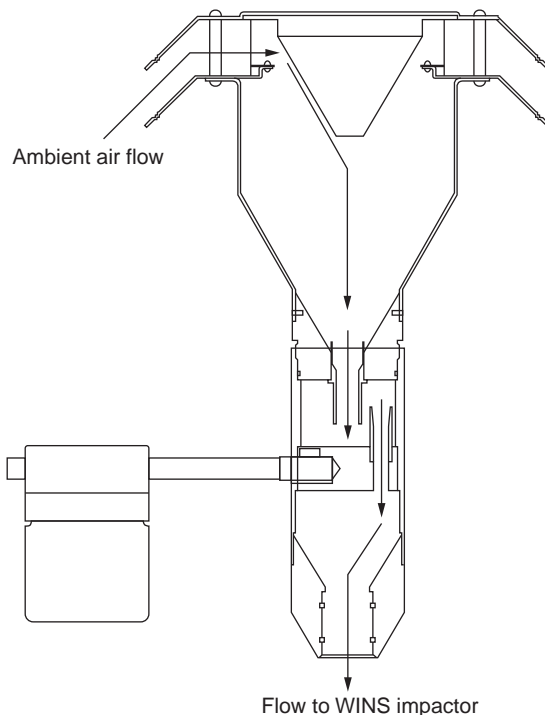


FIGURE 5.9 Flow of air through a sampler inlet head used to collect particulate matter with aerodynamic diameters $<2.5\ \mu\text{m}$ ($\text{PM}_{2.5}$). WINS = well impactor ninety-six, that is, the design of the particle impactor specified by the U.S. EPA for reference method samplers for $\text{PM}_{2.5}$. *Source:* U.S. EPA (1998).

from the surrounding airstream, remove particles with aerodynamic diameters $>10\ \mu\text{m}$, and move the remaining smaller particles to the next stage. Figure 5.10 illustrates the impactor and filter assembly for removing those particles $<10\ \mu\text{m}$ but $>2.5\ \mu\text{m}$ in diameter but allows particles of $2.5\ \mu\text{m}$ in diameter to pass and be collected on a filter surface. Particles $<10\ \mu\text{m}$ but $<2.5\ \mu\text{m}$ are removed downstream from the inlet by a single-stage, single-flow, single-jet impactor assembly. Aerosols are collected on filters that are weighed before and after sampling. This system uses 37 mm diameter glass filters immersed in low volatility, low viscosity diffusion oil. The oil is added to reduce the impact of “bounce,” that is, particle hit the filter and are not reliably collected (U.S. EPA, 1996).

The before and after weight difference represents the particulate mass in a given air volume, expressed in units of mass (e.g., ng) per units of volume (most often m^3), that is, ng/m^3 . This represents the mass concentration of aerosols in ambient air requires a sensitivity of $\pm 100\ \text{ng}$ for PM_{10} on standard $0.20 \times 0.25\ \text{m}^2$ filter and a sensitivity of $\pm 1\ \text{ng}$ for 3 mm or 47 mm diameter Teflon filters. The LD for the balance must be good enough to meet the atmospheric limits of detection of $60\ \text{ng}/\text{m}^3$ for PM_{10} and $40\ \text{ng}/\text{m}^3$ for $\text{PM}_{2.5}$. This assumes a 24-h sampling period on a standard filter for PM_{10} at a flow rate of $0.0189\ \text{m}^3/\text{s}$, and a 24-h sampling period on a 47 mm diameter filter for $\text{PM}_{2.5}$ at a flow rate of $2.78 \times 10^{-4}\ \text{m}^3/\text{s}$ (Solomon et al., 2001).

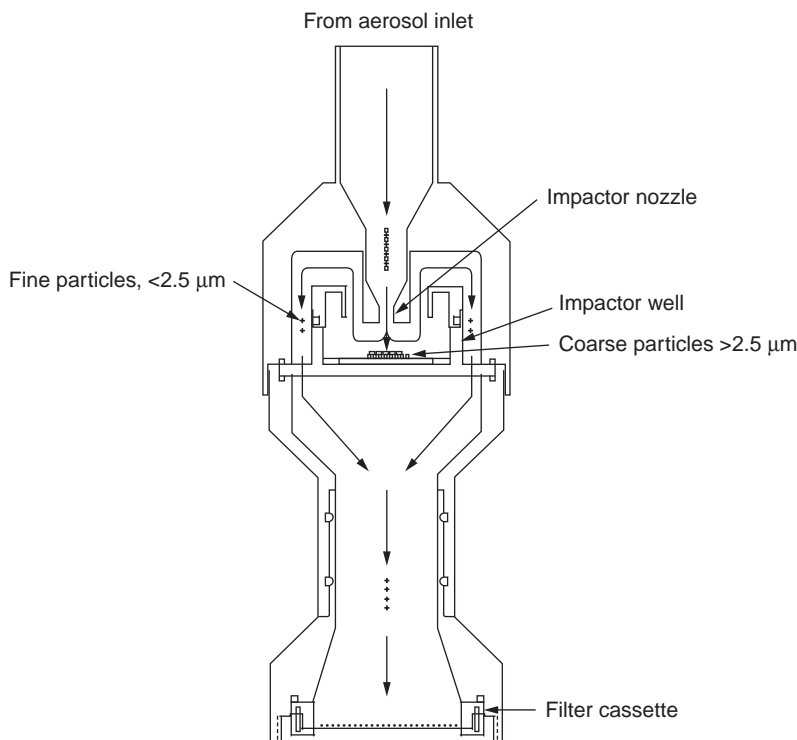


FIGURE 5.10 Flow of air through an impactor well and filter holder used to collect particulate matter with aerodynamic diameters $<2.5\ \mu\text{m}$ ($\text{PM}_{2.5}$). *Source:* U.S. EPA (1998).

Preweighing must account the relative humidity (i.e., water in the filter adds mass), not only during weighing but for a substantial amount of time (e.g., 24-h preceding period), as part of numerous calibration steps.

Precision and accuracy in aerosol measurements are greatly affected by collection efficiency. Accuracy is the same as that defined in the previous section, but precision for aerosols is defined the random variation among individual measurements of the same property, usually under prescribed identical conditions. For ambient particulate concentration measurements, precision is usually expressed in terms of a standard deviation estimated by collocated sampling or by reweighing filters and comparing reproducibility. Typical precision is within 5%.⁷

5.3.5 Microbial Limits of Detection

Regulatory agencies pursue ways to achieve lower levels of detection for microbes. As an example, the Invitrogen Corporation developed the PathAlert™ detection kits for the bacteria *Francisella tularensis* (*F. tularensis*), *Yersinia pestis* (*Y. pestis*), and *Bacillus anthracis* (*B. anthracis*) as part of the Environmental Technology Verification Program

⁷United States Code of Federal Regulations, Part 58, Appendix A, 1997; and U.S. Environmental Protection Agency (1998).

TABLE 5.6 Accuracy of Bacteria Test Kit Based on Percentage of Positive Results Compared to Total Replicates for Each Bacteria Species

Bacteria	Concentration Range of Samples Used in Accuracy Calculations (cfu/mL)	Overall Accuracy (Positive Results Out of Total Replicates)
<i>F. tularensis</i>	2×10^4 to 5×10^5	100% (20/20)
<i>Y. pestis</i>	2×10^2 to 5×10^3	100% (16/16)
<i>B. anthracis</i>	2×10^4 to 5×10^5	100% (16/16)

Source: U.S. Environmental Protection Agency and Battelle National Laboratory (2004).

(U.S. Environmental Protection Agency and Battelle National Laboratory, 2004). In the verification of the bacterial test, precision was determined from the overall percentage of consistent responses for all the sample sets. Responses were deemed to be consistent if all responses of the four replicates were the same. For *F. tularensis* replicates, 95% of the sample sets (20 out of 21) showed consistent results. Likewise, 95% of the sample sets (20 out of 21) were consistent for *Y. pestis*. For both, the inconsistency resulted from an inconclusive result for a drinking water (DW) replicate for each bacterium. *B. anthracis* also showed 95% consistency (20 out of 21), but the one sample set with inconsistent results was the infective dose in a performance test (PT) sample. Here, three of the four samples were inconclusive, but the fourth sample was positive for *B. anthracis*. The infective dose of *B. anthracis* was below the method level of detection (MLD) for this bacterium.

Accuracy was assessed by evaluating how often the results were positive in the presence of a concentration of contaminant above the MLD. Contaminant-only performance testing (PT) samples were used for this analysis. An overall percent agreement was determined by dividing the number of positive responses by the overall number of analyses of contaminant-only PT samples above the method LD. The results are presented in Table 5.6.

For *F. tularensis*, *Y. pestis*, and *B. anthracis*, all samples at concentration levels above the vendor-stated method LD generated positive responses for each set of replicates, resulting in 100% agreement for the overall accuracy of the detection kit for each bacterium. The infective/lethal dose for *Y. pestis* is 0.28 colony-forming units per milliliter (cfu/mL) and for *B. anthracis* is 200 cfu/mL. Both doses were below the MLD and not included in the accuracy calculations for those bacteria.

Specificity is the ability of a test to show a negative response when the contaminant is in fact absent. The specificity rate of this kit was determined by dividing the number of negative responses by the total number of unspiked samples. Unspiked interferent PT samples and unspiked DW samples were used to assess specificity. For *F. tularensis* and *Y. pestis*, one unspiked DW replicate for each bacterium produced an inconclusive response (Table 5.7).

A false positive response is defined as a detectable or positive test response when the agent is not present, in this case it was the interferent PT samples or DW samples that were not spiked. The false positive rate was the frequency of false positive results out of the total number of unspiked samples. The false negative response was defined as a negative response when the sample was spiked with a contaminant at a concentration greater than the MLD. Spiked PT (contaminant and interferent) samples and spiked DW samples were included in the analysis.

TABLE 5.7 Specificity of Bacteria Test Kit Based on Percentage of Negative Results Compared to Total Replicates for Each Bacteria Species

Bacteria	Overall Specificity (Negative Results Out of Total Replicates)
<i>F. tularensis</i>	96% (23/24)
<i>V. pestis</i>	96% (23/24)
<i>B. anthracis</i>	100% (22/22)

Source: U.S. Environmental Protection Agency and Battelle National Laboratory (2004).

TABLE 5.8 False Positive and False Negative Response Rates of Bacteria Test Kit

Bacteria	False Positive Rate	False Negative Rate
<i>F. tularensis</i>	0/24	0/60
<i>Y. pestis</i>	0/24	0/56
<i>B. anthracis</i>	0/22	0/56

Source: U.S. Environmental Protection Agency and Battelle National Laboratory (2004).

Conversely, the false negative rate was reported as the frequency of false negative results out of the total number of spiked samples for a particular contaminant. The results are presented in Table 5.8. No false positives or false negatives were found for any of the sample matrices for any bacteria for this kit. One replicate for unspiked DW in two different DW samples showed an inconclusive result for *F. tularensis* and one for *Y. pestis*. Two inconclusive results were reported; one for *F. tularensis* in one DW sample and one for *Y. pestis* in a different DW sample (See Table 5.8).

5.4 MEASUREMENT UNCERTAINTY

Contaminant assessments have numerous sources of uncertainty. There are two basic types of uncertainty: Types A and B. Type A uncertainties result from the inherent unpredictability of complex processes that occur in nature. These uncertainties cannot be eliminated by increasing data collection or enhancing analysis. The scientist and engineer must simply recognize that Type A uncertainty exists, but must not confuse it with Type B uncertainties, which can be reduced by collecting and analyzing additional scientific data.

The first step in an uncertainty analysis is to characterize potential uncertainties. Sources of Type B uncertainty take many forms (Finkel, 1990). There can be substantial uncertainty concerning the numerical values of the attributes being studied (e.g., contaminant concentrations, wind speed, discharge rates, groundwater flow, and other variables). Modeling generates its own uncertainties, including errors in selecting the variables to be included in the model, such as surrogate contaminants that represent whole classes of compounds (e.g., does benzene represent the behavior or toxicity of other aromatic compounds?). The application of the findings, even if the results themselves have tolerable uncertainty, leads to the propagation of uncertainties when ambiguity arises regarding their meaning. For example, a decision rule is a statement about which alternative will be selected, for example, for cleanup, based on the characteristics of the decision situation.

A “decision-rule uncertainty” occurs when there are disagreements or poor specification of objectives (i.e., is our study really addressing the client’s needs?).

Variability and uncertainty must not be confused. Variability consists of measurable factors that differ across populations such as soil type, vegetative cover, or body mass of individuals in a population. Uncertainty consists of unknown or not fully known factors that are difficult to measure, such as the inability to access an ideal site that would be representative because it is on private property.

Modeling uncertainties, for example, may consist of extrapolations from a single value to represent a whole population, that is, a point estimate (e.g., 70 kg as the weight of an adult male). Such estimates can be typical values for a population or an estimate of an upper end of the population’s value, for example, 70 years as the duration of exposure used as a “worse case” scenario. Another approach is known as the Monte Carlo technique (Figure 5.11). The Monte Carlo-type exposure assessments use probability distribution functions, which are statistical distributions of the possible values of each population characteristic according to the probability of the occurrence of each value. These are derived using iterations of values for each population characteristic. While the Monte Carlo technique may help to deal with the point estimate limitation, it can suffer from confusing variability with uncertainty.

Other data interpretation uncertainties can result from the oversimplification of complex entities. For example, assessments consist of an aggregation of measurement data, modeling, and combinations of sampling and modeling results. However, these

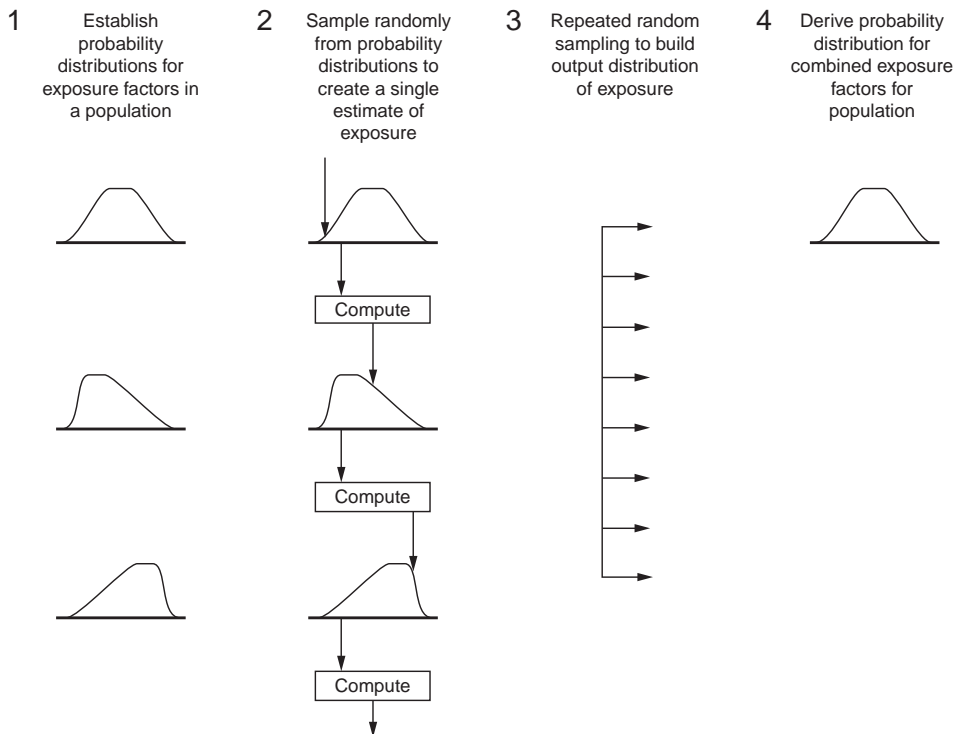


FIGURE 5.11 Principles of the Monte Carlo method for aggregating data. *Source:* Australian Department of Health and Ageing (2002).

complicated models are providing only a snapshot of highly dynamic human and environmental systems. The use of more complex models does not necessarily increase precision, and extreme values can be improperly characterized. For example, a 50th percentile value can always be estimated with more certainty than a 99th percentile value.

The bottom line is that uncertainty is always present in sampling, analysis, and data interpretation, so the monitoring and data reduction plan should be systematic and rigorous. The uncertainty analysis must be addressed for each step of the contaminant assessment process, including any propagation and enlargement of cumulative error (e.g., an incorrect pH value that goes into an index where pH is weighted heavily, and then used in another algorithm for sustainability). The characterization of the uncertainty of the assessment includes selecting and rejecting data and information ultimately used to make environmental decisions, and includes both qualitative and quantitative methods (See Table 5.9).

Uncertainty factors (UFs) are applied to address both the inherent and study uncertainties upon which to establish safe levels of exposure to contaminants. The UFs consider the uncertainties resulting from the variation in sensitivity among the members of the populations, including interhuman and intraspecies variability; the extrapolation of animal data to humans (i.e., interspecies variability); the extrapolation from data gathered

TABLE 5.9 Example of an Uncertainty Table for Exposure Assessment

<i>Assumption</i>	Effect on Exposure ^a		
	Potential magnitude for over-estimation of exposure	Potential magnitude for under-estimation of exposure	Potential magnitude for over- or under-estimation of exposure
<i>Environmental sampling and analysis</i>			
Sufficient samples may not have been taken to characterize the media being evaluated, especially with respect to currently available soil data			Moderate
Systematic or random errors in the chemical analyses may yield erroneous data			Low–high
<i>Exposure parameter estimation</i>			
The standard assumptions regarding body weight, period exposed, life expectancy, population characteristics, and lifestyle may not be representative of any actual exposure situation			Moderate
The amount of media intake is assumed to be constant and representative of the exposed population	Moderate		
Assumption of daily lifetime exposure for residents	Moderate to high		

Source: Australian Department of Health and Ageing (2002).

^aAs a general guideline, assumptions marked as “low” may affect estimates of exposure by less than one order of magnitude, assumptions marked as “moderate” may affect estimates of exposure by between one and two orders of magnitude, and assumptions marked as “high” may affect estimates of exposure by more than two orders of magnitude.

in a study with less-than-lifetime exposure to lifetime exposure, that is, extrapolating from acute or subchronic to chronic exposure; the extrapolation from different thresholds, such as the LOAEL rather than from a NOAEL; and the extrapolation from an incomplete data base is incomplete. Note that most of these sources of uncertainty have a component associated with measurement and analysis.

The numerical value uncertainties are directly related to the quality and representativeness of the sampling design and the analytical expressions described in Table 5.9. When these values are used in environmental models, they are known as “parameter uncertainties.” They account for imprecision and inaccuracy associated with the measurement and analytical equipment, systemic weaknesses in data gathering (i.e., bias).

5.5 MEASUREMENT DECISION MAKING

Environmental contaminants may be physical, chemical, or biological. Heat is physical contaminant since every species has a unique range of temperature tolerance. Chemical contaminants are probably the first type that comes to mind, since they are often measured in surface and well water, soil, air, and bodily fluids. Biological contaminants may be pathogenic microbes (e.g., fecal coliform bacteria in water), but they may also include organisms that upset environmental conditions, for example, animals such as the zebra mussels that destroy biodiversity in the Great Lakes or plants such as kudzu that cover large swaths of ecosystems in the southeastern United States.

The value of a water body can be directly related to water temperature since it is directly proportional to dissolved oxygen content, which is a limiting factor of the type of fish communities that can be supported by a water body (see Tables 5.10 and 5.11). A trout stream is a highly valued resource that is adversely impacted if mean temperatures increase. Rougher, less valued fish (e.g., carp and catfish) can live at much lower ambient water body temperatures than can salmon, trout and other coldwater fish populations. Thus, net increase in heat may directly stress the game fish population. That is, fish species vary in their ability to tolerate higher temperatures, meaning that the less tolerant, higher value fish will be inordinately threatened.

The threat may not be completely explained as heat stress due directly to the increase in temperature (U.S. EPA, 1997). Much can be explained by the concomitant decrease in the stream’s dissolved oxygen (DO) concentrations (see Figures 5.12 and 5.13), which deems the water body hostile to the fish. Even if the adult fish can survive at the reduced DO levels, their reproductive capacities decreases. Or, the reproduction is not adversely affected, but the survival of juvenile fish can be reduced.

The increased temperature can also increase the solubility of substances toxic to organisms, which increases the exposure. For example, greater concentrations of mercury and other toxic metals will occur with elevated temperatures. The lower DO concentrations will lead to a reduced environment where the metals and compounds will form sulfides, and other compounds that can be toxic to the fish. Thus, the change in temperature, the resulting decrease in DO and increasing metal concentrations, and the synergistic impact of the combining the hypoxic water and reduced metal compounds is a cascade of harm to the stream’s ecosystems (Figure 5.14).

Biota also plays a role in the heat-initiated effect. Combined abiotic and biotic responses occur. Notably, the growth and metabolism of the bacteria results in even more rapidly decreasing DO levels. Algae both consume DO for metabolism and produce DO

TABLE 5.10 Relationship Between Water Temperature and Maximum Dissolved Oxygen (DO) Concentration in Water (at 1 atm)

Temperature (°C)	Dissolved Oxygen (mg/L)	Temperature (°C)	Dissolved Oxygen (mg/L)
0	14.60	23	8.56
1	14.19	24	8.40
2	13.81	25	8.24
3	13.44	26	8.09
4	13.09	27	7.95
5	12.75	28	7.81
6	12.43	29	7.67
7	12.12	30	7.54
8	11.83	31	7.41
9	11.55	32	7.28
10	11.27	33	7.16
11	11.01	34	7.16
12	10.76	35	6.93
13	10.52	36	6.82
14	10.29	37	6.71
15	10.07	38	6.61
16	9.85	39	6.51
17	9.65	40	6.41
18	9.45	41	6.41
19	9.26	42	6.22
20	9.07	43	6.13
21	8.90	44	6.04
22	8.72	45	5.95

Sources: Vallero (2010) and U.S. Environmental Protection Agency (1997).

by photosynthesis. The increase in temperature increases their aqueous solubility, and the decrease in DO is accompanied by redox changes, for example, formation of reduced metal species, such as metal sulfides. This is also being mediated by the bacteria, some of which will begin reducing the metals as the oxygen levels drop (reduced conditions in the

TABLE 5.11 Normal Temperature Tolerances of Aquatic Organisms

Organism	Taxonomy	Range in Temperature Tolerance (°C)	Minimum Dissolved Oxygen (mg/L)
Trout	<i>Salma</i> , <i>Oncorhynchus</i> , and <i>Salvelinus</i> spp.	5–20	6.5
Smallmouth bass	<i>Micropertus dolomieu</i>	5–28	6.5
Caddisfly larvae	<i>Brachycentrus</i> spp.	10–25	4.0
Mayfly larvae	<i>Ephemerella invaria</i>	10–25	4.0
Stonefly larvae	<i>Pteronarcys</i> spp.	10–25	4.0
Catfish	Order Siluriformes	20–25	2.5
Carp	<i>Cyprinus</i> spp.	10–25	2.0
Water boatmen	<i>Notonecta</i> spp.	10–25	2.0
Mosquito larvae	Family Culicidae	10–25	1.0

Source: Vallero (2010) and Vernier Corporation (2009).

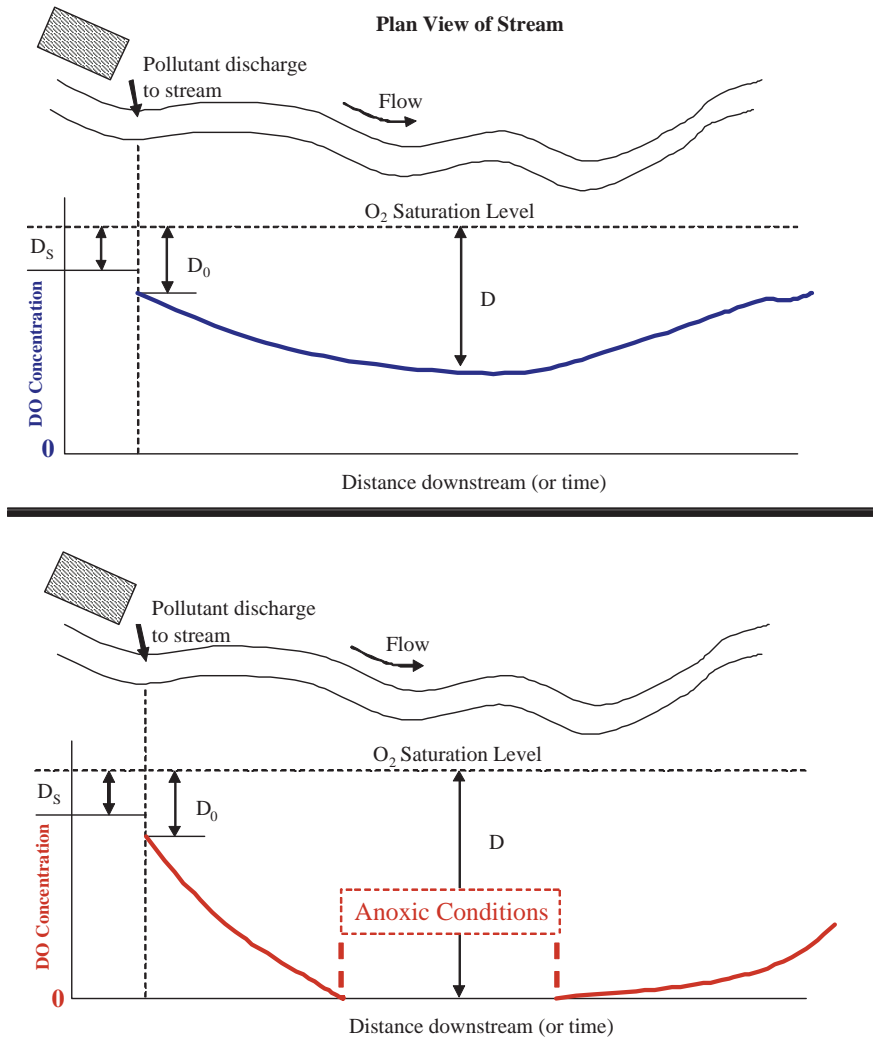


FIGURE 5.12 Dissolved oxygen (DO) deficit downstream from a heated effluent. The increased temperature can result in an increase in microbial kinetics, as well as more rapid abiotic chemical reactions, both consuming DO. The concentration of dissolved oxygen in the top curve remains above 0, so although the DO decreases, the overall system DO recovers. The bottom sags where dissolved oxygen falls to 0, and anoxic conditions result and continue until the DO concentrations begin to increase. D_s is the background oxygen deficit before the pollutants enter the stream. D_0 is the oxygen deficit after the pollutant is mixed. D is the deficit for contaminant A that may be measured at any point downstream. The deficit is overcome more slowly in the lower curve (smaller slope) because the re-oxygenation is dampened by the higher temperatures and changes to microbial system, which means the system has become more vulnerable to another insult, for example, another downstream source could cause the system to return to anoxic conditions.

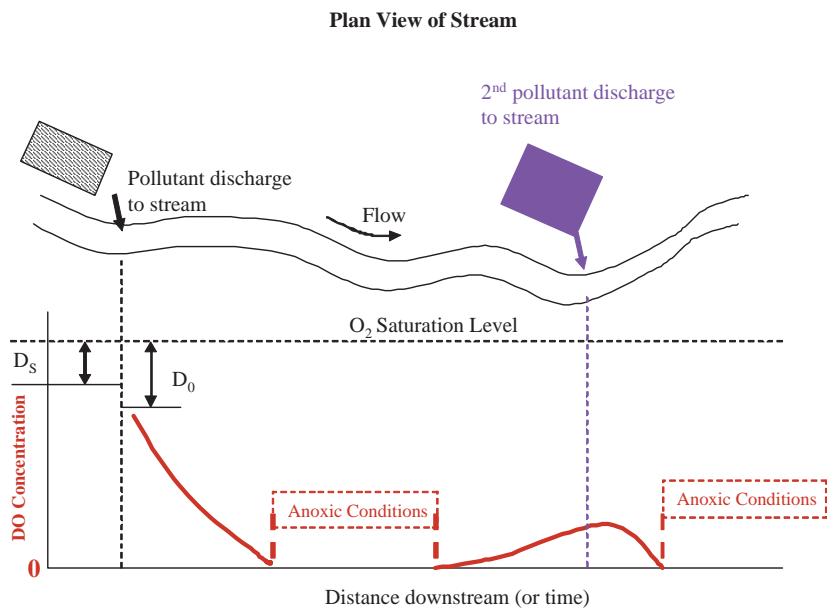


FIGURE 5.13 Cumulative effect of a second heat source, causing an overall system to become more vulnerable. The rate of re-oxygenation is suppressed, with a return to anoxic conditions.

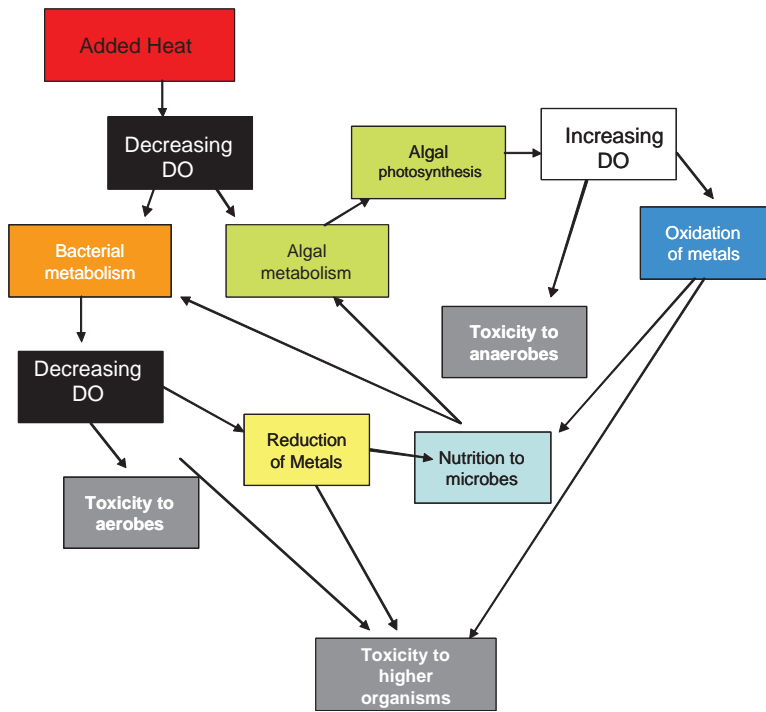


FIGURE 5.14 Adverse effects in the real world usually result from a combination of conditions. In this example, the added heat results in an abiotic response (i.e. decreased dissolved oxygen (DO) concentrations in the water). *Source:* Vallero (2010).

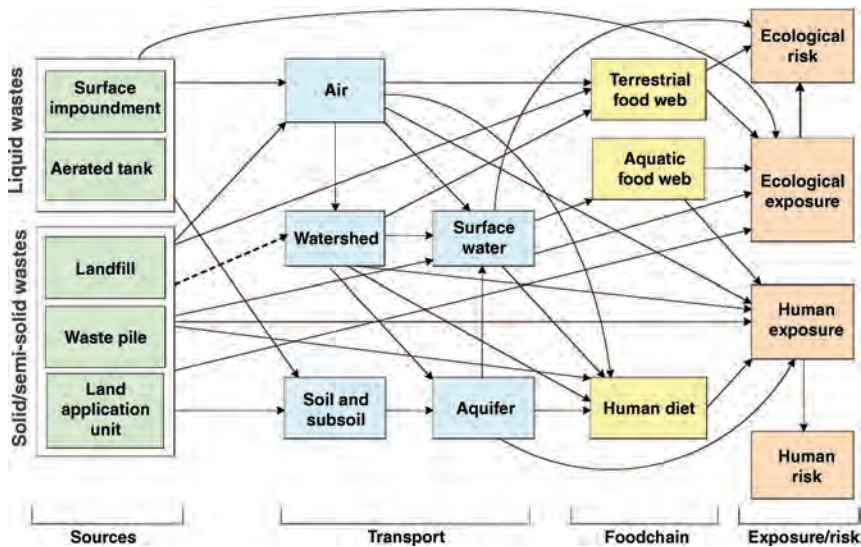


FIGURE 5.15 Environmental transport pathways can be affected by net heat gain. Compounds (nutrients, contaminants), microbes and energy (e.g., heat) follow the path through the environment indicated by arrows. The residence time within in any of the boxes is affected by conditions, including temperature. Adapted from Vallero et al. (2007).

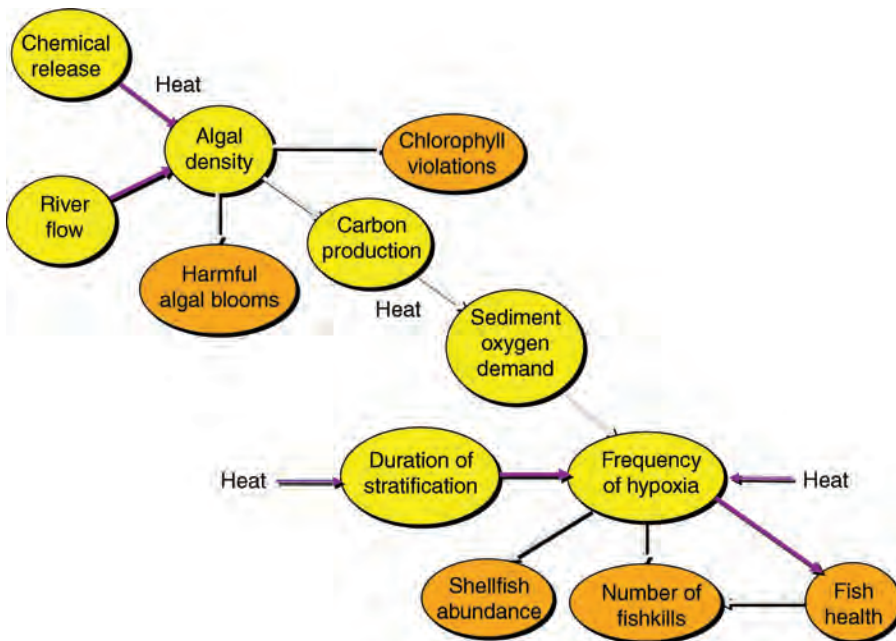


FIGURE 5.16 Flow of events and conditions leading to fish kills, indicating some of the points where added heat can exacerbate the likelihood of a fish kill or other adverse environmental event Vallero et al. (2007). Graphics adapted from U.S. Environmental Protection Agency (2007).

water and sediment). However, the opposite is true in the more oxidized regions, that is, the metals are forming oxides. The increase in the metal compounds combined with the reduced DO, combined with the increased temperatures can act synergistically to make the conditions toxic for higher animals, for example, a fish kill.

The first-order abiotic effect (i.e., increased temperature) results in an increased microbial population. However, the growth and metabolism of the bacteria results in decreasing the DO levels, but the growth of the algae both consume DO for metabolism and produce DO by photosynthesis. Meanwhile, a combined abiotic and biotic response occurs with the metals. The increase in temperature increases their aqueous solubility and the decrease in DO is accompanied by redox changes, for example, formation of reduced metal species, such as metal sulfides. This is also being mediated by the bacteria, some of which will begin reducing the metals as the oxygen levels drop (reduced conditions in the water and sediment). However, the opposite is true in the more oxidized regions, that is, the metals are forming oxides. The increase in the metal compounds combined with the reduced DO, combined with the increased temperatures can act synergistically to make the conditions toxic for higher animals, for example, a fish kill (Vallero et al., 2007). Predicting the likelihood of a fish kill can be quite complicated, with many factors that either mitigate or exacerbate the outcome (see Figures 5.15 and 5.16).

5.6 ENVIRONMENTAL INDICATORS

Measurements alone are seldom sufficient to determine environmental conditions. These data must often be combined with other information in the form of indicators and indices.

5.6.1 Oxygen Indicators

The biochemical oxygen demand (BOD) is the amount of oxygen that bacteria will consume in the process of decomposing organic matter under aerobic conditions. The BOD is measured by incubating a sealed sample of water for 5 days and measuring the loss of oxygen by comparing the O_2 concentration of the sample at time = 0 (just before the sample is sealed) to the concentration at time = 5 days (known specifically as BOD_5). Samples are commonly diluted before incubation to prevent the bacteria from depleting all of the oxygen in the sample before the test is complete (State of Georgia, 2003). BOD_5 is merely the measured DO at the beginning time, that is, the initial DO (D_1), measured immediately after it is taken from the source, minus the DO of the same water measured exactly 5 days after D_1 , that is, D_5 :

$$BOD = \frac{D_1 - D_5}{P} \quad (5.7)$$

where P = decimal volumetric fraction of water utilized. D units are in mg/L. If the dilution water is seeded, the calculation becomes

$$BOD = \frac{(D_1 - D_5) - (B_1 - B_5)f}{P} \quad (5.8)$$

where B_1 = initial DO of seed control; B_5 = final DO of seed control; and f = the ratio of seed in sample to seed in control = % seed in D_1 / % seed in B_1 . B units are in mg/L.

For example, to find the BOD₅ value for a 10 mL water sample added to 300 mL of dilution water with a measured DO of 7 mg/L and a measured DO of 4 mg/L 5 days later

$$P = \frac{10}{300} = 0.03$$

$$\text{BOD}_5 = \frac{7 - 4}{0.03} = 100 \text{ mg/L}$$

Thus, the microbial population in this water is demanding 100 mg/L dissolved oxygen over the 5-day period. So, if a conventional municipal wastewater treatment system is achieving 95% treatment efficiency, the effluent discharged from this plant would be 5 mg/L.

Chemical oxygen demand (COD) does not differentiate between biologically available and inert organic matter, and it is a measure of the total quantity of oxygen required to oxidize all organic material completely to carbon dioxide and water. COD values always exceed BOD values for the same sample. COD (mg/L) is measured by oxidation using potassium dichromate (K₂Cr₂O₇) in the presence of sulfuric acid (H₂SO₄) and silver. By convention, 1 g of carbohydrate or 1 g of protein accounts for about 1 g of COD. On average, the ratio BOD:COD is 0.5. If the ratio is <0.3, the water sample likely contains elevated concentrations of *recalcitrant* organic compounds, that is, compounds that resist biodegradation (Gerba and Pepper, 2009). That is, there are numerous carbon-based compounds in the sample, but the microbial populations are not efficiently using them for carbon and energy sources. This is the advantage of having both BOD and COD measurements. Sometimes, however, COD measurements are conducted simply because they require only a few hours compared with the 5 days for BOD.

Since available carbon is a limiting factor, the carbonaceous BOD reaches a plateau, that is, ultimate carbonaceous BOD (see Figure 5.17). However, carbonaceous compounds are the only substances demanding oxygen. Microbial populations will continue to demand O₂ from the water to degrade other compounds, especially nitrogenous

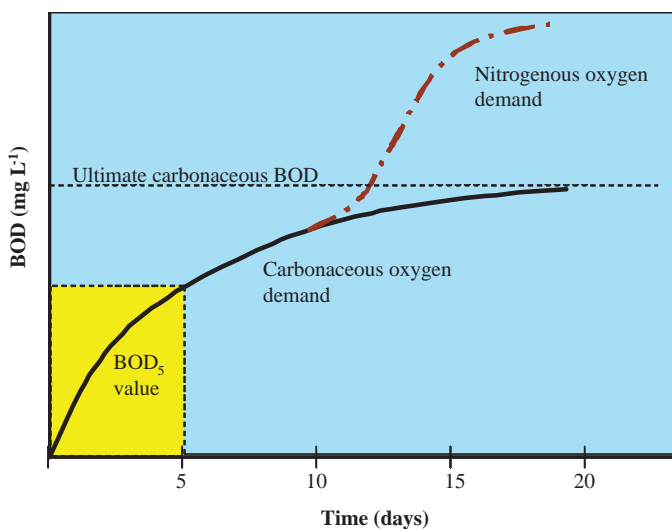


FIGURE 5.17 Biochemical oxygen demand (BOD) curve showing ultimate carbonaceous BOD and nitrogenous BOD. Adapted from Gerba and Pepper (2009).

compounds, which account for the bump in the BOD curve. Thus, in addition to serving as an indication of the amount of molecular oxygen needed for biological treatment of the organic matter, BOD also provides a guide to sizing a treatment process, assign its efficiency and giving operators and regulators information about whether the facility is meeting its design criteria and is complying with pollution control permits.

If effluent with high BOD concentrations reaches surface waters, it may diminish DO to levels lethal some fish and many aquatic insects. As the water body reaerates as a result of mixing with the atmosphere and by algal photosynthesis, O_2 is added to the water, the oxygen levels will slowly increase downstream. The drop and rise in DO concentrations downstream from a source of BOD is known as the DO sag curve, because the concentration of dissolved oxygen “sags” as the microbes deplete it. So, the falling O_2 concentrations fall with both time and distance from the point where the high BOD substances enter the water.

The stress from decreasing DO is usually indicated by the BOD. Similar to most environmental systems, the water bodies that receive sediment loads are complex in their response to increased input of materials. The DO will respond both positively and negatively to increased nutrient levels. Since the biota have unique optimal ranges of growth and metabolism that varies among species (e.g., algae will add some O_2 with photosynthesis, but use some O_2 for metabolism, whereas the bacteria will generally be net consumers of molecular oxygen).

5.6.2 Indices

The most widely applied environmental indices are those that follow the framework of an index of biological integrity. In biological systems, integrity is the capacity of a system to sustain a balanced and healthy community. This means the community of organisms in that system meets certain criteria for species composition, diversity and adaptability, often compared with a reference site that is a benchmark for integrity. As such, biological integrity indices are designed to integrate the relationships of chemical and physical parameters with each other and across various levels of biological organization. They are now used to evaluate the integrity of environmental systems using a range of metrics to describe system.

Thus, environmental indices combine attributes to determine a system's condition (e.g., diversity and productivity) and to estimate stresses. The original index of biotic integrity developed by Karr (1981) was based on fish fauna attributes and has provided predictions of how well a system will respond to a combination of stresses. In fact, the index is completely biological, with no direct chemical measurements. However, the metrics (see Table 5.12) are indirect indicators of physicochemical factors (e.g., the abundance of game fish is directly related to dissolved oxygen concentrations). The metrics provide descriptions of a system's structure and function.

An example of the data that is gathered to characterize a system is provided in Table 5.13. The information that is gleaned from these data is tailored to the physical, chemical, and biological conditions of an area, for example, for large spatial regions. The information from a biologically based index can be used to evaluate a system, as shown in Figure 5.18.

Systems involve scale and complexities in both biology and chemistry.

For example, a fish's direct aqueous exposure (AE in $\mu\text{g}/\text{day}$) is the product of the organism's ventilation volume, that is, the flow Q (in mL/day), and the compound's aqueous concentration, C_w ($\mu\text{g}/\text{mL}$). The fish's exposure by its diet (DE, in $\mu\text{g}/\text{day}$) is the

TABLE 5.12 Biological Metrics Used in the Original Index of Biological Integrity (IBI)

Integrity Aspect	Biological Metric
Species richness and composition	Total number of fish species (total taxa) Number of <i>Catostomidae</i> species (suckers) Number of darter species Number of sunfish species Number of intolerant or sensitive species
Indicator species metrics	Percent of individuals that are <i>Lepomis cyanellus</i> (Centrarchidae) Percent of individuals that are omnivores
Trophic function metrics	Percent of individuals that are insectivorous Cyprinidae Percent of individuals that are top carnivores or piscivores Percent of individuals that are hybrids
Reproductive function metrics	Abundance or catch per effort of fish
Abundance and condition metrics	Percent of individuals that are diseased, deformed, or that have eroded fins, lesions, or tumors (DELTs)

Source: Karr (1981).

product of its feeding rate, F_w (g wet weight/day), and the compound's concentration in the fish's prey, C_p ($\mu\text{g/g}$ wet wt). If the fish's food consist of single type of prey that is at equilibrium with the water, fish's aqueous and dietary exposures and the bioconcentration factor (BCF) can be calculated when they are equal:

$$AE = DE; QC_w = F_w C_p; BCF = \frac{Q}{F_w} \quad (5.9)$$

The ventilation-to-feeding ratio for a 1 kg trout has been found (Erickson and McKim, 1987; Stewart et al., 1983) to be on the order of $10^{4.3}$ mL/g. Assuming the quantitative structure activity relationship (QSAR) for the trout's prey is $BCF = 0.048$ times the octanol–water coefficient (K_{ow}); it appears that the trout's predominant route of exposure for any chemical with a $K_{ow} > 10^{5.6}$. Exposure must also account for the organism's assimilation of compounds in food, which for very lipophilic compounds, will probably account for the majority of exposure compared with that from the water. Even though chemical exchange occurs from both food and water through passive diffusion (Fick's law relationships), the uptake from food, unlike direct uptake from water, does not necessarily relax the diffusion gradient into the fish. The difference between digestion and assimilation of food can result in higher contaminant concentrations in the fish's gut.

Predicting uptake where the principal route of exchange is dietary is complicated in that most fish species exhibit well defined, size-dependent, taxonomic, and temporal trends regarding their prey. This means that a single bioaccumulation factor (BAF) may help to assess the risk to all fish species. In fact, the BAF may not even apply to different sizes of the same species.

The systematic biological exchange of materials between the organism, in this case various species of fishes, is known as uptake, which can be expressed by the following three differential equations for each age class or cohort of fish (Barber, 2008):

$$\frac{dB_f}{dt} = J_g + J_i + J_{bt} \quad (5.10)$$

TABLE 5.13 Biological Metrics that Apply to Various Regions of North America^a

Alternative IBI Metrics	Midwestern United States	Central Appalachians	Sacramento-San Joaquin	Colorado Front Range	Western Oregon	Ohio	Ohio Headwater Sites	Northwestern United States	Ontario	Central Corn Belt Plain	Wisconsin-Warmwater	Wisconsin-Coldwater	Maryland Coastal Plain	Maryland Non Tidal
1. Total number of species	X	X	X	X				X	X	X			X	X
Number of native fish species					X	X	X		X		X			
Number of salmonid age classes ^b				X	X									
2. Number of darter species	X	X		X		X				X	X			
Number of sculpin species					X									
Number of benthic insectivore species								X						
Number of darter and sculpin species							X				X			
Number of darter, sculpin, and madtom species														
Number of salmonid juveniles (individuals) ^b			X		X	X ^c		X						
Percentage of round-bodied suckers			X											
Number of sculpins (individuals)			X										X	
Number of benthic species				X		X				X	X			X
3. Number of sunfish species	X													
Number of cyprinid species					X									
Number of water column species								X						
Number of sunfish and trout species									X					
Number of salmonid species			X							X				
Number of headwater species							X							
Percentage of headwater species							X			X				
4. Number of sucker species	X				X	X		X		X	X			
Number of adult trout species ^b			X		X									
Number of minnow species				X			X			X				

(continued)

TABLE 5.13 (Continued)

Alternative IBI Metrics	Midwestern United States	Central Appalachians	Sacramento-San Joaquin	Colorado Front Range	Western Oregon	Ohio	Ohio Headwater Sites	Northwestern United States	Ontario	Central Corn Belt Plain	Wisconsin-Warmwater	Wisconsin-Coldwater	Maryland Coastal Plain	Maryland Non Tidal
Number of sucker and catfish species														
5. Number of intolerant species	X			X	X	X		X	X		X	X	X	X
Number of sensitive species							X			X				
Number of amphibian species		X												
Presence of brook trout									X			X		
Percentage of stenothermal cool and cold water species												X		
Percentage of salmonid ind. as brook trout												X		
6. Percentage of green sunfish	X													
Percentage of common carp					X									
Percentage of white sucker				X				X		X		X	X	X
Percentage of tolerant species						X								
Percentage of creek chub		X												
Percentage of dace species									X					
Percentage of eastern mudminnow													X	
7. Percentage of omnivores	X			X		X	X	X	X	X	X			
Percentage of generalist feeders		X												
Percentage of generalists, omnivores, and invertivores														X
8. Percentage of insectivorous cyprinids	X												X	
Percentage of insectivores					X			X		X	X		X	X ^d
Percentage of specialized insectivores		X		X										

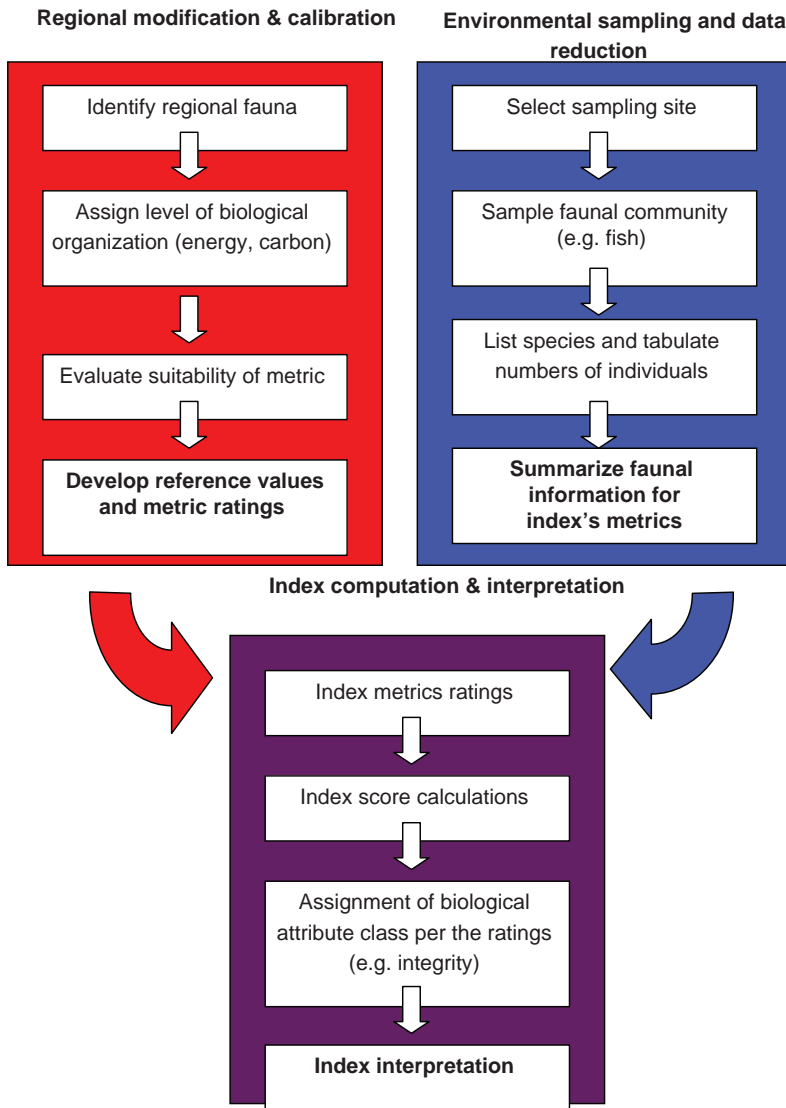


FIGURE 5.18 Sequence of activities involved in calculating and interpreting an Index of Biotic Integrity (IBI). Adapted from Barbour et al. (1999) and Karr (1987).

where J_g represents the net chemical exchange ($\mu\text{g/d}$) across the fish's gills from the water; J_i represents the net chemical exchange ($\mu\text{g/d}$) across the fish's intestine from food; and J_{bt} is the chemical's biotransformation rate ($\mu\text{g/d}$).

Physiologically based models for fish growth are often formulated in terms of energy content and flow (e.g., kcal/fish and kcal/day). Equation 5.4 is basically the same as such bioenergetic models because energy densities of fish depend on their dry weight (Kushlan et al., 1986; Hartman and Brandt, 1995; Schreckenbach et al., 2001). Obviously, feeding depends on the availability of suitable prey, so the mortality of the fish is a function of the individual feeding levels and population densities of its predators. Thus, the fish's dietary

exposure is directly related to the organism's feeding rate and the concentrations chemicals in its prey.

5.7 EXTENDING MEASUREMENT DATA USING MODELS

Models take many forms, from conceptual models that explain the way a system works, such as delineation of all the factors and parameters of how a particle moves in the air after release. In general, developing a model requires two main steps. First, a model of the domain and the processes being studied must be defined. Then, at the model boundaries, a model of the boundary conditions is especially needed to represent the influencing environment surrounding the study domain. Physical or dynamic models are used to estimate the location of chemical expected to move under controlled conditions, but on a much smaller scale than real life. These can include chambers to model the fate of chemical. Like all models, the dynamic model's accuracy is dictated by the degree to which the actual conditions can be simulated and the quality of the information that is used.

Numerical models apply mathematical expressions to approximate a system. Transport and fate models can be statistical and/or deterministic. Statistical models include the pollutant dispersion models, such as the Lagrangian models, which follow the movement of a substance starting from the source to the receptor locations. These often assume idealized Gaussian distributions from a point of release; that is, the pollutant concentrations are normally distributed in both the vertical and the horizontal directions from the source. The Lagrangian approach is common for atmospheric releases, and recent models based on this approach have incorporated additional descriptions of complex turbulence. Stochastic models are statistical models that assume that the events affecting the behavior of a chemical in the environment are random, so such models are based on probabilities. These are being commonly adopted in the modeling of human exposures.

Deterministic models are used when the physical, chemical, and other processes are sufficiently understood so as to be incorporated to reflect the movement and fate of chemicals. Often, they are difficult to develop because each process must be represented by a set of algorithms in the model. Also, the relationship between and among the systems, such as the kinetics and mass balances, must also be represented. Thus, the modeler must "parameterize" every important event following a chemical's release to the environment. Hybrid models use both statistical and deterministic approaches, for example, when one part of a system tends to be more random while another has a very strong basis in physical principles.

Numerous models are available to address the movement of chemicals through a single environmental medium, but increasingly, engineers applying multimedia models, such as compartmental models that help to predict the behavior and changes to chemicals as they move within and among reservoirs (e.g., soil in an ecosystem or carpet in a house), in the air as dust and vapors, and in exchanges with surfaces.

Measurement data can be used to evaluate these models. Conversely, once a model is considered to be reliable, measurements can be used as inputs to the model to represent larger areas than those actually measured. Measurements may also be input into models to predict different scenarios. The engineer may be able to use current and previous measurements to extrapolate environmental conditions in both space and time. This is very useful when deciding on ways to clean up pollution or to consider possible problems that may be encountered if a substance is used. In this way, models can be very useful to the environmental engineer who must choose from various interventions or who must consider the life cycle of an ingredient in a product. Thus, measurement interpolation and extrapolation with models is an important tool to predicting reliability, risk and in support of optimization.

5.8 SUMMARY

The environment consists of very complex systems ranging in scale from the cell to the planet. These systems are comprised of matrices of nonliving (i.e., abiotic) and living (biotic) components. Various means of measurement are required to determine the condition of such systems. Many of these measurements are direct physical and chemical measurements, for example, temperature, density, and pH of soil and water. Others are indirect, such as light scattering as indication of the number of aerosols in the atmosphere. This chapter provides an overview of some of the most important measurement methods in use today. In addition, the chapter introduces some of the techniques available for sampling, analysis, and extrapolation and interpolation (e.g., models) of measured results.

NOMENCLATURE

<i>A</i>	Absorption
<i>AA</i>	Atomic absorption
<i>AADT</i>	Annual average daily traffic
<i>AE</i>	Aquatic exposure
<i>B</i>	Final DO of seed control
<i>b</i>	Path length of light
<i>B₁</i>	Initial DO of seed control
<i>BAF</i>	Bioaccumulation factor
<i>BCF</i>	Bioconcentration factor
<i>BOD</i>	Biochemical oxygen demand
<i>cfu</i>	Colony-forming units
<i>cm</i>	Centimeter
<i>COD</i>	Chemical oxygen demand
<i>C_p</i>	Chemical concentration in fish prey
<i>C_w</i>	Compound's aqueous concentration
<i>[C]</i>	Chemical concentration
<i>°C</i>	Degrees Celsius
<i>D</i>	Aerodynamic diameter
<i>D₁</i>	Initial DO measurement on day 1
<i>D₅</i>	DO measurement on day 5
<i>DE</i>	Exposure by diet
<i>DELT</i>	Diseased, deformed, eroded fins, lesions, or tumors
<i>DO</i>	Dissolved oxygen
<i>D_{ps}</i>	Stokes particle diameter
<i>DQO</i>	Data quality objective
<i>DW</i>	Drinking water
<i>e</i>	Molar absorptivity
<i>ECD</i>	Electron capture detection
<i>f</i>	Ratio of seed in sample to seed in control
<i>FID</i>	Flame ionization detection
<i>F_w</i>	Feeding rate
<i>g</i>	Gram
<i>GC</i>	Gas chromatography
<i>GIS</i>	Geographic information system

HPLC	High-performance liquid chromatography
IBI	Index of biological integrity
ICP	Inductively coupled plasma
IDL	Instrument detection limit
J_{bt}	Biotransformation rate
J_g	Net chemical exchange across fish gill
J_i	Net chemical exchange across the fish's intestine from food
K_{ow}	Octanol–water coefficient
K	Kelvin
L	Liter
LC	Liquid chromatography
LDR	Land disposal regulation
LoD	Limit of detection
LoQ	Limit of quantitation
MDL	Method detection limit
mL	Milliliter
ML	Minimum limit
mmHg	Millimeters of mercury (pressure)
MS	Mass spectroscopy
MSAT	Mobile source air toxic
μg	Microgram
μm	Micron (micrometer)
NAAQS	National Ambient Air Quality Standards
ng	Nanogram
P	Decimal volumetric fraction of water
PM	Particulate matter
PM ₁₀	Particulate matter ≤ 10 micron aerodynamic diameter
PM _{2.5}	Particulate matter ≤ 2.5 micron aerodynamic diameter
ppb	Parts per billion
ppt	Parts per trillion
PQL	Practical quantitation limit
PT	Performance testing
PUF	Polyurethane foam
ρ_p	Particle density
Q	Flow
QA	Quality assurance
QC	Quality control
QSAR	Quantitative structure activity relationship
S	Standard deviation
LCL	Lower 95% confidence limit
S^2	Variance
SFE	Supercritical fluid extraction
SOP	Standard operating procedures
SPE	Solid phase extraction
TSP	Total suspended particulates
UCL	Upper 95% confidence limit
UF	Uncertainty factor
XRF	X-ray fluorescence

REFERENCES

- Australian Department of Health and Ageing. Environmental health risk assessment: Guidelines for assessing human health risks from environmental hazards; 2002.
- Barber MC. Bioaccumulation and Aquatic System Simulator (BASS). User's Manual, Version 2.2. Report No. EPA 600/R-01/035, update 2.2, March 2008. Athens, GA: U.S. Environmental Protection Agency; 2008.
- Barbour MT, Gerritsen J, Snyder BD, Stribling JB. *Rapid Bioassessment Protocols for Use in Streams and Wadeable Rivers: Periphyton, Benthic Macroinvertebrates and Fish*. 2 ed. Report No. EPA 841-B-99-002. Washington, DC: U.S. Environmental Protection Agency. Office of Water; 1999.
- Barbour MT, Stribling JB, Karr JR. Multimetric approach for establishing biocriteria and measuring biological condition. In: Davis WS, Simon TP, editors. *Biological Assessment and Criteria. Tools for Water Resource Planning and Decision Making*. Boca Raton, Florida: Lewis Publishers; 1995. p. 63–77.
- Ekhtera M, Mansoori G, Mensinger M, Rehmat A, Deville B. Supercritical fluid extraction for remediation of contaminated soil. In: Abraham M, Sunol A. editors. *Supercritical Fluids: Extraction and Pollution Prevention*. ACSSS, Vol. 670. Washington, DC: American Chemical Society; 1997. p. 280–298.
- Erickson RL, McKim JM. A model for exchange of organic chemicals at fish gills: flow and diffusion limitations. *Aquatic Toxicology* 1990;18:175–198.
- ESRI. *Understanding GIS: The Arc/Info Method*. 3rd ed. Redlands (CA): ESRI; 1995.
- Fausch DO, Karr JR, Yant PR. Regional application of an index of biotic integrity based on stream fish communities. *Transactions of the American Fisheries Society* 1984;113:39–55.
- Finkel A. *Confronting Uncertainty in Risk Management: A Guide for Decision-Makers*. Washington, DC: Center for Risk Management, Resources for the Future; 1990.
- Gerba CP, Pepper IL. Wastewater treatment and biosolids reuse. In: Maier RM, Pepper IL, Gerba CP, editors. *Environmental Microbiology*. 2nd ed. Burlington (MA): Elsevier Academic Press; 2009.
- Hall LW, Scott MC, Killen WD. *Development of biological indicators based on fish assemblages in Maryland coastal plain streams*. Annapolis, Maryland: Maryland Department of Natural Resources, Chesapeake Bay and Watershed Programs; CBWP-MANTA-EA-96-1. 1996.
- Hanlon JA. Office of Wastewater Management, U.S. Environmental Protection Agency. Memorandum to Water Division Directors, Regions 1–10. August 27, 2007.
- Hartman KJ, Brandt SB. Estimating energy density of fish. *Transactions of the American Fisheries Society*. 1995;124:347–355.
- Hughes RM, Gammon JR. Longitudinal changes in fish assemblages and water quality in the Willamette River, Oregon. *Transactions of the American Fisheries Society* 1987;116:196–209.
- Karr JR. Assessment of biotic integrity using fish communities. *Fisheries* 1981;6:21–27.
- Karr JR, Fausch KD, Angermeier PL, Yant PR, Schlosser IJ. *Assessment of Biological Integrity in Running Waters: A Method and Its Rationale*. Champaign, Illinois: Illinois Natural History Survey Special Publication 5; 1986.
- Kimbrough S, Vallero D, Shores R, Vette A, Black AK, Martinez V. Multi-criteria decision analysis for the selection of a near road ambient air monitoring site for the measurement of mobile source air toxics. *Transportation Research, Part D: Transport and Environment* 2008;13(8):505–515.
- Kushlan JA, Voorhees SA, Loftus WF, Frohring PC. Length, mass, and calorific relationships of Everglades animals. *Florida Scientist*. 1986;49:65–79.

- Leonard PM, Orth DJ. Application and testing of an index of biotic integrity in small, coolwater streams. *Transactions of the American Fisheries Society* 1986;115:401–414.
- Letcher T, Vallero D. *Waste: A Managers Handbook* Amsterdam (NV): Elsevier; 2011.
- Lyons J. *Using the Index of Biotic Integrity (IBI) to Measure Environmental Quality in Warmwater Streams of Wisconsin*. General Technical Report NC-149. St. Paul, Minnesota: North Central Forest Experiment Station, U.S. Department of Agriculture; 1992.
- Lyons J, Wang L, Simonson TD. Development and Validation of an Index of Biotic Integrity for Coldwater Streams in Wisconsin. *North American Journal of Fisheries Management* 1996;16:241–256.
- Malczewski J. *GIS and Multicriteria Decision Analysis*. John Wiley & Sons: New York, NY; 2009.
- Miller DL, Leonard PM, Hughes RM, Karr JR, Moyle PB, Schrader LH, Thompson BA, Daniel RA, Fausch KD, Fitzhugh GA, Gammon JR, Halliwell DB, Angermeier PL, Orth DJ. Regional applications of an Index of Biotic Integrity for use in water resource management. *Fisheries* 1988;13:12–20.
- Moyle PB, Brown LR, Herbold B. Final Report on Development and Preliminary Tests of Indices of Biotic Integrity for California. Final Project Report submitted to the U.S. Corvallis, Oregon: Environmental Protection Agency; 1986.
- Ohio Environmental Protection Agency (Ohio EPA). 1987. Biological Criteria for the Protection of Aquatic Life: Volumes I–III. Volume I. The role of biological data in water quality assessment. Volume II: Users manual for biological field assessment of Ohio surface waters. Volume III. Standardized biological field sampling and laboratory methods for assessing fish and macroinvertebrate communities. Ohio EPA, Division of Water Quality Monitoring and Assessment, Surface Water Section, Columbus, Ohio (updated 1988, 1989).
- Pacific Northwest National Laboratory and Mercury Analytical Laboratory. Determination of total mercury. Available at <http://marine.pnl.gov/resources/determination.stm>. Accessed September 23, 2009.
- Roth NE, Southerland MT, Chaillou JC, Vølstad JH, Weisberg SB, Wilson HT, Heimbuch DG, Seibel JC. *Maryland Biological Stream Survey: Ecological status of non-tidal streams in six basins sampled in 1995*. Annapolis, Maryland: Maryland Department of Natural Resources, Chesapeake Bay and Watershed Programs, Monitoring and Non-tidal Assessment; CBWP-MANTA-EA-97-2. 1997.
- Rubow KL. Filtration: fundamentals and applications In: *Aerosol and Particle Measurement Short Course*. Minneapolis (MN): University of Minnesota; 2004.
- Schreckenbach K, Knösche R, Ebert K. Nutrient and energy content of freshwater fishes. *Journal of Applied Ichthyology* 2001;17:142–144.
- Simon and Lyons “Application of the Index of Biotic Integrity to Evaluate Water Resource Integrity in Freshwater Ecosystems,” Chapter 16, in Davis and Simon. 1995. Biological Assessment and Criteria - Tools for Water Resource Planning and Decision Making.).
- Simon TP. Development of Ecoregion Expectations for the Index of Biotic Integrity. I. Central Corn Belt Plain. EPA 905-9-91-025. Chicago, Illinois: U.S. Environmental Protection Agency, Region 5; 1991.
- Solomon P, Norris G, Landis M, Tolocka M. Chemical analysis methods for atmospheric aerosol components. In: Baron PA, Willeke K, editors. *Aerosol Measurement: Principles, Techniques, and Applications*. 2nd ed. Hoboken (NJ): Wiley-Intersciences, Inc.; 2001.
- Steedman RJ. Modification and assessment of an index of biotic integrity to quantify stream quality in southern Ontario. *Canadian Journal of Fisheries and Aquatic Sciences* 1988;45:492–501.
- Sumathi VR, Natesan U, Sarkar C. GIS-based approach for optimized siting of municipal solid waste landfill. *Waste Management*. 2008;28(11): 2146–60.
- State of Georgia, Watershed Protection Plan Development Guidebook. 2003.

- Stewart DJ, Weininger D, Rottiers DV, Edsall TA. An energetics model for lake trout *Salvelinus namaycush*: application to the Lake Michigan population. *Canadian Journal of Fisheries and Aquatic Sciences* 1983;40:681–698.
- U.S. EPA. 1992. NPDES storm water sampling guidance document. EPA 833B92001. Chapter 3; 1992.
- U.S. EPA. 1994. Method 1613: Tetra-through Octa-Chlorinated Dioxins and Furans by Isotope Dilution HRGC/HRMS (Rev. B). Washington, DC.
- U.S. Environmental Protection Agency. 1997. *Volunteer Stream Monitoring Methods Manual*. EPA 841-B-97-003. Chapter 5. Monitoring and assessing water quality: 5.2.
- U.S. Environmental Protection Agency. 1998. Quality Assurance Guidance Document 2.12. Monitoring PM_{2.5} in Ambient Air Using Designated Reference or Class I Equivalent Methods. Research Triangle Park, North Carolina. November 1998.
- U.S. EPA. 1999. Method TO-9A in Compendium of Methods for the Determination of Toxic Organic Compounds in Ambient Air. 2nd ed. EPA/625/R-96/010b.
- U.S. Environmental Protection Agency. 2001. D. Crumbling, 2001. Clarifying DQO Terminology Usage to Support Modernization of Site Cleanup Practice. EPA 542-R-01-014.
- U.S. Environmental Protection Agency. 2002. Guidance for the Data Quality Objectives Process. EPA QA/G-4, EPA/600/R-96/055; Washington, DC.
- U.S. Environmental Protection Agency and Battelle National Laboratory. 2004. ETV Joint Verification Statement. Rapid Polymerase Chain Reaction: Detecting Biological Agents and Pathogens in Water; Available at <http://www.epa.gov/ordnhsrc/pubs/vsInvitrogen121404.pdf>. Accessed September 24, 2009.
- U.S. Environmental Protection Agency. 2006. *Data Quality Objectives Guidance*. EPA/240/B-06/001. Washington, DC.
- U.S. Environmental Protection Agency. 2011a. *Hazardous waste: test methods*. Available at <http://www.epa.gov/osw/hazard/testmethods/index.htm>. Accessed March 8, 2011.
- U.S. Environmental Protection Agency. 2011b. Hazardous waste: Test methods: Composite sampling. Available at http://www.epa.gov/osw/hazard/testmethods/faq/faqs_sampl.htm#Composite. Accessed March 8, 2011.
- U.S. Environmental Protection Agency. 2011c. *RCRA SW846 Method 8290: Polychlorinated Dibenzodioxins (PCDDs) and Polychlorinated Dibenzofurans (PCDFs) by High Resolution Gas Chromatograph/High Resolution Mass Spectrometry (HRGC/HRMS)*; Available at <http://www.epa.gov/osw/hazard/testmethods/sw846/>. Accessed March 7, 2011.
- United States Code of Federal Regulations: 40 CFR 131.6 Appendix B—Definition and procedure for determination of the method detection limit—Revision 1.11. [49 FR 43430, Oct. 26, 1984; 50 FR 694, 696, Jan. 4, 1985, as amended at 51 FR 23703, June 30, 1986.]
- Vallero D. *Engineering the Risks of Hazardous Wastes*. Boston (MA): Butterworth-Heinemann; 2003.
- Vallero DA, Reckhow KH, Gronewold AD. Application of multimedia models for human and ecological exposure analysis. International Conference on Environmental Epidemiology and Exposure. October 17. Durham (NC): Graphic adapted from U.S. Environmental Protection Agency; 2007.
- Vallero DA. *Fundamentals of Air Pollution*. 4th ed. Burlington (MA): Elsevier Academic Press; 2008.
- Vallero DA. *Environmental Biotechnology: A Biosystems Approach*. Amsterdam (NV): Elsevier Academic Press; 2010.
- Vernier Corporation. *Computer 19: Dissolved oxygen in water*. Available at http://www2.vernier.com/sample_labs/BWV-19-COMP-dissolved_oxygen.pdf. Accessed October 19, 2009.

6

HYDROLOGY MEASUREMENTS

TODD C. RASMUSSEN

- 6.1 Introduction
 - 6.1.1 Stage
 - 6.1.2 Velocity
 - 6.1.3 Water budget
 - 6.1.4 Residence time
- 6.2 Precipitation
 - 6.2.1 Rain gauges
 - 6.2.2 Snow accumulation
 - 6.2.3 Remote sensing
- 6.3 Evapotranspiration
 - 6.3.1 Pan evaporation
 - 6.3.2 Watershed studies
 - 6.3.3 Weighing lysimeters
 - 6.3.4 Enclosures
 - 6.3.5 Eddy covariance
 - 6.3.6 Indirect approaches
 - 6.3.7 Remote sensing
- 6.4 Surface flow
 - 6.4.1 Topographic and bathymetric maps
 - 6.4.2 Control structures
 - 6.4.3 Rating curves
- 6.5 Groundwater
 - 6.5.1 Types of aquifers
 - 6.5.2 Hydraulic properties
 - 6.5.3 Flow and transport
- 6.6 Soil water
 - 6.6.1 Infiltration
 - 6.6.2 Percolation
 - 6.6.3 Hydraulic properties
- 6.7 Water quality
 - 6.7.1 Physical
 - 6.7.2 Chemical
 - 6.7.3 Biological

Suggested Readings

Handbook of Measurement in Science and Engineering. Edited by Myer Kutz.
Copyright © 2013 John Wiley & Sons, Inc.

6.1 INTRODUCTION

Hydrology is the study of the occurrence, movement, and quality of water in the environment, including both surface and subsurface systems. Surface water hydrology focuses on aboveground flow and transport, including overland, channel, lake, and wetland environments, whereas subsurface hydrology considers both groundwater (saturated) and soil-water (unsaturated) environments.

Sciences closely related to hydrology include geology, atmospheric sciences, glaciology, ecology, and oceanography. Engineering, law, economics, political science, and management sciences also rely on hydrologic knowledge because of water's important role in human systems, such as thermoelectric and hydroelectric power production; navigation; and municipal, industrial, and agricultural water uses.

This section summarizes hydrologic measurements in natural systems. The goal of gathering these measurements is to develop a quantitative understanding of the hydrologic system being investigated. This understanding—with relevant data—can then be used for design, prediction, and management purposes.

6.1.1 Stage

Stage is used to describe the water-surface elevation relative to an arbitrary datum. While stage could be referenced to mean absolute sea level, reference to any local datum can be used.¹ In saturated, subsurface environments, the *hydraulic head* is used instead of stage to describe the height of water in a well. The head can also be inferred for water held in saturated pores near the well by assuming equilibrium between the well and the formation.²

Manual stage measurements include the use of a *staff gauge*, which is a precision measurement scale mounted vertically in the water body so that the water stage can be visually determined from the gauge. The staff gauge is permanently placed in a location where it can be observed over a wide range of stages, either directly or through a remote video feed. If a single staff gauge is not visible or submerged at high flows (or dry during low flows), then additional staff gauges can be placed uphill or downhill of each other.

A *sounder* or *depth-indicator* is used to obtain manual water-level measurements in a well or below a bridge. The sounder probe is lowered until it reaches the water surface, at which point either a sound or a light is triggered. The water-surface elevation is determined by subtracting the depth to water from the measurement elevation.

A *hook-gage* is commonly used for more precise water-level measurements (e.g., in an evaporation pan). These gauges can measure water levels to a precision of 20 μm (<0.001 in.), but require a stilling well that has been accurately leveled.

Automated stage (depth) sensors, include vibrating wires, strain-type pressure transducers, bubblers, and acoustic level indicators, which can be used to continuously monitor water levels. These devices are mounted either below the water surface (e.g., vibrating wires, transducers, bubbler hoses), or above the water surface looking downward to measure fluid pressure (e.g., acoustic sensors). Sensors that monitor fluid

¹ If a geo-referenced elevation is required, then the elevation of the local datum can be found and stages can be converted to an absolute elevation.

² In unsaturated, subsurface environments, a negative pressure (or *soil water tension*) is used instead of hydraulic head, as discussed in a subsequent section.

pressure may not provide accurate readings of water levels if a significant water velocity is present or if the fluid density changes due to temperature, entrained gasses, sediments, and dissolved solids.

6.1.2 Velocity

Water velocities, v , are determined using various types of flow meters, such as a *Doppler*, *electromagnetic*, or *Gurly-type*. *Acoustic Doppler Velocity* (ADV) meters generate a sound source that produces an echo from particulates in the moving water. The frequency (Doppler) shift of the echo relative to the source is used to find the two- or three-dimensional velocity field in the vicinity of the flow meter. An electromagnetic flow meter generates an electromagnetic field that is distorted by a moving electrolyte. The Gurly-type flow meter employs rotating vanes that spin as the fluid moves past the device.

Tracers can also be used to measure water velocity. *Natural*, or *environmental*, tracers are ones that are already present within a system. For example, the presence of synthetic compounds that have been generated recently (e.g., chlorofluorocarbons, radionuclides from atomic testing) can be used to distinguish between older and newer waters in the subsurface. Also, radiocarbon (^{14}C) dating can be used to determine the time since groundwater was recharged.

Introduced tracers are added intentionally for the purpose of determining travel times and velocities. In this case, care must be taken that the tracer does not cause environmental harm. Ideally, the tracer should be conservative, for example, it does not decay or sorb along the way. Examples include dissolved noble gasses (e.g., helium, argon, krypton) and fluorobenzoic acids. Tracers can be added as a single spike or at a constant rate over a longer period of time. They can also be added at a point, along a line, or over an area.

In surface water, the total flow, Q , is found by multiplying the flow velocity, v , by the cross-sectional area perpendicular to flow, A ($Q = vA$), or if the total flow and area are known, then $v = Q/A$. Alternatively, the velocity can be found using the time of travel, τ , of a tracer along a travel path of length L ($v = L/\tau$), which leads to $Q = V/\tau$, where $V = AL$ is the total volume of water moving along the travel path.

Within the subsurface, the *Darcian velocity* or *flux*, q , is used to characterize the volumetric flow rate, that is, $q = Q/A$, where Q is the total rate and A is the cross-sectional area normal to flow. The fluid velocity can be determined by dividing the flux by the porosity, n ($v = q/n$). Note that the flux only equals the velocity when a unit porosity is present. The flux can be difficult to measure directly, so *Darcy's Law* ($\mathbf{q} = -K\nabla h$) is commonly used to infer the flux using the hydraulic conductivity, K , and the hydraulic gradient, ∇h , where h is the hydraulic head.

A third type of velocity is the *celerity*, *kinematic*, or *wave velocity*, c , which is used to describe dynamic changes in stage or head. The celerity can be found either by observing the time of arrival, τ , of flood peaks across a distance, L ($c = L/\tau$), or estimated using $c = dQ/dA$, which is the marginal change in discharge, dQ , per unit change in cross-sectional area, dA .

6.1.3 Water Budget

The *water budget* is an accounting of the stores and flows of water in a hydrologic system, and provides an overview of the important elements of the hydrologic system

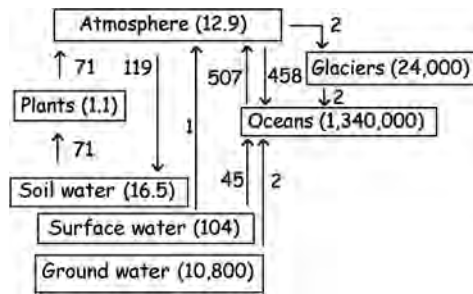


FIGURE 6.1 Annual global hydrologic budget. Volumes in PL. Flows in PL/year (1 PL = 10¹⁵L).

being studied. The budget is used to account for quantities of water in storage, as well as the movement of water between storage elements. As such, it can become the organizing tool for guiding hydrologic investigations.

For example, Figure 6.1 shows a global water budget that quantifies the volume of water in various global storages. Note that the oceans dominate the total inventory, followed by glaciers, and then groundwater. Also note the rate of flow between the storage elements. The water budget for a specific study area will likely be unique, and understanding the volumes and exchanges between elements helps to understand the hydrologic behavior of the study area.

The water budget can also be used to show changes in storage over time within storage elements. For dynamic systems, changes in storage are found using $\Delta S/\Delta t = I - O$, where $\Delta S/\Delta t$ is the change in storage per unit time, and I and O are the inflows and outflows to a storage element, respectively. These quantities can be expressed either as a volume, V , or as a depth, D ($D = V/A$), where A is the catchment area.

While simple in concept, accounting for the spatial and temporal complexity of water storage and movement in natural systems requires that careful attention be paid to monitoring design. Figure 6.2 illustrates a simplified conceptualization of hydrologic monitoring components, which are more fully discussed in subsequent sections.

Note that the conceptual model can be organized in different ways. For example, atmospheric exchanges of water can be divided into precipitation inputs and evapotranspiration outputs. Interactions within the subsurface can also be lumped or divided into various

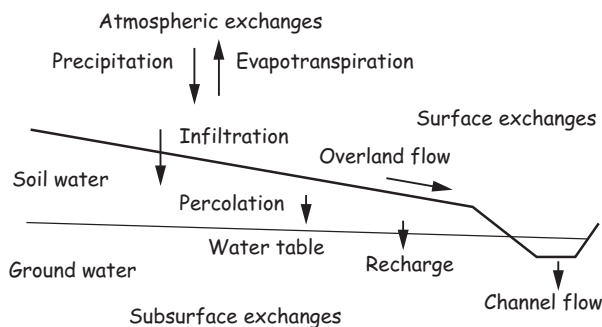


FIGURE 6.2 Simplified conceptual model of hydrologic system showing water storage elements and exchanges between elements.

regions (e.g., soil water, shallow groundwater). And the surface water can be divided into overland, channel, and water bodies (e.g., wetlands, lakes, reservoirs).

6.1.4 Residence Time

Another useful tool for organizing data collection efforts is the definition of the water *residence time*, which is the average length of time water spends in each reservoir. Many chemical and biological conditions are influenced by the residence time, and many of the measurements described in the following sections are used to quantify this attribute of the hydrologic system.

The concept of the residence time is relatively simple; a small reservoir with a large flow through it has a shorter residence time than a large reservoir with small flows. The mean residence time, τ , is calculated by dividing the storage volume, V , by the average flow rate, Q ($\tau = V/Q$). Another means for calculating the residence time in a flowing system (such as a river) is to use $\tau = L/v$, where L is the distance between two points and v is the fluid velocity.

Although a system with stable flow rates and volumes has a constant residence time, this would not be true if flows change rapidly with time. In this case, the average residence time can be dynamically updated using $\tau(t) = \Delta t + \alpha \tau(t - \Delta t)$, where $\tau(t - \Delta t)$ was the mean residence time during the previous time, Δt is the time interval, $\alpha = 1 - \Delta V/V$ is the fraction of “old” water carried over from one time step to the next, and where $\Delta V = Q(t) \Delta t$ is the volume of “new” water entering the system.

6.2 PRECIPITATION

Atmospheric precipitation is the primary input to hydrologic systems, and can occur as rain, snow, sleet, or hail. Precipitation often varies rapidly in both space and time and is a challenge to measure accurately over large areas, especially in mountainous regions where orographic conditions may influence accumulation. Precipitation is commonly measured using ground- and satellite-based systems.

6.2.1 Rain Gauges

A rain gauge is a standard method for collecting atmospheric precipitation. Rain gauges should be placed away from taller objects (e.g., trees, buildings) that might obstruct precipitation collection. Gauges should also be placed sufficiently high so that *ground splash* does not affect the reading. Locating the gauge near shrubbery, which attenuates wind near the gauge but does not obstruct the rain, should minimize wind effects.

Both manual (nonrecording) and automated (recording) rain gauges are available, the selection of which depends on the temporal frequency of required data. A *manual rain gauge* normally consists of a collection funnel that concentrates collected rainfall into a smaller, inner collection tube (Figure 6.3). The purpose of funneling water into the inner tube is to enhance the measurement resolution of the device. When the capacity of the inner tube is exceeded, the excess rainfall overflows into a larger diameter, outer collection tube. The depth of rainfall is determined by measuring the depth of rainfall collected within the inner tube, as well as the excess in the outer tube, if present. The excess water



FIGURE 6.3 Manual rain gauge showing collection funnel, inner collection tube (1 in. capacity), outer collection tube (10 in. capacity), and mounting post. Credit: www.agry.purdue.edu/turf/tips/2006/raingauge.jpg

in the outer tube is measured by pouring it into the inner tube at the time of measurement. Manual rain gages are commonly read on a daily basis.

Automated rain gauges typically use a funnel to route precipitation to a pair of tipping buckets that alternately collect increments of precipitation (Figure 6.4). An electronic signal is generated for each bucket tip (generally every 0.2 mm or 0.01 in.). Depending on the recording device used, the recorded data could consist of either the total number of bucket tips over a specified time increment (e.g., every 5 min) or the time of each tip.

6.2.2 Snow Accumulation

Frozen precipitation (e.g., snow, sleet, freezing rain) is more difficult to monitor using rain gauges, both because of wind blowing snow past the gauge as well as the accumulation of frozen precipitation within the funnel. While a *wind shield* can be placed around the gauge to minimize blowing snow, and a *heating element* can be used to melt frozen precipitation, these methods are still prone to error.



FIGURE 6.4 Automated rain gauge showing tipping bucket used to measure precipitation increments.

Alternatively, the amount of snow on an exposed horizontal surface can be recorded. The *snow water equivalent*, D_{sw} , can be directly measured using *snow pillows* that measure the weight of snow that accumulates over time. A snow pillow consists of an inflated chamber, the pressure within being directly related to the weight of the overlying snow.

For cases where only the snow depth, D_s , is known, a core sample should be collected to determine the *snow density*, ρ_s . The density is determined by weighing the mass of

snow collected within a core barrel of known volume. The snow water equivalent is then found using $D_{\text{sw}} = \rho_s D_s$. While the snow water equivalent is useful for predicting the potential runoff volume, it does not directly provide a measure of the likelihood for snow-melt. The *snow cold content*, E_s , is the energy required to raise the snow temperature to 0°C , $E_s = H_T T_s D_{\text{sw}}$, where H_T is the specific heat of ice ($0.5 \text{ cal/g}^\circ\text{C}$), and T_s is the snow temperature below freezing. The snow temperature can be measured manually or using automated techniques.

6.2.3 Remote Sensing

Remotely sensed precipitation estimates are also available, including ground and satellite sources. Remote sensing uses active (e.g., radar) or passive microwave signals to determine the atmospheric density of precipitation. The Tropical Rainfall Monitoring Mission³ is a satellite-based platform that provides both active and passive estimates of precipitation between the latitudes of $\pm 50^\circ$. Ground-based radar sources are also available in many countries (e.g., NEXRAD). The advantage of these estimates is that they can provide improved spatial and temporal resolution over large areas, but may be less accurate in mountainous terrain.

6.3 EVAPOTRANSPIRATION

Evapotranspiration describes the process of water movement from the surface to the atmosphere. *Evaporation* is an abiotic process that accounts for water loss from the water surfaces (e.g., lakes, ponds) and from unvegetated soil or paved surfaces. *Transpiration* consists of water loss from plant surfaces, either from intercepted precipitation on plant surfaces or from the translocation of subsurface water through the roots, stems, and leaves of the vegetation.

Measuring and estimating evapotranspiration is difficult, and indirect methods are commonly used. That is, variables other than evapotranspiration are measured and used to make inferences about evapotranspiration. While direct estimates of evapotranspiration are preferred for different kinds of crops, forests, and soils, this is often expensive and time-consuming.

When characterizing evapotranspiration, it is important to distinguish between the *potential evapotranspiration* (PET) and the *actual evapotranspiration* (AET). PET is the maximum possible transpiration by plants with unlimited soil moisture. For a given plant community, PET is determined by climatic variables such as temperature, solar radiation, humidity, and wind speed. Because water is often limiting, AET is a better estimate of water loss to the atmosphere. AET depends on PET, as well as soil moisture conditions and the type of crop (or vegetation) and its stage of growth. Estimating AET on a daily basis is useful for determining crop irrigation needs.

6.3.1 Pan Evaporation

Pan evaporation is considered to be the simplest way to estimate PET. The *Class A evaporation pan* has a cylindrical shape with a diameter of 120.7 cm and is filled with water to a

³ TRMM, <http://trmm.gsfc.nasa.gov/>



FIGURE 6.5 Class A evaporation pan showing animal exclusion cage, anemometer for wind speed determination, and water level measurement device. Credit: http://upload.wikimedia.org/wikipedia/commons/e/e5/Evaporation_Pan.jpg

depth of 25 cm (Figure 6.5). The pan is placed on a level (wooden) base and should be enclosed to prevent water loss from animals such as livestock, birds, and bats. Control of debris, mosquitos, and algae within the pan is also important. Water loss due to evaporation, E_p , is expressed as a change in water depth each day, $E_p = dh/dt$. Measurements are not possible when the water surface is frozen or when large precipitation events overtop the pan.

An empirical *pan coefficient* is commonly employed to determine lake evaporation as well as AET. In these applications, the lake evaporation or AET is found as the product of the pan coefficient, C_p , and the pan evaporation rate, $AET = C_p E_p$.

6.3.2 Watershed Studies

A common procedure for determining AET over large areas uses a *water balance equation*, $AET = P - Q - \Delta S/\Delta t$, where P is the measured average precipitation over the

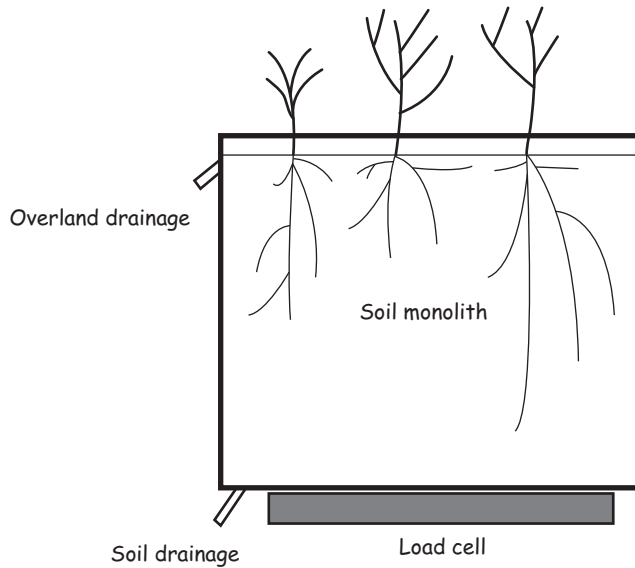


FIGURE 6.6 Weighing lysimeter showing the plant-containing soil monolith, load cell, water drainage through monolith, and overland flow from monolith.

watershed, Q is the measured total discharge (surface and groundwater) from the watershed, and $\Delta S/\Delta t$ is the change in water storage within the watershed. Note that measurement errors directly affect the AET estimate, so that precision is essential in the determination of the water balance components.

6.3.3 Weighing Lysimeters

Direct measurements of AET for specific crops over time can be obtained using *weighing lysimeters*, which are large soil monoliths that are placed on top of load cells (Figure 6.6). Weighing lysimeters are expensive but provide accurate, time-varying estimates of evapotranspiration. Daily changes in load are used to account for water loss from the soil monolith during the life cycle of the crop. AET is determined using the water balance equation, except that P is the sum of precipitation and irrigation inputs, Q is the sum of overland flow and deep seepage out of the lysimeter, and $\Delta S/\Delta t$ is found using the change in load (mass) of the lysimeter.

A novel approach for determining landscape-scale evapotranspiration uses groundwater levels in monitoring wells to record changes in soil moisture content. This approach, termed a *geological weighing lysimeter*, relies on fluid pressures within isolated geologic units in place of the traditional load cell. This technique has been demonstrated in a large number of regions around the world, but requires a suitable geologic unit.

6.3.4 Enclosures

Another method for determining evapotranspiration uses gas flux measurements from enclosed crops. A transparent tent is placed over the target vegetation, and air is blown



FIGURE 6.7 Eddy covariance station used to measure turbulent gas (e.g., water vapor) movement within the atmospheric boundary layer. Credit: www.campbellsci.com/news-2010-q3/ec150-ec155

through the tent. Evapotranspiration is determined by measuring the humidity and flow rate of the incoming and outgoing air. Care must be taken, however, to ensure that temperatures and gas concentrations within the enclosure are consistent with outdoor conditions.

6.3.5 Eddy Covariance

The eddy covariance technique measures vertical turbulent fluxes within the atmospheric boundary layer to find the water loss to the atmosphere (Figure 6.7). An added benefit is the ability to determine other gaseous exchanges, such as CO_2 and CH_4 . The method requires a large, horizontal, homogenous expanse of vegetation to ensure uniform conditions over the area.

6.3.6 Indirect Approaches

The *Penman–Monteith* equation estimates evapotranspiration using daily mean temperature, relative humidity, wind speed, and solar radiation. The *Blaney–Criddle* method is an alternative method that only requires air temperature at a site and is more commonly used in the western United States, where relative humidity is frequently low. For both methods, soil moisture availability and crop stage are additional factors that must be considered when trying to estimate AET.

6.3.7 Remote Sensing

Remote sensing of evapotranspiration losses is also possible using an *energy budget*. In this method, the incoming short-wave energy from the sun is offset by short-wave reflection, long-wave emission, surface conduction, and latent heat losses from evapotranspiration. Both short-wave and long-wave components can be monitored remotely, and surface conduction can be estimated, leaving the evapotranspiration loss as the residual: $\text{ET} = (\text{SW} + \text{LW} + G)/\delta$, where ET is the estimated evapotranspiration rate, SW is the net incoming short-wave radiation, LW is the net incoming long-wave radiation, G is the net transfer of sensible heat from the subsurface to the ground surface, and δ is the latent heat of evaporation.

6.4 SURFACE FLOW

Surface flow consists of *overland flow* (e.g., flow across the ground surface that is not contained in a channel, including *sheet*, *rill*, and *gully* flow, Figure 6.8), *lotic* flow (e.g., fast moving water in defined channels), and *lentic* flow (e.g., slow flow in wetlands, lakes, and reservoirs).

Quantifying overland flow can be challenging due to its disseminated and ephemeral nature. Berms can be used to intercept overland flow, concentrating and routing it to a control structure (described below). Water within a defined channel is often easier to measure because it is more concentrated, although shallow, braided channels can be challenging. Measuring flows in lentic systems can also be challenging because internal circulation is slow and varies with depth and location.

6.4.1 Topographic and Bathymetric Maps

Topographic maps provide detailed geomorphic information (e.g., slope, aspect) about elevations within the system. Features such as stream channels, wetlands, and lakes can be determined, as well as artificial structures such as reservoirs, highways, and developed areas. These maps provide a starting point for determining the size and nature of the watershed contributing to surface water flow.

While published maps and Digital Elevation Models are available for most areas, more detailed maps may be required for some purposes. In these cases, *photogrammetric* and *LIDAR* tools can be used to develop high-resolution topographic maps.

Bathymetric maps display the topography of the bed of large water bodies, such as a wetland, lake, or reservoir. In these systems, lines of equal elevation are used to show the bottom contours. The bathymetric map can be used to determine the relationship between

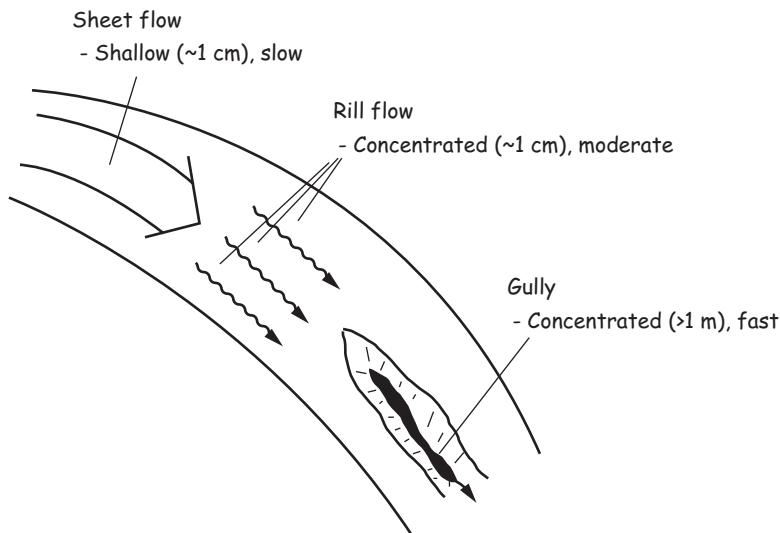


FIGURE 6.8 Overland flow showing sheet (or saturation-excess) flow, as well as concentrated flow in rills and gullies.

the elevation, area, and volume of water of the system. Estimation of the area within each contour interval, A , provides a starting point for estimating the water storage volume, V , by integration: $V = \int A(z) dz \approx \sum A_i \Delta z$, where z is the elevation. The summation provides a useful approximation for small changes in elevation, Δz .

Depth profilers using *echo-sounders* (sonar) with concomitant *Global Positioning System* locations can be used to map the shape of the feature. For wadeable systems, survey transects can be run across the feature to generate a profile which can be interpolated to create the bathymetric map. The bathymetry of an artificial reservoir can also be taken from topographic surveys created before the feature was filled. Changes in bathymetry over time may occur due to sediment accumulation or dredging.

6.4.2 Control Structures

Engineered structures often provide the most accurate method for determining flow rates. This is because a simple geometry and known flow hydraulics can be imposed that reduces uncertainties in the flow measurement. One simple control structure would be an empty container (e.g., barrel, reservoir) that is filled over time. The flow rate into the container, Q , would be $Q = A \Delta h / \Delta t$, where A is the cross-sectional area of the container, and $\Delta h / \Delta t$ is the rate of change of water level over time in the container.

Another type of control structure is the *weir* that consists of an upstream ponding basin and a weir blade (Figure 6.9). Discharge from the weir is directly related to the stage in the weir basin. The weir *bulkhead* (or dam) is constructed using a concrete, timber, or metal structure. A weir *notch* is placed in the bulkhead to allow water to pass. The weir *crest* is the edge or surface over which the discharged water flows. Types of outlets include triangular (or v-notch), rectangular, and trapezoidal (or cipolletti). A *nappe* forms as water spills over the crest. The nappe can be classified as either *submerged* or *free*. A submerged nappe forms when the crest is flooded, whereas a free nappe forms when the water discharges freely into air.

A *sharp-crested* weir is constructed so that the flowing water passes over a vertical, knife-edge blade. This design minimizes friction across the weir blade and provides the greatest accuracy. A *broad-crested* weir consists of a crest across which water flows for

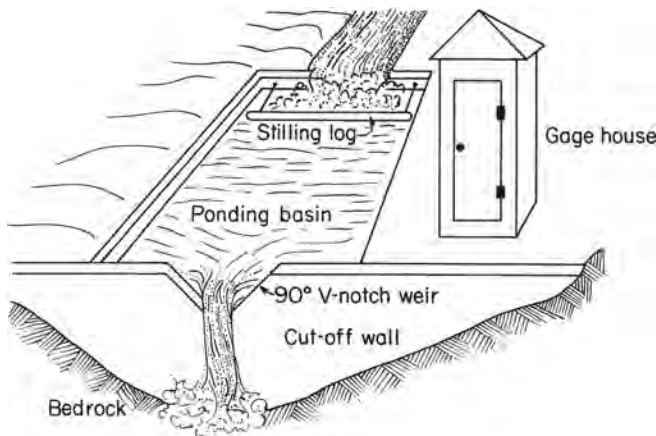


FIGURE 6.9 General components of a weir.

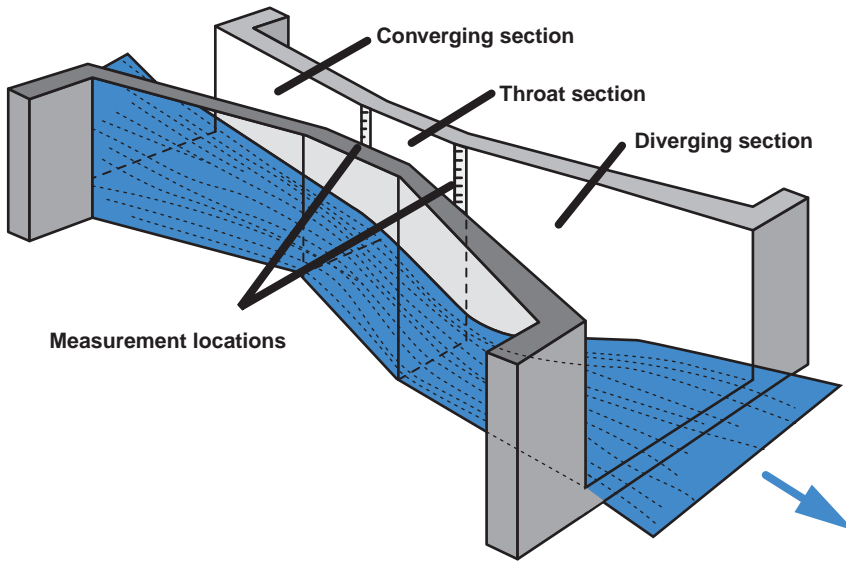


FIGURE 6.10 General components of a Parshall flume.

some distance before falling over the downstream edge. Spillways of dams are often constructed in the shape of a broad-crested weir.

Weirs may not provide accurate estimates when the weir blade becomes blocked by ice or floating debris, such as leaves and branches. Another source of error arises when the weir basin fills with sediments.

Flumes are another type of device for estimating discharge. Flumes require no upstream stilling pond and allow sediment to pass unimpaired through the structure. Ice, leaves, and other debris may still affect readings if they block the flume outlet. *H-type* flumes developed by the U.S. Department of Agriculture are useful for measuring discharge in sediment-laden streams. A small drop is required downstream of H-type flumes, which can be difficult to achieve in level channels. Another popular flume design is the *Parshall* flume, which uses a constriction in the width and depth of the channel (Figure 6.10). No drop is needed downstream of the structure, allowing its use in level channels.

Rather than having to construct a control structure, a *culvert* may already exist to route water under roads. Culverts are often round or rectangular, providing a regular section for measuring discharge. There are four combinations of two general conditions, upstream flooded (or open) and downstream flooded (or open).

The simplest flow condition occurs when both the upstream and downstream conditions are flooded. These are fully submerged conditions that are consistent with flow through a pipe. In this case, the velocity is estimated using the difference in elevation between the upstream and downstream water levels, along with the culvert length, roughness, and diameter. When both the upstream and downstream conditions are open, then channel flow conditions exist. In this case, the discharge can be determined using Manning's equation (described below). Necessary information includes the culvert roughness, hydraulic radius, and slope. The accuracy of this approach is degraded by debris or other obstructions within the culvert.

6.4.3 Rating Curves

Rating curves—also called a *stage–discharge relationship*—are used to relate river *stage* to *discharge* so that measurements of water levels can be used to estimate stream discharge. Creating the rating curve requires that periodic measurements of stream discharge be obtained, and related to the river stage at the time of measurement. Sufficient flow measurements over a range of stages (low, medium, and high) should be collected so that the full range of anticipated discharges can be predicted. Also, repetitive measurements should be collected to ensure that channel flow conditions do not change over time due to scouring or sedimentation within the channel.

Stream discharge can be determined using field measurements of water velocity within the channel. The water velocity is usually not constant in a channel, with the maximum velocity commonly occurring at or near the surface in the middle of the channel, with slower velocities near the bottom and along the banks of the channel. The actual distribution of channel velocities can vary greatly, however, due to bends in the river or obstructions such as woody debris or rocks.

Because water velocity can be highly variable within a channel, multiple measurements of water velocity are needed at different points within the channel. The stream discharge, Q , is calculated using the sum of discharges within specific sections within the channel, $Q = \sum_{i=1}^n Q_i$, where $Q_i = v_i A_i$ is the flow in each section, v_i is the water velocity in each section, and A_i is the channel section cross-sectional area. To reduce measurement error, an ideal channel section should be straight, with few obstructions, and with uniform depth and velocity. A minimum of 10 subsections is then used to estimate the total flow, with approximately equal flows in each subsection, if possible. That is, faster, deeper parts of the channel should have more measurements than shallow, calm sections.

Rather than collecting field measurements of water velocities, *Manning's equation* provides an empirical relationship to estimate the mean channel water velocity for conditions where engineered structures and flow measurements are not available, $v = (\alpha/n)R^{2/3}S^{1/2}$, where v is the average channel velocity, $\alpha = 1$ in metric units and 1.49 in English units, n is the *Manning roughness coefficient*, which increases with increasing channel roughness, $R = A/P$ is the *hydraulic radius*, equal to the ratio of the channel cross-sectional area, A , to the channel *wetted perimeter*, P , and S is the *energy gradient*. The total stream discharge is found using $Q = vA$. The energy gradient can be approximated using the regional channel slope (taken from topographic maps). The hydraulic radius can be approximated for a wide, shallow stream using the channel depth.

6.5 GROUNDWATER

Groundwater traditionally refers to subsurface water located in saturated geologic materials. The upper boundary of the saturated zone lies at or near the *water table*, which is the elevation of water in a well tapping the uppermost part of the saturated zone.

The saturated zone may extend above the water table due to *capillary rise* in small pores. The capillary rise is caused by *surface tension forces*, consisting of both *cohesive forces* between water molecules and *adhesive forces* between the water and mineral surfaces. The zone of saturation above the water table is called the *capillary fringe*.

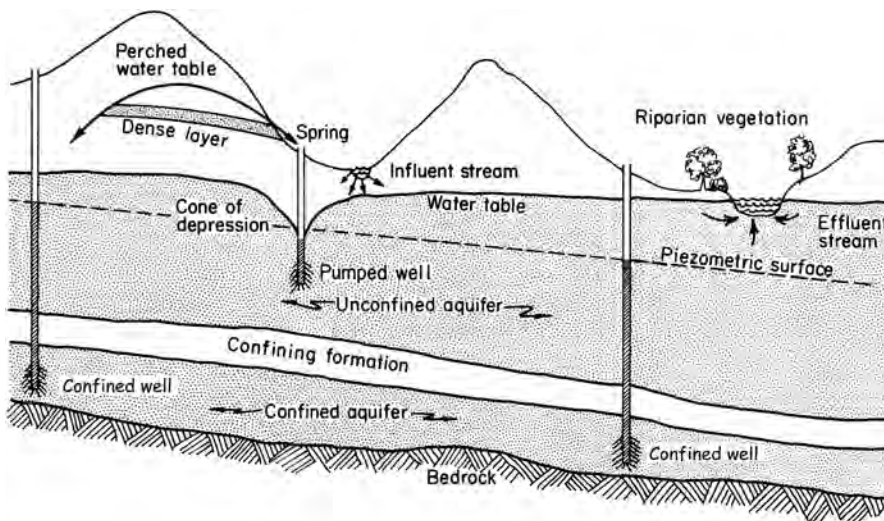


FIGURE 6.11 Subsurface hydrology showing unconfined (surficial) and confined aquifers.

6.5.1 Types of Aquifers

Aquifers are defined as geologic units that transmit useful quantities of water to wells. An *unconfined aquifer* refers to an aquifer bounded above by the water table and is the uppermost aquifer if more than one aquifer is present. Contour lines of water table elevations can be drawn just like topographic contours.

In some geologic regions, multiple aquifers may be present, with layers of lower permeability separating each of the aquifers (Figure 6.11). The lower aquifers are referred to as *confined aquifers*. A *confining layer* (or *aquitard*) of lower permeability separates confined aquifers from overlying aquifers.

Water levels in confined aquifers do not correspond to the water table; instead, they are termed *piezometric surfaces*. This is because water levels in these wells rise above the top of the aquifer. *Flowing* or *artesian* wells occur when the piezometric surface is higher than the ground surface. Contour lines of equal water-level elevations, called *equipotentials*, can be drawn for each aquifer. Figure 6.12 illustrates how *piezometers* (small-diameter wells

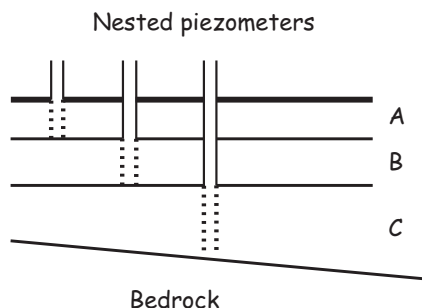


FIGURE 6.12 Piezometers can be installed in different soil horizons to monitor the vertical change in hydraulic head at a site.

used for monitoring water levels) can be installed at different depths for the purpose of monitoring the vertical distribution of hydraulic head within a profile.

Drilling logs and *borehole surveys* can be used to determine the locations of aquifers and confining layers. In fractured rock, *drill cores* or *downhole telev viewers* can be used to identify water-yielding intervals. Water quality measurements can also be used to identify different hydrogeologic units.

6.5.2 Hydraulic Properties

Pores, or *voids*, within the saturated zone are entirely filled with water. The volume of water per unit volume of solids is referred to as the *porosity*, n . The porosity of geologic materials can be found by extracting samples of the media, making sure that they are fully saturated, and then measuring the water lost by drying in an oven. The volume of water lost per unit volume of sample is their porosity.

The ability of geologic materials to conduct water is called the *hydraulic conductivity*, K . A similar term is the *permeability*, k , which is a more general term that can be used for determining the movement of any fluid. The relationship between the two is found using $K = (\gamma/\mu)k$, where γ is the *specific weight* of the fluid used to measure the fluid pressure, and μ is the *dynamic viscosity* of the moving fluid. For a system in which freshwater is used to measure both the oil and water pressure, then the ratio of conductivities is equal to the ratio of their viscosities; the permeability is approximately the same for both.

The source of water pumped from an unconfined aquifer is water draining from pores as the water table falls. In this case, not all of the water drains because of water held by capillary forces within the capillary fringe. The volume of water drained per unit area of aquifer per unit decline of water table is called the *specific yield*, S_y , of the aquifer and is usually less than the porosity, $S_y < n$.

Water does not drain from a confined aquifer as the piezometric surface is lowered because the air–water interface is higher than the top of the aquifer. In this case, the water source is the slight expansion of water due to the small compressibility of water and the aquifer itself. The *specific storage*, S_s , is used to describe the water released from a confined aquifer.

Two additional hydraulic parameters are used for confined aquifers, the *transmissivity*, T , and *storativity*, S . The transmissivity is the product of the *aquifer thickness*, b , and the hydraulic conductivity of the aquifer, $T = bK$. Similarly, the storativity is the product of the thickness and the specific storage, $S = bS_s$. Like the specific yield, the storativity represents the volume of water drained per unit area of aquifer per unit decline in water level, but is much smaller, $S_y \gg S$.

Vertical flow between aquifers through the confining layer is termed *leakage*. The *hydraulic conductance*, $C = K'/b'$, is used to account for this leakage and is a function of the hydraulic conductivity of the confining layer, K' , and the thickness of the layer, b' .

Slug tests are commonly used to determine the hydraulic conductivity by rapidly adding or removing a known volume of water (or slug) and noting the resulting change in water level over time. The advantage of this test is that only a single borehole is required, yet the hydraulic properties are generally biased toward those in the vicinity of the borehole and may not accurately represent regional behavior.

Aquifer hydraulic properties are best estimated using *aquifer tests*. A well is pumped at a constant rate and nearby *observation wells* are monitored for their *drawdown*, which is the decline in water levels as a cone of depression forms around the pumping well.

Required data include the pumping rate, Q ; the distance from the pumping well, r ; and the drawdown record as a function of time, $s(t)$. These data are interpreted using graphical or computer techniques to determine the aquifer hydraulic properties. If drawdown data are collected with sufficient frequency, then the time rate of change of drawdowns can be determined, ds/dt , and analyzed using *derivative curve techniques*.

Alternatively, *periodic* (or *sinusoidal*) aquifer tests can also be used to determine aquifer hydraulic properties. In this approach, a periodic pumping rate is used in conjunction with the observed responses in observation wells to determine the lag and attenuation of the pumping signal. These values are then used to infer hydraulic properties. The primary advantages of the periodic approach are the minimization of formation water removed (which may be contaminated and require expensive treatment), as well as the minimization of background interference from regional disturbances.

6.5.3 Flow and Transport

The *hydraulic head*, h , is a variable used to determine the magnitude and direction of water flow in the subsurface; water moves from high to low head. For example, water moves from zones with high water tables in an unconfined aquifer toward lower areas. If a stream is lower in elevation than the surrounding water table, then water is moving from the aquifer toward the stream. In a confined aquifer, water also moves from higher to lower head. And water moves between aquifers across the confining layer in the direction of higher to lower head.

Monitoring regional hydraulic heads require a sufficiently dense network of monitoring wells to account for natural heterogeneities in aquifer properties as well as variations in aquifer recharge, discharge, and pumping. Corrections to hydraulic head due to barometric and tidal influences are also needed.

The magnitude of water movement is a function of the *hydraulic gradient*, ∇h , or the change in hydraulic head with distance, $\nabla h = [\partial h/\partial x, \partial h/\partial y, \partial h/\partial z]$ in three dimensions, or $\nabla h = dh/dx$ in one dimension. To simplify the analysis of groundwater flow, vertical flow within an aquifer is often ignored, as is horizontal flow within a confining layer.

The hydraulic gradient between two wells, $\nabla h = \Delta h/\Delta x$, is found by taking the difference in water surface elevation between the two wells, Δh , divided by the distance between the two, Δx . Water levels in four different wells are the minimum required to find the three-dimensional hydraulic gradient *vector*.

The water *flux*, or *darciian flux*, is the rate of water movement and is governed by *Darcy's law*: $q = -K\nabla h$. Note that K incorporates the fluid and geologic properties, whereas ∇h incorporates the driving force. The equation contains a minus sign because flow is in the opposite direction of the hydraulic gradient.

Anisotropy refers to geologic media that have greater hydraulic conductivity in one direction than another. In this case, the hydraulic conductivity must be treated as a *tensor* rather than a *scalar*.

A general conservation equation can be written for groundwater flow, $K [\partial^2 h/\partial x^2 + \partial^2 h/\partial y^2 + \partial^2 h/\partial z^2] = S_y \partial h/\partial t$, for unconfined aquifers, and $T [\partial^2 h/\partial x^2 + \partial^2 h/\partial y^2] = S \partial h/\partial t$ for confined aquifers without leakage.

These equations are used in computer models to predict the water table and piezometric surface in areas where field observations are lacking, or for evaluating alternative management strategies. Calibration of these models to field data is a critical part of the modeling effort and is complicated by unknown variations in subsurface properties

(e.g., faults, fractures), as well as limited understanding of discharge and recharge behavior. Field-testing strategies to reduce these uncertainties should be an integral part of evaluating model performance.

The pore-water velocity, v , is of interest when examining transport of pollutants. Unlike in surface water, the water velocity is not the same as the water flux. This is because the porosity affects the water velocity by limiting the cross-sectional area available for flow. That is, water flowing through a medium with higher porosity moves slower than the same amount of water flowing through lower porosity.

As noted previously, the water velocity is equal to the darcian flux, q , divided by the porosity, n , $v = q/n$. (Note that the two velocities are equal when the porosity equals one.) Field-testing strategies to determine aquifer porosity usually relies on core samples, or on cross-hole tracer tests. In general, however, preferential flow that bypasses the bulk of the porosity leads to more rapid migration of tracers and pollutants. The *effective porosity*, n_e , is a measure of the voids that dominate fluid flow and is less than the bulk porosity, $n_e < n$. Environmental tracers that have migrated over longer distances and time are likely to be better indicators of preferential flow patterns than artificial tracer tests conducted over shorter distances and time.

6.6 SOIL WATER

Between the ground surface and the saturated zone lies a region of partially saturated geologic media, which is called the *unsaturated* or *vadose zone* (Figure 6.13). This region

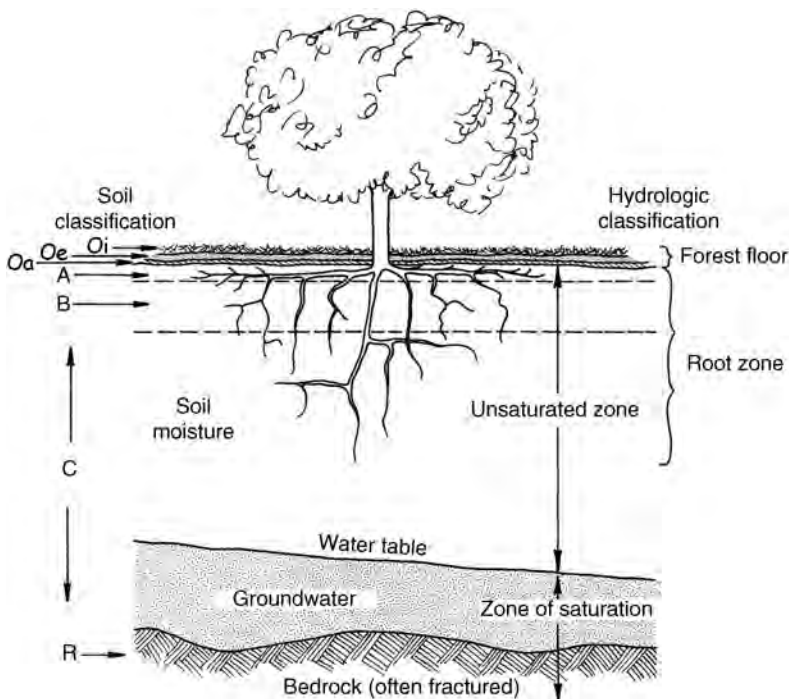


FIGURE 6.13 Subsurface hydrology showing unsaturated (vadose) zone.

contains *soil water*, which is bound by capillary forces to soil surfaces. Many pores contain air because they are not fully saturated with water.

6.6.1 Infiltration

Soils naturally absorb precipitation as long as their saturation and permeability are not limiting. One goal of soil management is to determine the maximum rate of soil infiltration over time and to find ways to increase and maintain this rate. Typically, the maximum rate of water uptake occurs at the beginning of a precipitation (or irrigation) event, and declines over time. This occurs because the hydraulic gradient that drives water flow decreases over time as the soil becomes saturated.

Field measurements of infiltration rates can be used to provide estimates of soil hydraulic properties, especially the saturated and unsaturated hydraulic conductivities and diffusivity. Two types of methods are commonly used to determine infiltration rates, *infiltrimeters* (also called *permeameters*) and *rainfall simulators*. Infiltrimeters are surface or borehole tools that apply water at the soil surface, much like the way surface irrigation might occur. Rainfall simulators generate water droplets that mimic natural rainfall conditions, and better reproduce the effects of raindrop impact on the soil surface.

The *double-ring infiltrimeter* is a simple method for determining soil infiltration rates for surface-saturated conditions (Figure 6.14). The purpose of the outer ring is to ensure that vertical water movement into the inner ring is maintained. Two methods are used to determine the water infiltration rate. For the *falling-head method*, the rate of water-level decline in the inner ring is used. For the *constant-head method*, water is added to maintain a uniform level in both rings, and the amount of water required to maintain the inner ring is measured. The constant-head method yields a higher infiltration rate than the falling-head method. One property of the double-ring method is that the water depth is greater than would occur during most rainfall events, but may be consistent with conditions during flood or furrow irrigation.

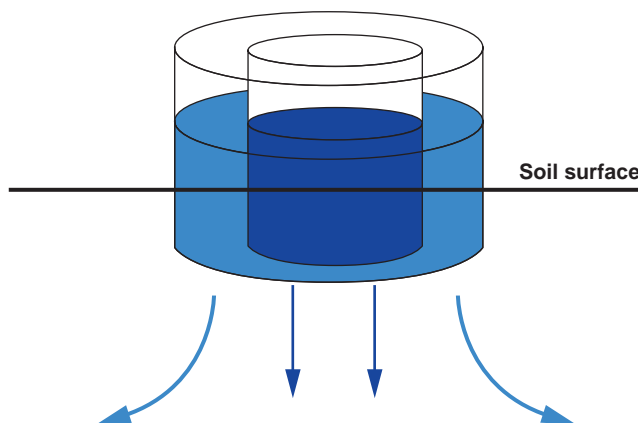


FIGURE 6.14 Double-ring infiltrimeter used to measure the maximum rate of water infiltration in soils. Water levels in both inner and outer rings are maintained at constant levels. Flow into soil from outer ring is divergent, and not measured. Flow into the soil through the inner ring is measured and is assumed vertical.

Alternatively, a *disk permeameter* can be employed to maintain a more stable and natural hydraulic potential at the soil surface. Both small positive and negative potentials can be maintained to ensure that infiltration curves more closely resemble those anticipated under natural rainfall conditions.

The *borehole permeameter* is useful for determining water movement from shallow boreholes into unsaturated soils. Rather than examining vertical infiltration from surface applications, this method relies on the infiltration through the borehole walls, which is primarily horizontal near the walls. Care must be taken, however, to remove any smearing of clays that may form on the borehole wall during its installation.

Infiltrimeters and permeameters do not account for surface disturbances associated with raindrop impact. During a rainstorm, the kinetic energy of falling raindrops can disturb the soil surface, resulting in clogging of pores, and an associated decrease in the rate of infiltration.

Rainfall simulators are commonly used to generate water droplets to simulate natural rainfall conditions. The simulator consists of a sprinkler head that sprays water over an area with a well-defined drop size and energy content. The infiltration rate is found by subtracting the measured runoff from the application rate.

6.6.2 Percolation

Although infiltration refers to the movement of water into soil across the soil surface, *percolation* refers to the movement throughout the unsaturated zone. Water can be removed from the soil surface by evaporation, and from deeper in the soil column by plant root uptake for transpiration. Water that moves downward past the root zone is termed *deep percolation*. Deep percolation continues to move downward until it reaches the saturated zone, where it contributes to groundwater *recharge*.

Percolation rates can be measured by installing *horizontal collection pans* that intercept percolating water as it moves downward. The rate can also be estimated by measuring the soil-water content using *time domain reflectometers* or the matric tension using *tensiometers* or *psychrometers*, and then using these observations to estimate the unsaturated hydraulic conductivity, $K(\psi)$, where ψ is the *matric tension*, or negative pressure head. The unsaturated hydraulic conductivity is independently determined from laboratory or field measurements. The flow rate is determined by assuming a unit vertical hydraulic gradient, $dh/dz = 1$, so that $q = K(\psi)$.

6.6.3 Hydraulic Properties

The *water content*, θ , is the volume of water per unit volume of soil and is less than or equal to the porosity, $n \geq \theta \geq 0$. The *relative saturation*, Θ , is the ratio of the soil water content to the porosity, $\Theta = \theta/n$, so that $1 \geq \Theta \geq 0$.

The *matric* or *soil-water tension*, ψ , is a measure of the force which binds water to the soil surface. Water films near the mineral surface are bound more tightly than water at a greater distance. For example, the height of fluid rise in a capillary tube is greater in smaller diameter tubes, $\psi = (2\sigma \cos\alpha)/r$, where σ is the solid-liquid-air interfacial (surface) tension, α is the solid-liquid contact angle, and r is the pore radius. As a result, the matric tension decreases as larger pores are filled with water.

The relationship between the soil-water content and the matric tension is described using the soil-water *retention* or *characteristic curve*, $\theta(\psi)$. Certain tensions are especially important and are referred to as

Saturation (SAT): The water content when the pores are entirely filled with water. Saturation corresponds to matric potentials of zero (or at positive pressures). At saturation, the volumetric water content equals the porosity. The saturated water content is determined by oven drying the sample and finding the weight gain after immersion.

Field Capacity (FC): The water content held after rapid gravitational drainage has occurred, corresponding to matric tensions between 100 and 333 mbar.

Wilting Point (WP): The amount of water held when plant roots can no longer extract water from the soil. This tension is commonly assumed to be 15 bars.

Air Dry (AD): The amount of water held by soil when it is exposed to the atmosphere, a function of the relative humidity (moist air causes soils to be wetter than soils left in dry air).

Oven Dry (OD): The matric tension that corresponds to a soil that has been in an oven until the weight stabilizes, assumed to be approximately 10 kbar.

Plant Available Water (AW): The difference between the field capacity and the wilting point, $AW = FC - WP$.

Each soil has to be characterized to define its characteristic curve due to unique variations in pore and particle size distributions. The curve can be determined using *porous plates*, *psychrometry*, *hanging columns*, or *centrifuge methods*.

A prominent feature of these curves is their *hysteresis*, that is, wetting curves have a different shape than drying curves. This leads to *scanning curves* that trace the water saturation as the matric tension is reversed.

The *unsaturated hydraulic conductivity*, $K(\psi)$, is a measure of the ability of water to move through unsaturated soils held at a constant saturation. Wet soils display a higher $K(\psi)$ than dry soils because more of the pores are filled with water and contribute to fluid flow. The *relative conductivity*, K_r , is defined using $K_r = K(\psi)/K$, so that $1 \geq K_r \geq 0$ where $K_r = 1$ for fully saturated conditions ($\psi = 0$). The relationship between K_r and matric potential, ψ , is called the *relative permeability curve* and can be determined using *Tempe cells*, *hanging columns*, or *centrifuge methods*.

The slope of the retention curve, $d\theta/d\psi$, is called the *specific water capacity curve*, C . This function is used to determine the *soil-water diffusivity*, $D(\psi) = K(\psi)/C(\psi)$, which is used to describe dynamic flow behavior in unsaturated soils. The diffusivity can be found using the same methods as used for determining the unsaturated conductivity.

6.7 WATER QUALITY

The need for water to sustain human and ecosystem health drives efforts to measure and understand water quality behavior in all aspects of hydrology, from atmospheric precipitation (e.g., acid rain), to surface water (e.g., fish health), groundwater (e.g., toxic chemicals), and soil water (e.g., organics and metals that might accumulate in crop tissue).

This section takes a broad view of water quality; one that focuses on the *physical*, *chemical*, and *biological* integrity of water resources. By integrity, we mean the suitability of the water for designated uses such as fishing, swimming, and water supply.

6.7.1 Physical

Physical water quality refers to the ability of the channel habitat to support native plants and animals. Most streams support natural riparian and hyporheic zones, floodplains, and riverine wetlands (Figure 6.15). These provide key feeding areas, refugia, and substrate for the entire life cycle of the resident aquatic species. Disruption of the habitat by dams, channeling, riparian encroachment, and loss of wetland habitat can alter the natural flow rhythms that support a healthy ecosystem (Figure 6.16).

The *channel structure* refers to the structural and hydraulic properties of the channel. For example, a *pool-and-riffle habitat* refers to an alternating sequence of calm, deep segments with fast, shallow segments. The *stream-bank condition* is another important feature, in that banks may collapse due to disturbance or excessive stream velocity, altering the natural refugia that stream banks provide. The amount and type of large woody debris and leaf litter are important attributes that favor specific ecologic communities. The natural sinuosity (or tortuosity) of a stream channel also affects channel hydraulics and bank condition. Characterizing the type of channel structure requires stream surveys that map the important channel components.

Benthic sediments affect not only hyporheic flows but also fish nesting behavior and plant establishment. Cobbles and gravels provide habitat for macroinvertebrates, while coarse sediments with few fines allow water to migrate through channel sediments, maintaining benthic oxygen concentrations, and moderating stream temperatures. Increasing fine sediments may adversely affect native biological productivity, and may favor the establishment of invasive species. Surveys of benthic sediments using *pebble counts* or *particle size distributions* can be used to characterize stream conditions.



FIGURE 6.15 Sandy Creek, a stream in good physical condition. Note natural riparian vegetation and stream substrate.



FIGURE 6.16 Tanyard Branch, a stream in poor physical condition. Note lack of riparian vegetation and artificial substrate.

Riparian zones and *wetlands* help to sequester upland sediment and nutrient inputs, and also moderate hydraulic disturbances such as droughts and floods. Riparian zones and forested wetlands also moderate stream temperatures by providing a canopy that buffers the daily and seasonal temperature cycles. Water temperatures can be measured manually using a thermometer, or using *thermistors* or *thermocouples* that provide electronic outputs.

6.7.2 Chemical

The chemical condition of water includes those dissolved and suspended components that favorably or adversely affect the integrity for designated uses. In many cases, the desirable conditions should mimic natural conditions, that is, the absence of synthetic chemicals, or within a normal range of natural concentrations.

Many of the chemical water quality measurements can be obtained manually in the field, using *peristaltic pumps* that collect *discrete* or *composite* water samples, or using sensors connected to display units (Figure 6.17) or *dataloggers* to control sensors that provide data. Other measurements must be performed using laboratory equipment.

Samples for laboratory analysis should be collected so that a good representation of hydrologic conditions is achieved. For example, the *DH-48 integrated sampler* is used for collecting surface water from across the full depth and width of a stream. Wells should be thoroughly purged using *bailers* or *peristaltic pumps* to ensure that fresh groundwater is sampled. *Soil suction samplers* are useful for extracting pore waters. Regardless of the method employed, standard methods for collecting data should be used.

pH is a measure of the proton, or *hydronium*, H_3O^+ , *activity*, or *concentration*. pH is commonly measured using a glass *ion selective electrode* or a *pH indicator* that changes color with pH. pH is related to the *alkalinity* and *acidity*. Alkalinity is determined for water samples with a pH greater than 4.5, whereas acidity is determined for samples less



FIGURE 6.17 Example of a portable water quality probe that monitors multiple water quality parameters, including temperature, specific conductance, pH, dissolved oxygen, turbidity, and water depth.

than 4.5. For alkalinity, acid is titrated with the water sample to lower the pH to 4.5, whereas a base is titrated to raise it to 4.5 for acidity. Acid mine drainage may have high acidities, whereas carbonate waters are likely to have high alkalinities.

Although plants manufacture their own oxygen through photosynthesis, respiration by aquatic animals requires sufficient oxygen in the water to maintain their metabolic processes. Oxygen dissolves in water in small amounts, on the order of 10 mg/L (ppm), but is a function of water temperature and atmospheric pressure. Colder water contains higher concentrations than warmer water, and water at higher elevations has lower concentrations than at sea level (when both are at the same temperature). Dissolved oxygen concentrations are measured using *membrane galvanic sensors* or *luminescent probes*.

Dissolved oxygen concentrations may fall to *anoxic* levels (<0.5 mg/L) if respiration rates are high, or in geologic conditions where reducing conditions dominate. In these cases, the *redox* or *oxidation–reduction potential* (ORP) can be monitored using an ORP probe, which can be combined with the pH sensor.

The *electrical conductivity* is a measure of the dissolved solids in solution. Increasing concentrations of dissolved solids leads to increasing conductivities. A common unit for conductivity is $\mu\text{S}/\text{cm}$ (S is Siemen, equivalent to the older unit, mho) and is also reported as *specific conductance* when the conductivity is temperature compensated. The conductivity probe consists of two electrodes a fixed distance apart. The probe is immersed in water, and the electric current and voltage are monitored between the electrodes. The conversion from specific conductance to dissolved solids concentration should be calibrated using the types of ions likely to be present in the water sample.

Turbidity is a measure of the clarity of water. Higher concentrations of *suspended* solids, and *dissolved* and *particulate* organic matter can increase the water turbidity. Turbidity is commonly measured using a portable *nephelometric turbidimeter*, which determines the amount of scattering from a light source. The amount of *filterable solids*



FIGURE 6.18 Example of a portable water-quality testing device for determining nutrient concentrations.

can be found by filtering the water samples through a fine medium. *Centrifuging* can also be used to concentrate the heavier solids from the water, and then oven dried to obtain the total solids. For deeper water bodies such as lakes and reservoirs, a *Secchi disk* can be lowered below the water surface. The depth at which the disk is no longer visible is called the *Secchi depth*.

Nutrients, primarily nitrogen and phosphorus, contribute to lake and estuary *eutrophication*. Multiple forms of these nutrients might be present, such as sorbed to sediments, as part of organic molecules (e.g., amino acids) both living and dead, and in multiple inorganic forms (especially for nitrogen, which could be found as nitrate, ammonium, and so on). Field estimation of nutrients using ion selective electrodes is possible, but *colorimetric* techniques (e.g., the Murphy–Riley technique for phosphorus) are likely to be more reliable (Figure 6.18). Laboratory estimates of nutrients also use colorimetry as well as *chromatography*.

Samples for *inorganic* (e.g., heavy metals) and *organic* (e.g., fuels, pesticides) analysis are normally collected in the field, stabilized, and promptly returned to a laboratory for further analysis. The use of special sampling techniques and equipment is critical for many constituents, especially for those at very low concentrations.

6.7.3 Biological

A wide range of organisms can survive in both the surface and subsurface environments. While most of these organisms are desirable, such as native fishes and plankton, others are not.

Water-borne pathogens and invasive species can be disruptive to humans and ecosystems. Microbial pathogens include viruses, bacteria, fungi, and protozoa. Sampling requires collection and laboratory culturing and analysis. Invasive species determination requires field surveys and traps to determine the presence or absence of the species.

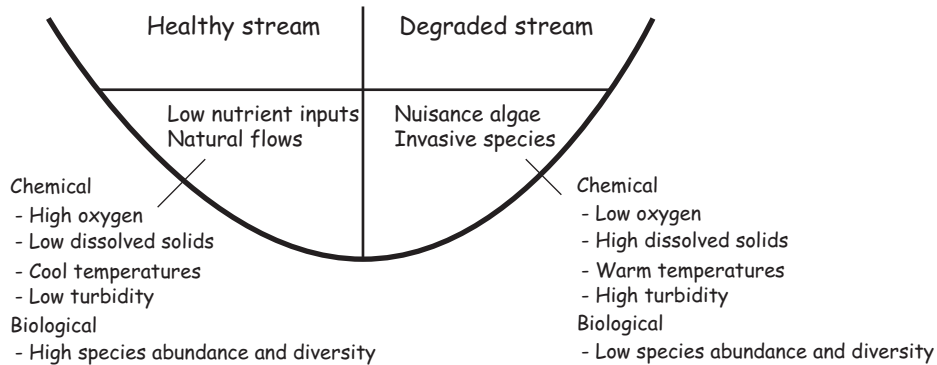


FIGURE 6.19 Example of the use of biotic indicators with chemical water quality measurements.

Besides pathogens, other measures of stream health have been established that utilize an *index of biological integrity* (IBI), which is an effort to use aquatic surveys to determine the status of ecosystem health (Figure 6.19). Various indicators can be used, including algae, macroinvertebrates, and fish. These indicators are usually geographically specific and are commonly limited to specific times of the year when the target index is most effectively estimated. IBIs rely on the preference of some organisms for natural, or undisturbed, conditions, while other organisms are more tolerant of pollution and disturbance. The advantage of biological indicators over water chemistry lies in their long-term presence that reacts to occasional disturbances, which may not be observed using standard chemical sampling.

SUGGESTED READINGS

- Allen RG, Walter IA, Elliot RL, Howell TA, Itenfisu D, Jensen ME, Snyder RL, editors. *The ASCE Standardized Reference Evapotranspiration Equation*. American Society of Civil Engineers; 2005. p. 216.
- Allen RG, Pereira LA, Raes D, Smith M. *Crop Evapotranspiration Guidelines for Computing Crop Water Requirements*. Irrigation and Drainage Paper No. 56. Food and Agriculture Organization of the United Nations; 2005. p. 328.
- Bear J. *Dynamics of Fluids in Porous Media*. Dover Publications; 1988. p. 784.
- Boyd CE. *Water Quality: An Introduction*. Springer; 2000. p. 330.
- Conti ME, editor. Bioindicators and biomarkers for environmental quality and human exposure assessment. *Biological Monitoring: Theory and Applications*. WIT Press; 2008. p. 256.
- Driscoll FG. *Groundwater and Wells*. 2nd ed. Johnson Screens; 1986. p. 1089.
- Ertud K, Mirza I, editors. *Water Quality: Physical, Chemical and Biological Characteristics*. Nova Science Publication; 2010. p. 277.
- Fetter CW. *Applied Hydrogeology*. 4th ed. Prentice Hall; 2000. p. 598.
- Fetter CW. *Contaminant Hydrogeology*. 2nd ed. Waveland Press; 2008. p. 500.
- Gupta RS. *Hydrology and Hydraulic Systems*. 3rd ed. Waveland Press; 2007. p. 896.
- Hartmann DL. *Global Physical Climatology*. Academic Press; 1994. p. 411.
- Hillel D. 1998. *Environmental Soil Physics*. Academic Press; 1994. p. 771.

- Jury WA, Horton R. *Soil Physics*. 6th ed. Wiley; 2004. p. 384.
- Loeb SL, Spacie A, editors. *Biological Monitoring of Aquatic Systems*. CRC Press; 1994. p. 400.
- Maidment DR. *Handbook of Hydrology*. McGraw-Hill Professional; 1993. p. 1424.
- Mitsch WJ, Gosselink JG. *Wetlands*. 4th ed. Wiley; 2007. p. 600.
- Robinson PJ, Henderson-Sellers A. *Contemporary Climatology*. 2nd ed. Prentice Hall; 1999. p. 352.
- Sparks DL. *Environmental Soil Chemistry*. 2nd ed. Academic Press; 2002. p. 352.
- Stumm W, Morgan JJ. *Aquatic Chemistry*. 3rd ed. Wiley-Interscience; 1996. p. 1040.
- Todd DK, Mays LW. *Groundwater Hydrology*. 3rd ed. Wiley; 2008. p. 656.
- Ward AD, Trimble SW. *Environmental Hydrology*. 2nd ed. CRC Press; 2003. p. 504.

Additional information can also be found online at the U.S. Geological Survey website, “Techniques of Water Resources Investigations Reports.”⁴

⁴<http://pubs.usgs.gov/twri/>

MOBILE SOURCE EMISSIONS TESTING

MOHAN VENIGALLA

- 7.1 Testing for regulatory compliance
 - 7.1.1 The federal test procedure
 - 7.1.2 Testing for inspection and maintenance
 - 7.1.3 Emission testing with on-board portable emissions monitoring systems
 - 7.1.4 Testing for special purpose emissions studies

References

This chapter provides a broad perspective on emission testing of new and in-use vehicles, particularly in the United States. Under a mandate from the U.S. Congress, The U.S. Environmental Protection Agency (US EPA or EPA) sets emissions standards for vehicles. These standards mainly differ among engine and/or vehicle classes, year of manufacturing. Once EPA sets emission standards manufacturers are required to produce engines that meet those standards within a given timeframe of the corresponding implementation schedule.

The Clean Air Act (CAA) requires every engine and motor vehicle within the chain of commerce in the United States to meet a set of emission standards and conformity requirements. Anyone wishing to sell an engine or vehicle within the United States must demonstrate compliance with the CAA and all applicable EPA regulations. Upon adequate representation of conformity by the manufacturer and possibly confirmatory testing by EPA, EPA may issue a certificate of conformity that provides authorization for production and sales within the United States (CA-ARB, 2002).

The certification process begins when a manufacturer submits an application for certification to EPA for a group of vehicles or engines having similar design and emission characteristics. EPA requires manufacturers to provide detailed information to show that they have met all of the applicable requirements to qualify for the certificate of conformity. The application for certification describes those vehicles or engines

specifically covered by the certificate of conformity. The certificate is a license to produce and sell the vehicle and covers only those vehicles or engines specifically described in the application.

All of EPA's emission regulations specify test procedures to measure engine or vehicle emission levels. EPA uses the test results to determine compliance with the applicable emission standards.

The number and types of tests vary according to the regulated sector. Certification testing is a form of compliance testing that is required as a condition of certification and is generally performed prior to issuing a certificate. In-use testing occurs after the vehicles or engines have been certified, generally on privately used vehicles or engines. Production line (or assembly line) testing audits emission levels of vehicles or engines that are in production, but not yet in service, to confirm that the manufacturer is building compliant vehicles (CA-ARB, 2002).

7.1 TESTING FOR REGULATORY COMPLIANCE

Vehicle testing is usually performed in laboratory environment under simulated real-world driving conditions. Drive train is placed on a dynamometer mechanism where the circular motion of the drive train is transferred to spinning steel cylinders of the dynamometer. A dynamometer is a machine that measures the torque and power produced by an engine. It applies various loads to an engine and is usually connected to a computer that can analyze and calculate all the aspects of engine operation measured.

Dynamometers are particularly useful in designing and refining engine technology. They can help identify how an engine or its drive train needs to be modified or tuned to achieve more efficient power transfer.

There are two types of dynamometers. Engine dynamometer measures engine performance only. The engine is removed from the vehicle and mounted onto a special frame for this type of dynamometer. It is coupled directly to the engine flywheel and measures performance independent of the vehicle's drive train—such as its gearbox, transmission, or differential.

The second type of dynamometer is called chassis dynamometer, which measures the power from the engine through a vehicle's driven wheels. The drive train is mounted on rollers, and the vehicle is fixed to the ground to prevent jumping when it is driven during testing. As the vehicle is driven in gear, it turns the rollers without moving. The power output and emissions are measured using various techniques.

7.1.1 The Federal Test Procedure

One of the fundamental assumptions underlying on-road emissions inventory development is that driving behavior can be captured and reproduced in a laboratory setting. The Federal Test Procedure or FTP was developed in the late 1960s and early 1970s as a model of a typical commute in an urban area. The FTP was neither Federal nor typical. It is actually the composite of several trips to and from the then headquarters of the California Air Resources Board in Downtown Los Angeles.

When first utilized, the FTP (also referred to as CVS72), was a two mode or two-bag test. The term "bag" refers to the Tedlar bags in which exhaust is stored for analysis. The first mode of the test is referred to as a "cold start." Cold start refers to a physical turn of the

ignition when the engine is actually at ambient temperature after several hours of turn off. Cold starts are associated with excessive cranking or starting emissions of a vehicle. The cold-start mode is followed by a cold transient model for 505 s in duration with a maximum speed of 56.7 miles/h and an average speed of 25.6 miles/h. The maximum acceleration and deceleration rates of 5.6 and 5.2 miles/h/s were mitigated to 3.3 miles/h/s because of the limitations of the belt-driven dynamometers used at that time (CA-ARB, 2002).

Mode two, or bag two, of the FTP is referred to as the “hot stabilized” portion of the test. With a maximum speed of 34 miles/h and an average speed of 16, bag two is slower than bag one because bag one contains freeway driving. Again the maximum and minimum acceleration and deceleration rates of 5.4 and 8.0 miles/h/s were clipped at plus or minus 3.3 miles/h/s. Taken together, the two modes of the CVS72 comprise a driving cycle, which is 1372 s in duration with an average speed of 19.6 miles/h (CA-ARB, 2002).

The CVS75 cycle added a third mode to the test. Upon completion of bag two, the vehicle is switched off and allowed to sit or soak for 10 min after which the vehicle is restarted and mode one of the test is repeated. This is referred to as the hot-start portion of the test. The acceleration, deceleration, cruise, and idle cycle of this cycle are illustrated in Figure 7.1.

The laboratory setup includes instrumentation that can control engine in such a way that the drive train will mimic this driving cycle. Also included in the setup are instrumentation for emission collection and analysis. Figures 7.2 and 7.3 illustrate the schematic and actual setup of a laboratory set up for the FTP Test.

The parameters of LA4 cycle used in FTP are summarized in Table 7.1. The FTP test data for new vehicles are used by EPA to develop emissions models such as the MOBILE model and its successor MOVES model.

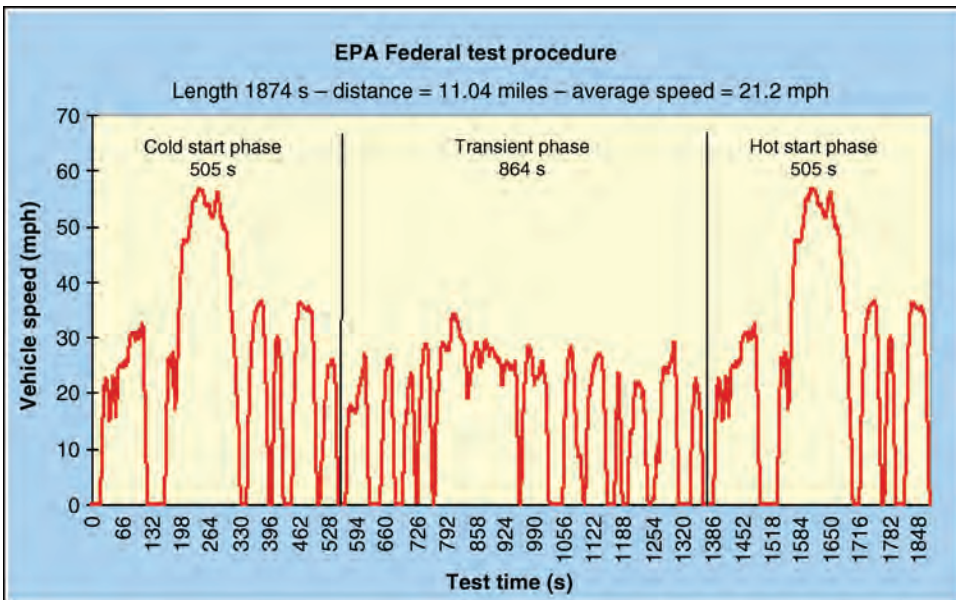


FIGURE 7.1 The CVS75 (LA4) cycle with three phases. *Source:* USEPA.

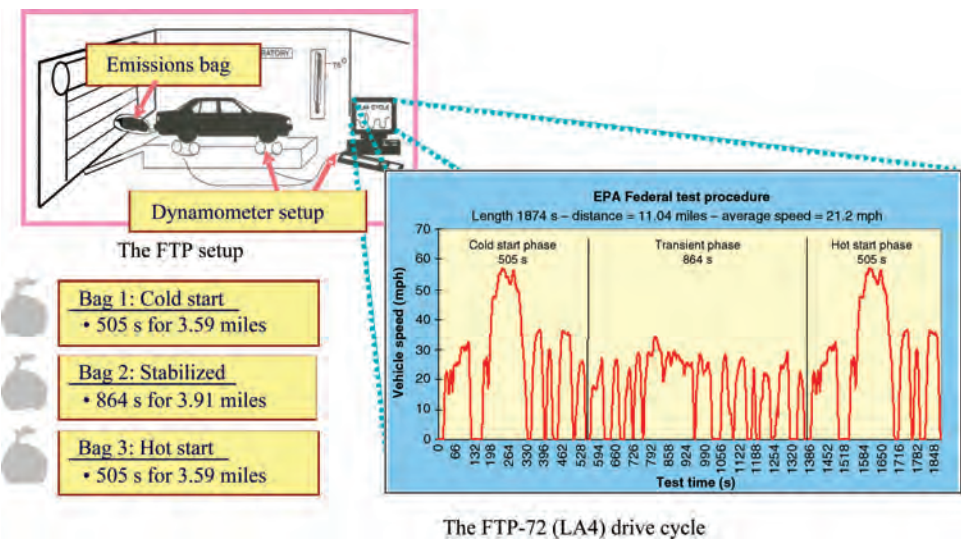


FIGURE 7.2 Schematic representation of FTP.

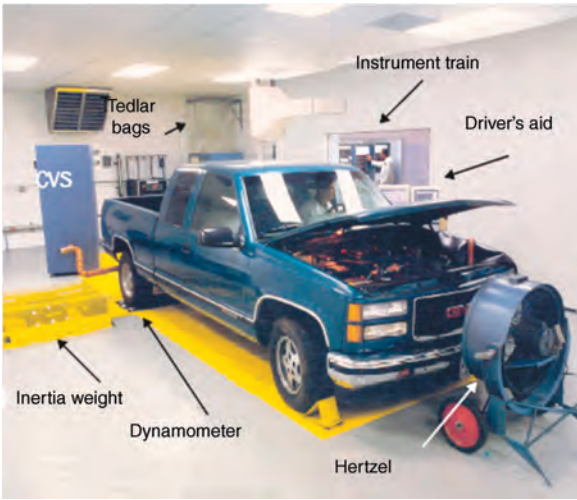


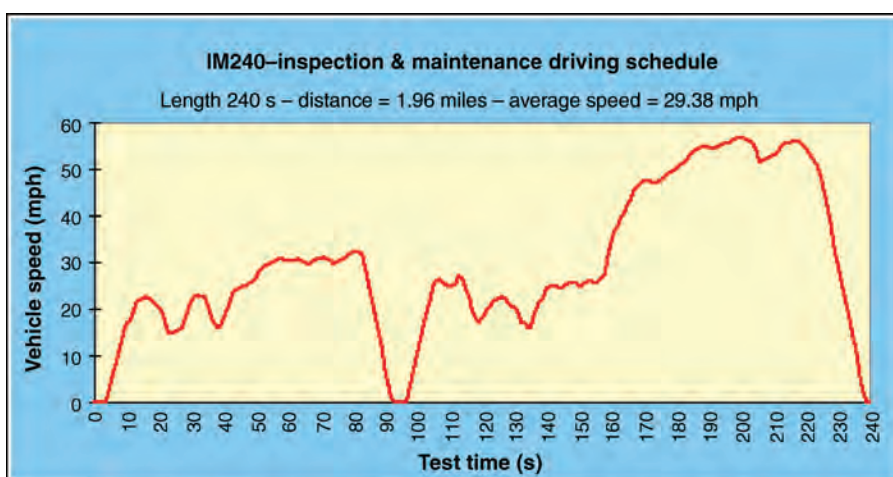
FIGURE 7.3 Laboratory setup of the FTP (Illustration courtesy: The Federal Test Procedure & Unified Cycle California Air Resource Board, Volume 1, Issue 9, June 2002).

7.1.2 Testing for Inspection and Maintenance

Vehicles registered in specific classifications of nonattainment areas are periodically required to go through testing for compliance with Inspection and Maintenance

TABLE 7.1 FTP Parameters

Drive Cycle	LA4 Cycle
Average speed	19.6 mph
Maximum speed	56.7 mph
Average maximum speed	N/A
Percent idle	19.0
Stops per mile	2.41
Maximum acceleration	1.48 m/s ²
Average maximum acceleration	N/A
Cycle length	7.5 miles

**FIGURE 7.4** IM240 drive cycle used in several I&M programs.

(I&M) programs. While the FTP is used for testing new vehicles, authorized emission-testing stations with chassis dynamometer facilities use a test cycle for shorter duration. Most commonly used driving cycle for I&M programs is IM240 Drive Cycle (Figure 7.4).

7.1.3 Emission Testing with On-Board Portable Emissions Monitoring Systems

Although the FTP presents a standardized method to assess the emissions performance of new vehicles, the driving schedule used in the FTP test does not reflect real-world driving conditions. Dynamometer tests are often used in regulatory procedures to check compliance of new vehicles with emission standards or to inspect in-use vehicles. The data obtained from driving cycles are also used to develop emission estimation models, such as EMFAC7F, MOBILE6 (also in MOVES), and UC Riverside's modal emissions models (Roughail et al., 2000; Frey, 2003).

Remote sensing (RS) is another method for measuring vehicle emissions. RS uses infrared (IR) and, in some cases, ultraviolet (UV) spectroscopy to measure the concentrations of pollutants in exhaust emissions as an on-road vehicle passes a sensor on the roadway. There are several applications of remote sensing in mobile emissions determination. These include monitoring of emissions to evaluate the overall effectiveness of inspection and maintenance programs; identification of high emitting vehicles for inspection or enforcement purposes; and development of emission factors (Roughail et al., 2000).

The major advantage of remote sensing is that it is possible to measure a large number of on-road vehicles (e.g., thousands per day). The major disadvantage of remote sensing is that it only gives an instantaneous estimate of emissions at a specific location. There are constraints on the siting of remote sensing devices (RSDs) that make it impractical to use remote sensing as a means for measuring vehicle emissions at many locations of practical interest, such as close to intersections or across multiple lanes of heavy traffic. Furthermore, remote sensing is more or less a fair weather technology (Roughail et al., 2000).

On-board emissions measurement is widely recognized as a desirable approach for quantifying emissions from vehicles since data are collected under real-world conditions at any location traveled by the vehicle. Variability in vehicle emissions as a result of variation in facility (roadway) characteristics, vehicle location, vehicle operation, driver, or other factors can be represented and analyzed more reliably than with the other methods. This is because measurements are obtained during real-world driving, eliminating the concern about nonrepresentativeness that is often an issue with dynamometer testing, and at any location, eliminating the siting restrictions inherent in remote sensing.

Once such device in use, OEM-2100, is capable of testing electronically controlled light-duty passenger vehicles and light trucks 1996 and newer with on-board diagnostics (OBD). The unit includes a durable touch-screen computer and comes standard in powder-coated aluminum housing.

The system connects to the vehicle via three interfaces. The vehicle cigarette lighter provided electrical power to the PEMS units; the vehicle OBD port provided access to the engine data stream; and the PEMS sample line and exhaust probe were inserted into the tailpipe. Exhaust gas and particulate concentrations were reported by their respective analyzers. The Portable Emission Measurement Systems (PEMS) in use captured HC, CO, CO₂, and NO_x emissions from the gasoline-fueled vehicles on a second-by-second basis, correlating the emissions captured with engine-operating characteristics such as engine RPM, fuel flow, ignition timing advance, throttle position, mass air flow, and vehicle speed (Hart 2002; USEPA, 2003). These operating characteristics were gathered from each test vehicle's engine control unit via the on-board diagnostics—OBD II—port. Global positioning system (GPS) units were used concurrently with PEMS, providing positional data, and cross-referencing speed and acceleration data with that gathered by the PEMS unit. The PEMS units were powered by each vehicle's 12 V system (via the cigarette lighter or auxiliary power ports). Processing the emissions data were managed by the PEMS on-board computer, and stored to the PEMS hard disk.

For example, PEMS were used in a study to measure differences in vehicle emissions between HOV and GP facilities. Several hundred miles of on-road emissions data were collected with vehicles instrumented with PEMS operated in HOV and GP lanes during

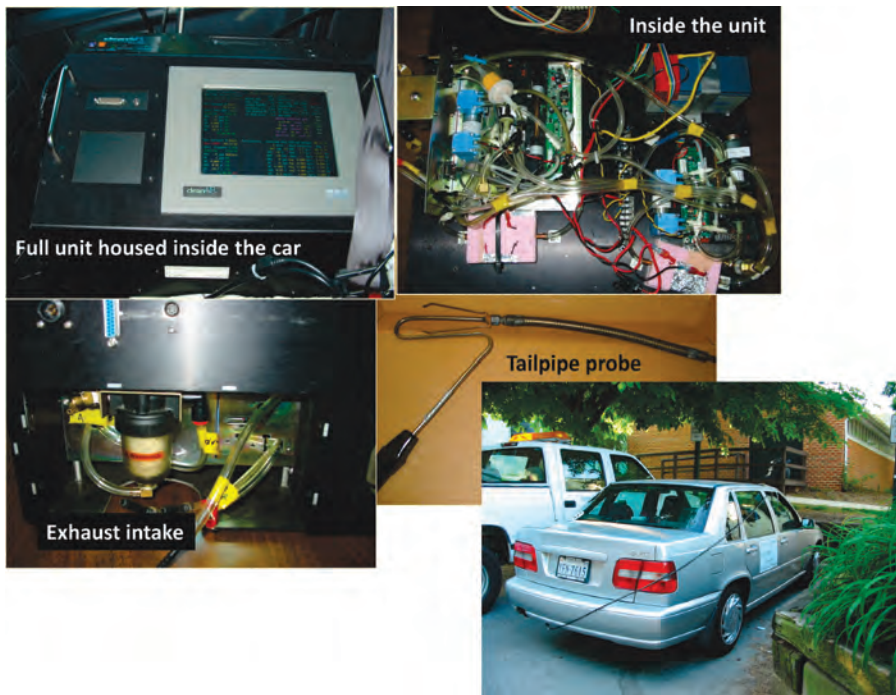


FIGURE 7.5 Setup of a portable emission monitor system for field emissions measurement study.

peak hours in the metropolitan Washington, DC area (Krimmer and Venigalla, 2006). Illustrated in Figure 7.5 is the OEM-2100 setup for this study.

However, on-board emissions measurement has not been widely used because it has been prohibitively expensive (Frey, 2003).

7.1.4 Testing for Special Purpose Emissions Studies

Whether the testing is done for regulatory compliance (FTP, IM240, remote sensing, etc.), or for testing on-road traffic and transportation strategies using portable emissions monitors, and emissions testing is generally an expensive proposition. In addition to requiring expensive equipment, acquiring and maintaining such systems also requires skilled staff and high maintenance costs. A few public universities have laboratory facilities to facilitate design and execution of various types of emissions testing studies.

For example, the Center for Environmental Research and Technology (CE-CERT) in Burns College of Engineering at University of California, Davis has state-of-the-art laboratory equipment to facilitate various types of emissions studies such as heavy-duty chassis and engine dynamometers, a light-duty chassis dynamometer, a portable emissions laboratory, and a chemical analysis laboratory (CE-CERT, 2011). Figure 7.6 shows the heavy-duty chassis dynamometer and mobile emissions laboratory at CE-CERT.



FIGURE 7.6 Heavy-duty Chassis Dynamometer and Mobile Emissions Laboratory at CE-CERT, University of California, Riverside.

REFERENCES

- CA-ARB. The Federal Test Procedure & Unified Cycle. *California Air Resources Board* 2002; 1 (9): 1–2.
- CE-CERT. *Emissions and Fuel Laboratories*. Riverside, CA: Burns School of Engineering, University of California, Riverside;2011.
- Frey C. *Use of On-Board Tailpipe Emissions Measurements for Development of Mobile Source Emission Factors*. US EPA;2003.
- Hart CJ. *EPA's Onboard Analysis Shootout: Overview and Results*. Assessment and Standards Division. U.S. Environmental Protection Agency;2002.
- Krimmer M, Venigalla AM. *Measuring Impacts of High Occupancy Vehicle (HOV) Lane Operations on Light Duty Vehicle Emissions: An Experimental Study with Instrumented Vehicles*. Journal of Transportation Research, National Research Council;2006.
- Roughail N, Frey H, Unal A, Dalton R. *ITS Integration of Real-Time Emissions and Traffic Management Systems*. Transportation Research Board, National Research Council;2000.
- USEPA. *Environmental Technology Verification Program Joint Statement-CATi*. Washington, DC: AMS Center, USEPA;2003.

PART II

MECHANICAL AND BIOMEDICAL ENGINEERING

8

DIMENSIONS, SURFACES, AND THEIR MEASUREMENT

MIKELL P. GROOVER

- 8.1 Dimensions, tolerances, and related attributes
 - 8.1.1 Dimensions and tolerances
 - 8.1.2 Other geometric attributes
- 8.2 Conventional measuring instruments and gages
 - 8.2.1 Precision gage blocks
 - 8.2.2 Measuring instruments for linear dimensions
 - 8.2.3 Comparative instruments
 - 8.2.4 Fixed gages
 - 8.2.5 Angular measurements
- 8.3 Surfaces
 - 8.3.1 Characteristics of surfaces

In addition to mechanical and physical properties of materials, other factors that determine the performance of a manufactured product include the dimensions and surfaces of its components. *Dimensions* are the linear or angular sizes of a component specified on the part drawing. Dimensions are important because they determine how well the components of a product fit together during assembly. When fabricating a given component, it is nearly impossible and very costly to make the part to the exact dimension given on the drawing. Instead, a limited variation is allowed from the dimension, and that allowable variation is called a *tolerance*.

The surfaces of a component are also important. They affect product performance, assembly fit, and aesthetic appeal that a potential customer might have for the product. A *surface* is the exterior boundary of an object with its surroundings, which may be another

object, a fluid, or space, or combinations of all these. The surface encloses the object's bulk mechanical and physical properties.

This chapter discusses dimensions, tolerances, and surfaces—three attributes specified by the product designer and determined by the manufacturing processes used to make the parts and products. It also considers how these attributes are assessed using measuring and gaging devices. A closely related topic is inspection, which is covered in Chapter 42.

8.1 DIMENSIONS, TOLERANCES, AND RELATED ATTRIBUTES

The basic parameters used by design engineers to specify sizes of geometric features on a part drawing are defined in this section. The parameters include dimensions and tolerances, flatness, roundness, and angularity.

8.1.1 Dimensions and Tolerances

ANSI (American National Standards Institute, Inc., 1982) defines a *dimension* as “a numerical value expressed in appropriate units of measure and indicated on a drawing and in other documents along with lines, symbols, and notes to define the size or geometric characteristic, or both, of a part or part feature.” Dimensions on part drawings represent nominal or basic sizes of the part and its features. These are the values that the designer would like the part size to be, if the part could be made to an exact size with no errors or variations in the fabrication process. However, there are variations in the manufacturing process, which are manifested as variations in the part size. Tolerances are used to define the limits of the allowed variation. Quoting again from the ANSI standard (American National Standards Institute, Inc., 1982), a *tolerance* is “the total amount by which a specific dimension is permitted to vary. The tolerance is the difference between the maximum and minimum limits.”

Tolerances can be specified in several ways, illustrated in Figure 8.1. Probably the most common is the *bilateral tolerance*, in which the variation is permitted in both positive and negative directions from the nominal dimension. For example, in Figure 8.1a, the nominal dimension = 2.500 linear units (e.g., mm, in.), with an

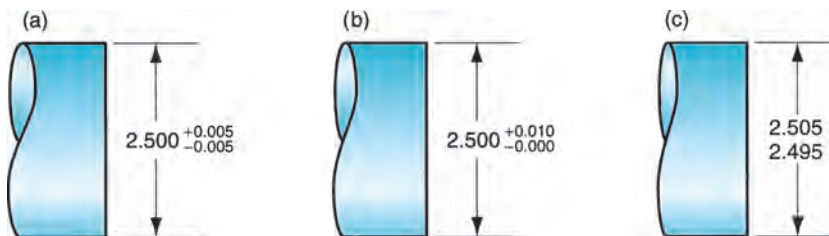


FIGURE 8.1 Three ways to specify tolerance limits for a nominal dimension of 2.500: (a) bilateral, (b) unilateral, and (c) limit dimensions.

TABLE 8.1 Definitions of Geometric Attributes of Parts

<i>Angularity</i> —The extent to which a part feature such as a surface or axis is at a specified angle relative to a reference surface. If the angle = 90°, then the attribute is called perpendicularity or squareness.	<i>Cylindricity</i> —The degree to which all points on a surface of revolution such as a cylinder are equidistant from the axis of revolution.
<i>Circularity</i> —For a surface of revolution such as a cylinder, circular hole, or cone, circularity is the degree to which all points on the intersection of the surface and a plane perpendicular to the axis of revolution are equidistant from the axis. For a sphere, circularity is the degree to which all points on the intersection of the surface and a plane passing through the center are equidistant from the center.	<i>Flatness</i> —The extent to which all points on a surface lie in a single plane.
<i>Concentricity</i> —The degree to which any two (or more) part features such as a cylindrical surface and a circular hole have a common axis.	<i>Parallelism</i> —The degree to which all points on a part feature such as a surface, line, or axis are equidistant from a reference plane or line or axis.
	<i>Perpendicularity</i> —The degree to which all points on a part feature such as a surface, line, or axis are 90 from a reference plane or line or axis.
	<i>Roundness</i> —Same as circularity.
	<i>Squareness</i> —Same as perpendicularity.
	<i>Straightness</i> —The degree to which a part feature such as a line or axis is a straight line.

allowable variation of 0.005 U in either direction. Parts outside these limits are unacceptable. It is possible for a bilateral tolerance to be unbalanced; for example, 2.500, +0.010, −0.005 dimensional units. A *unilateral tolerance* is one in which the variation from the specified dimension is permitted in only one direction, either positive or negative, as in Figure 8.1b. *Limit dimensions* are an alternative method to specify the permissible variation in a part feature size; they consist of the maximum and minimum dimensions allowed, as in Figure 8.1c.

8.1.2 Other Geometric Attributes

Dimensions and tolerances are normally expressed as linear (length) values. There are other geometric attributes of parts that are also important, such as flatness of a surface, roundness of a shaft or hole, parallelism between two surfaces, and so on. Definitions of these terms are listed in Table 8.1.

8.2 CONVENTIONAL MEASURING INSTRUMENTS AND GAGES

Measurement is a procedure in which an unknown quantity is compared with a known standard, using an accepted and consistent system of units. Two systems of units have evolved in the world: (1) the U.S. customary system (U.S.C.S.), and (2) the International System of Units (or SI, for Systeme Internationale d’Unites), more popularly known as

the metric system. Both systems are used in parallel throughout this book. The metric system is widely accepted in nearly every part of the industrialized world except the United States, which has stubbornly clung to its U.S.C.S. Gradually, the United States is adopting SI.

Measurement provides a numerical value of the quantity of interest, within certain limits of accuracy and precision. *Accuracy* is the degree to which the measured value agrees with the true value of the quantity of interest. A measurement procedure is accurate when it is absent of systematic errors, which are positive or negative deviations from the true value that are consistent from one measurement to the next. *Precision* is the degree of repeatability in the measurement process. Good precision means that the random errors in the measurement procedure are minimized. Random errors are usually associated with human participation in the measurement process. Examples include variations in the setup, imprecise reading of the scale, round-off approximations, and so on. Nonhuman contributors to random error include temperature changes, gradual wear and/or misalignment in the working elements of the device, and other variations.

Closely related to measurement is gaging. *Gaging* (also spelled *gauging*) determines simply whether the part characteristic meets or does not meet the design specification. It is usually faster than measuring, but scant information is provided about the actual value of the characteristic of interest. The video clip on measurement and gaging illustrates some of the topics discussed in this chapter.

Video Clip

Measurement and Gaging. This clip contains three segments: (1) precision, resolution, and accuracy; (2) how to read a vernier caliper; and (3) how to read a micrometer.

This section considers the variety of manually operated measuring instruments and gages used to evaluate dimensions such as length and diameter, as well as features such as angles, straightness, and roundness. This type of equipment is found in metrology labs, inspection departments, and tool rooms. The logical starting topic is precision gage blocks.

8.2.1 Precision Gage Blocks

Precision gage blocks are the standards against which other dimensional measuring instruments and gages are compared. Gage blocks are usually square or rectangular. The measuring surfaces are finished to be dimensionally accurate and parallel to within several millionths of an inch and are polished to a mirror finish. Several grades of precision gage blocks are available, with closer tolerances for higher precision grades. The highest grade—the *master laboratory standard*—is made to a tolerance of ± 0.00003 mm (± 0.00001 in.). Depending on degree of hardness desired and price the user is willing to pay, gage blocks can be made out of any of several hard materials, including tool steel, chrome-plated steel, chromium carbide, or tungsten carbide.

Precision gage blocks are available in certain standard sizes or in sets, the latter containing a variety of different-sized blocks. The sizes in a set are systematically determined so they can be stacked to achieve virtually any dimension desired to within 0.0025 mm (0.0001 in.).

For best results, gage blocks must be used on a flat reference surface, such as a surface plate. A *surface plate* is a large solid block whose top surface is finished to a flat plane. Most surface plates today are made of granite. Granite has the advantage of being hard, non-rusting, nonmagnetic, long-wearing, thermally stable, and easy to maintain.

Gage blocks and other high-precision measuring instruments must be used under standard conditions of temperature and other factors that might adversely affect the measurement. By international agreement, 20°C (68°F) has been established as the standard temperature. Metrology labs operate at this standard temperature. If gage blocks or other measuring instruments are used in a factory environment in which the temperature differs from this standard, corrections for thermal expansion or contraction may be required. Also, working gage blocks used for inspection in the shop are subject to wear and must be calibrated periodically against more precise laboratory gage blocks.

8.2.2 Measuring Instruments for Linear Dimensions

Measuring instruments can be divided into two types: graduated and nongraduated. *Graduated measuring devices* include a set of markings (called *graduations*) on a linear or angular scale to which the object's feature of interest can be compared for measurement. *Nongraduated measuring devices* possess no such scale and are used to make comparisons between dimensions or to transfer a dimension for measurement by a graduated device.

The most basic of the graduated measuring devices is the *rule* (made of steel, and often called a *steel rule*), used to measure linear dimensions. Rules are available in various lengths. Metric rule lengths include 150, 300, 600, and 1000 mm, with graduations of 1 or 0.5 mm. Common U.S. sizes are 6, 12, and 24 in., with graduations of 1/32, 1/64, or 1/100 in.

Calipers are available in either nongraduated or graduated styles. A nongraduated caliper (referred to simply as a *caliper*) consists of two legs joined by a hinge mechanism, as in Figure 8.2. The ends of the legs are made to contact the surfaces of the object being measured, and the hinge is designed to hold the legs in position during use. The contacts point either inward or outward. When they point inward, as in Figure 8.2, the instrument is an *outside caliper*, and is used for measuring outside dimensions such as a diameter. When the contacts point outward, it is an *inside caliper*, which is used to measure the distance between two internal surfaces. An instrument similar in configuration to the caliper is a *divider*, except that both legs are straight and terminate in hard, sharply pointed contacts. Dividers are used for scaling distances between two points or lines on a surface, and for scribing circles or arcs onto a surface.

A variety of graduated calipers are available for various measurement purposes. The simplest is the *slide caliper*, which consists of a steel rule to which two jaws are added, one fixed at the end of the rule and the other movable, shown in Figure 8.3. Slide calipers can be used for inside or outside measurements, depending on whether the inside or outside jaw faces are used. When in use, the jaws are forced into contact with the part surfaces to be measured, and the location of the movable jaw indicates the dimension of interest. Slide calipers permit more accurate and precise measurements than simple rules. A refinement of the slide caliper is the *vernier*

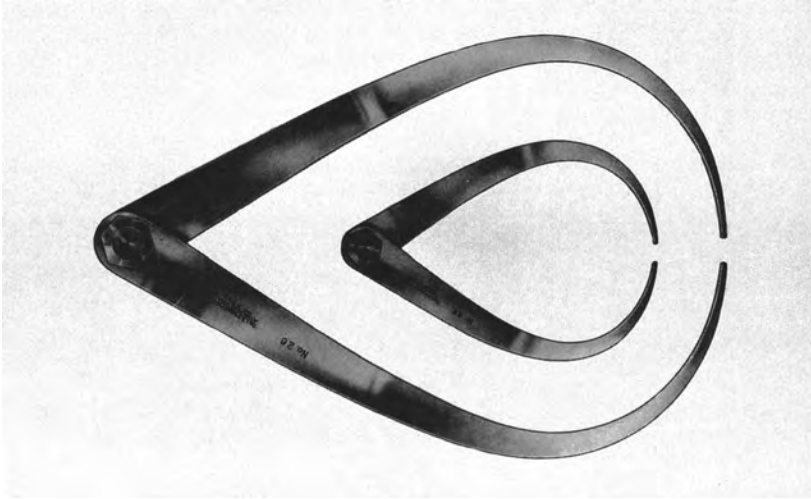


FIGURE 8.2 Two sizes of outside calipers (Courtesy of L.S. Starrett Co.).

caliper, shown in Figure 8.4. In this device, the movable jaw includes a vernier scale, named after P. Vernier (1580–1637), a French mathematician who invented it. The vernier provides graduations of 0.01 mm in the SI (and 0.001 in. in the U.S. customary scale), much more precise than the slide caliper.

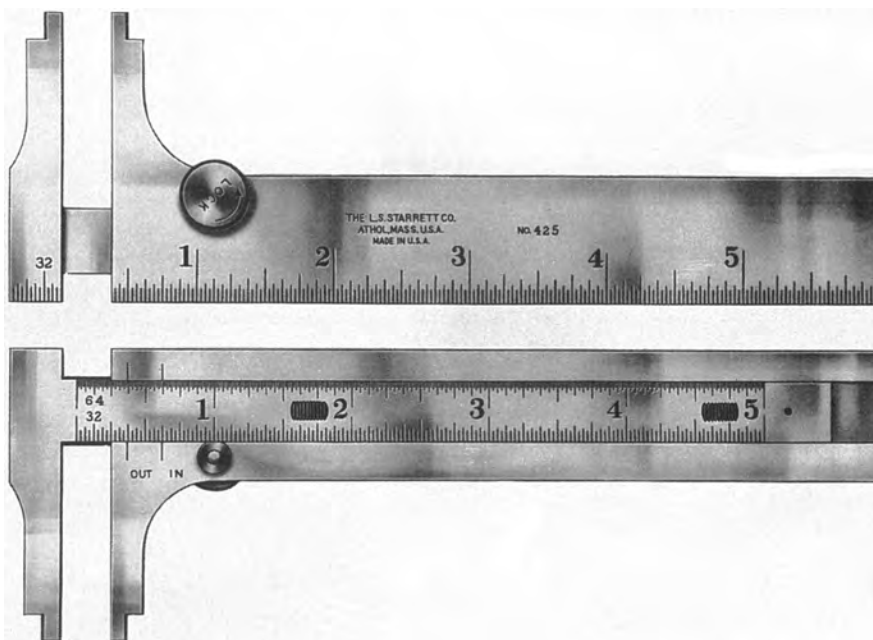


FIGURE 8.3 Slide caliper, opposite sides of instruments shown (Courtesy of L.S. Starrett Co.).

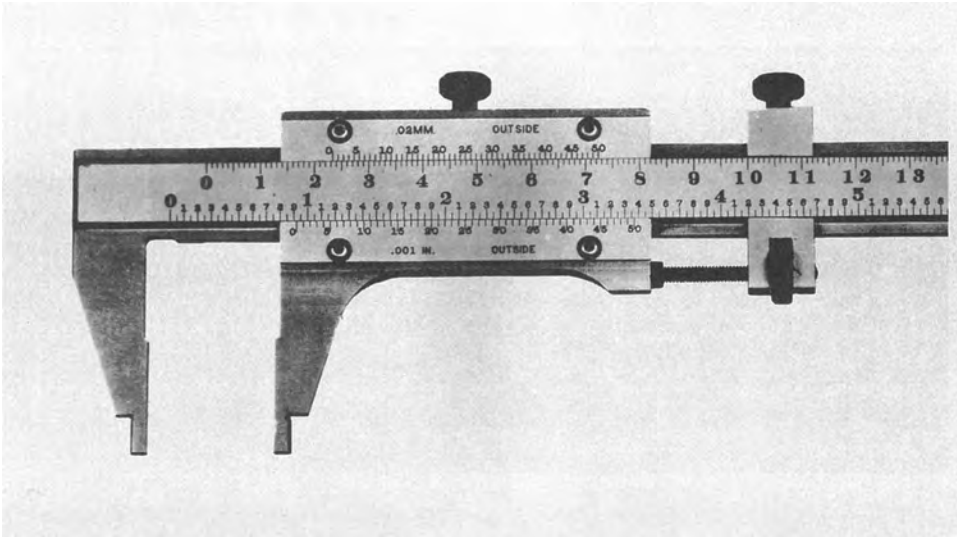


FIGURE 8.4 Vernier caliper (Courtesy of L.S. Starrett Co.).

The *micrometer* is a widely used and very accurate measuring device, the most common form of which consists of a spindle and a C-shaped anvil, as in Figure 8.5. The spindle is moved relative to the fixed anvil by means of an accurate screw thread. On a typical U.S. micrometer, each rotation of the spindle provides 0.025 in.



FIGURE 8.5 External micrometer, standard 1-in. size with digital readout (Courtesy of L.S. Starrett Co.).

of linear travel. Attached to the spindle is a thimble graduated with 25 marks around its circumference, each mark corresponding to 0.001 in. The micrometer sleeve is usually equipped with a vernier, allowing resolutions as close as 0.0001 in. On a micrometer with metric scale, graduations are 0.01 mm. Modern micrometers (and graduated calipers) are available with electronic devices that display a digital readout of the measurement (as in the figure). These instruments are easier to read and eliminate much of the human error associated with reading conventional graduated devices.

The most common micrometer types are (1) *external micrometer*, Figure 8.5, also called an *outside micrometer*, which comes in a variety of standard anvil sizes; (2) *internal micrometer*, or *inside micrometer*, which consists of a head assembly and a set of rods of different lengths to measure various inside dimensions that might be encountered; and (3) *depth micrometer*, similar to an inside micrometer but adapted to measure hole depths.

8.2.3 Comparative Instruments

Comparative instruments are used to make dimensional comparisons between two objects, such as a workpart and a reference surface. They are usually not capable of providing an absolute measurement of the quantity of interest; instead, they measure the magnitude and direction of the deviation between two objects. Instruments in this category include mechanical and electronic gages.

8.2.3.1 Mechanical Gages: Dial Indicators *Mechanical gages* are designed to mechanically magnify the deviation to permit observation. The most common instrument in this category is the *dial indicator* (Figure 8.6), which converts and amplifies the linear

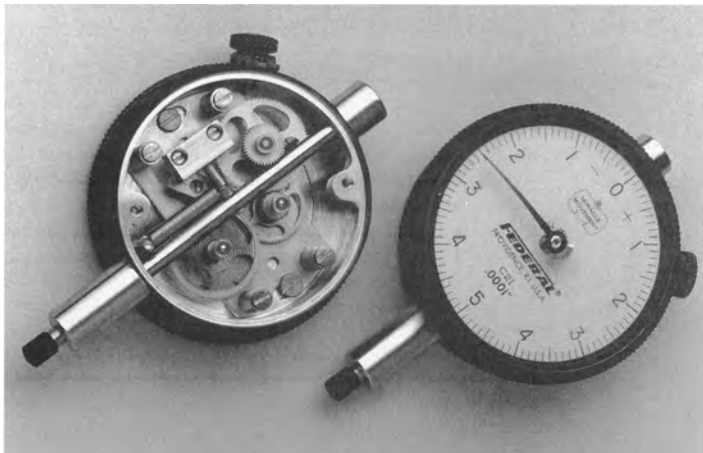


FIGURE 8.6 Dial indicator: top view shows dial and graduated face; bottom view shows rear of instrument with cover plate removed (Courtesy of Federal Products Co., Providence, RI).

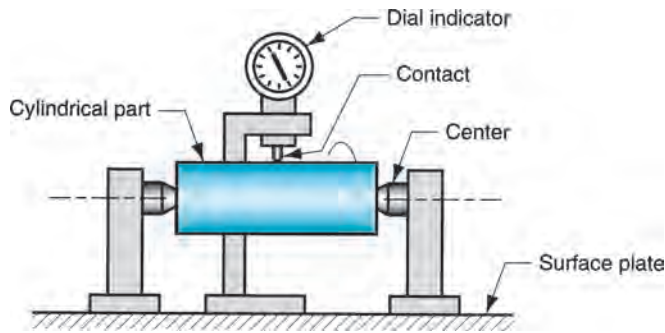


FIGURE 8.7 Dial indicator setup to measure runout; as part is rotated about its center, variations in outside surface relative to center are indicated on the dial.

movement of a contact pointer into rotation of a dial needle. The dial is graduated in small units such as 0.01 mm (or 0.001 in.). Dial indicators are used in many applications to measure straightness, flatness, parallelism, squareness, roundness, and runout. A typical setup for measuring runout is illustrated in Figure 8.7.

8.2.3.2 Electronic Gages Electronic gages are a family of measuring and gaging instruments based on transducers capable of converting a linear displacement into an electrical signal. The electrical signal is then amplified and transformed into a suitable data format such as a digital readout, as in Figure 8.5. Applications of electronic gages have grown rapidly in recent years, driven by advances in microprocessor technology. They are gradually replacing many of the conventional measuring and gaging devices. Advantages of electronic gages include (1) good sensitivity, accuracy, precision, repeatability, and speed of response; (2) ability to sense very small dimensions—down to $0.025\text{ }\mu\text{m}$ ($1\text{ }\mu\text{in.}$); (3) ease of operation; (4) reduced human error; (5) electrical signal that can be displayed in various formats; and (6) capability to be interfaced with computer systems for data processing.

8.2.4 Fixed Gages

A fixed gage is a physical replica of the part dimension to be assessed. There are two basic categories: master gage and limit gage. A *master gage* is fabricated to be a direct replica of the nominal size of the part dimension. It is generally used for setting up a comparative measuring instrument, such as a dial indicator; or for calibrating a measuring device.

A *limit gage* is fabricated to be a reverse replica of the part dimension and is designed to check the dimension at one or more of its tolerance limits. A limit gage often consists of two gages in one piece, the first for checking the lower limit of the tolerance on the part dimension, and the other for checking the upper limit. These gages are popularly known as *GO/NO-GO gages*, because one gage limit allows the part to be inserted, whereas the other limit does not. The *GO limit* is used to check the dimension at its maximum material condition; this is the minimum size for an internal feature such as a hole, and it is the maximum size for an external feature

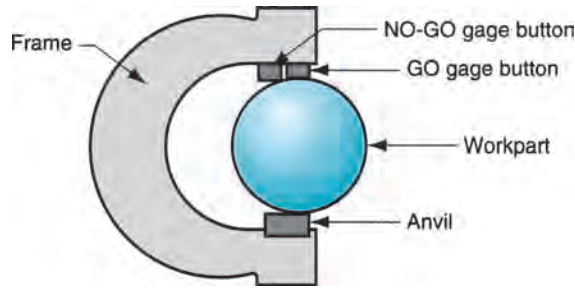


FIGURE 8.8 Snap gage for measuring diameter of a part; difference in height of GO and NO-GO gage buttons is exaggerated.

such as an outside diameter. The *NO-GO limit* is used to inspect the minimum material condition of the dimension in question.

Common limit gages are snap gages and ring gages for checking outside part dimensions, and plug gages for checking inside dimensions. A *snap gage* consists of a C-shaped frame with gaging surfaces located in the jaws of the frame, as in Figure 8.8. It has two gage buttons, the first being the GO gage, and the second being the NO-GO gage. Snap gages are used for checking outside dimensions such as diameter, width, thickness, and similar surfaces.

Ring gages are used for checking cylindrical diameters. For a given application, a pair of gages is usually required, one GO and the other NO-GO. Each gage is a ring whose opening is machined to one of the tolerance limits of the part diameter. For ease of handling, the outside of the ring is knurled. The two gages are distinguished by the presence of a groove around the outside of the NO-GO ring.

The most common limit gage for checking hole diameter is the *plug gage*. The typical gage consists of a handle to which are attached two accurately ground cylindrical pieces (plugs) of hardened steel, as in Figure 8.9. The cylindrical plugs serve as the GO and NO-GO gages. Other gages similar to the plug gage include *taper gages*, consisting of a tapered plug for checking tapered holes; and *thread gages*, in which the plug is threaded for checking internal threads on parts.

Fixed gages are easy to use, and the time required to complete an inspection is almost always less than when a measuring instrument is employed. Fixed gages were a

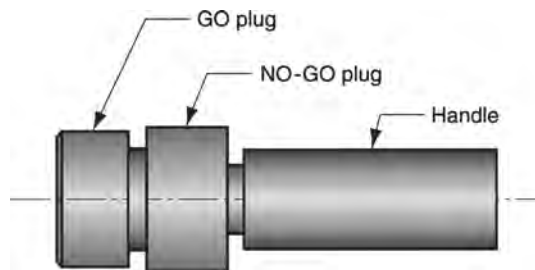


FIGURE 8.9 Plug gage; difference in diameters of GO and NO-GO plugs is exaggerated.

fundamental element in the development of interchangeable parts manufacturing (Historical Note 1.1). They provided the means by which parts could be made to tolerances that were sufficiently close for assembly without filing and fitting. Their disadvantage is that they provide little if any information on the actual part size; they only indicate whether the size is within tolerance. Today, with the availability of high-speed electronic measuring instruments, and with the need for statistical process control of part sizes, use of gages is gradually giving way to instruments that provide actual measurements of the dimension of interest.

8.2.5 Angular Measurements

Angles can be measured using any of several styles of *protractor*. A *simple protractor* consists of a blade that pivots relative to a semicircular head that is graduated in angular units (e.g., degrees, radians). To use, the blade is rotated to a position corresponding to some part angle to be measured, and the angle is read off the angular scale. A *bevel protractor* (Figure 8.10) consists of two straight blades that pivot relative to each other. The pivot assembly has a protractor scale that permits the angle formed by the blades to be read. When equipped with a vernier, the bevel protractor can be read to about 5 min; without a vernier the resolution is only about 1 degree.

High precision in angular measurements can be made using a *sine bar*, illustrated in Figure 8.11. One possible setup consists of a flat steel straight edge (the sine bar), and two precision rolls set a known distance apart on the bar. The straight edge is aligned with the part angle to be measured, and gage blocks or other accurate linear measurements are made to determine height. The procedure is carried out on a surface plate to

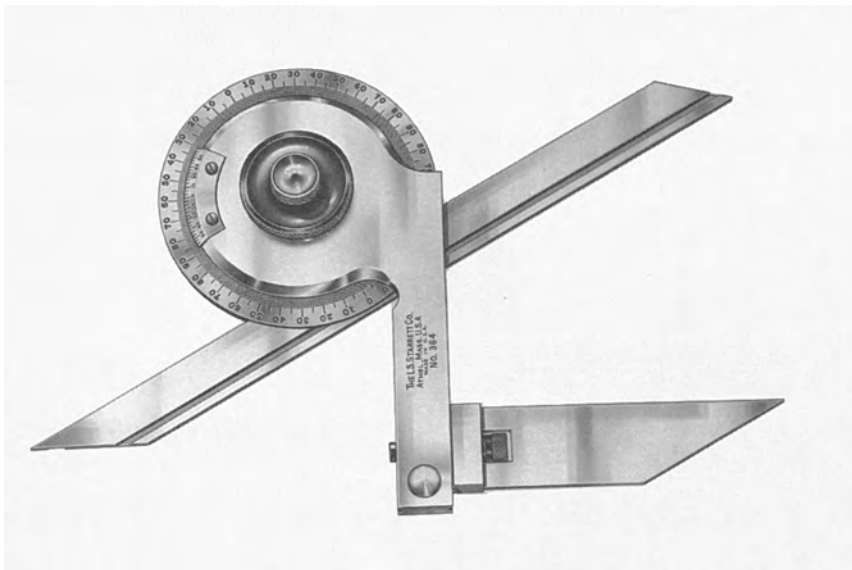


FIGURE 8.10 Bevel protractor with vernier scale (Courtesy of L.S. Starrett Co.).

achieve most accurate results. The height H and the length L of the sine bar between rolls are used to calculate the angle A using

$$\sin A = \frac{H}{L} \quad (8.1)$$

8.3 SURFACES

A surface is what one touches when holding an object, such as a manufactured part. The designer specifies the part dimensions, relating the various surfaces to each other. These *nominal surfaces*, representing the intended surface contour of the part, are defined by lines in the engineering drawing. The nominal surfaces appear as absolutely straight lines, ideal circles, round holes, and other edges and surfaces that are geometrically perfect. The actual surfaces of a manufactured part are determined by the processes used to make it. The variety of processes available in manufacturing result in wide variations in surface characteristics, and it is important for engineers to understand the technology of surfaces.

Surfaces are commercially and technologically important for a number of reasons, different reasons for different applications: (1) aesthetic reasons—surfaces that are smooth and free of scratches and blemishes are more likely to give a favorable impression to the customer, (2) surfaces affect safety, (3) friction and wear depend on surface characteristics, (4) surfaces affect mechanical and physical properties; for example, surface flaws can be points of stress concentration, (5) assembly of parts is affected by their surfaces; for example, the strength of adhesively bonded joints (Section 31.3) is increased when the surfaces are slightly rough, and (6) smooth surfaces make better electrical contacts.

Surface technology is concerned with (1) defining the characteristics of a surface, (2) surface texture, (3) surface integrity, and (4) the relationship between manufacturing processes and the characteristics of the resulting surface. The first three topics are covered in this section; the final topic is presented in Section 8.5.

8.3.1 Characteristics of Surfaces

A microscopic view of a part's surface reveals its irregularities and imperfections. The features of a typical surface are illustrated in the highly magnified cross section of the surface of a metal part in Figure 8.12. Although the discussion here is focused on metallic surfaces, based on the process capability for the particular manufacturing operation, as defined in Section 42.2. The tolerance that should be specified is a function of part size; larger parts require more generous tolerances. The table lists tolerance for moderately sized parts in each processing category.

The manufacturing process determines surface finish and surface integrity. Some processes are capable of producing better surfaces than others. In general, processing cost increases with improvement in surface finish. This is because additional operations and more time are usually required to obtain increasingly better surfaces. Processes noted for providing superior finishes include honing, lapping, polishing, and superfinishing (Chapter 25). Table 8.2 indicates the usual surface roughness that can be expected from various manufacturing processes.

TABLE 8.2 Surface Roughness Values Produced by the Various Manufacturing Processes^a

Process	Typical Finish	Roughness Range ^b	Process	Typical Finish	Roughness Range ^b
<i>Casting</i>			<i>Abrasive</i>		
Die casting	Good	1–2 (30–65)	Grinding	Very good	0.1–2 (5–75)
Investment	Good	1.5–3 (50–100)	Honing	Very good	0.1–1 (4–30)
Sand casting	Poor	12–25 (500–1000)	Lapping	Excellent	0.05–0.5 (2–15)
<i>Metal forming</i>			Polishing	Excellent	0.1–0.5 (5–15)
Cold rolling	Good	1–3 (25–125)	Superfinish	Excellent	0.02–0.3 (1–10)
Sheet metal draw	Good	1–3 (25–125)	<i>Nontraditional</i>		
Cold extrusion	Good	1–4 (30–150)	Chemical milling	Medium	1.5–5 (50–200)
Hot rolling	Poor	12–25 (500–1000)	Electrochemical	Good	0.2–2 (10–100)
<i>Machining</i>			Electric discharge	Medium	1.5–15 (50–500)
Boring	Good	0.5–6 (15–250)	Electron beam	Medium	1.5–15 (50–500)
Drilling	Medium	1.5–6 (60–250)	Laser beam	Medium	1.5–15 (50–500)
Milling	Good	1–6 (30–250)	<i>Thermal</i>		
Reaming	Good	1–3 (30–125)	Arc welding	Poor	5–25 (250–1000)
Shaping and planing	Medium	1.5–12 (60–500)	Flame cutting	Poor	12–25 (500–1000)
Sawing	Poor	3–25 (100–1000)	Plasma arc cutting	Poor	12–25 (500–1000)
Turning	Good	0.5–6 (15–250)			

^aCompiled from American National Standards Institute, Inc. (1978, 1986) and other sources.

^bRoughness range values are given, in μm ($\mu\text{in.}$). Roughness can vary significantly for a given process, depending on process parameters.

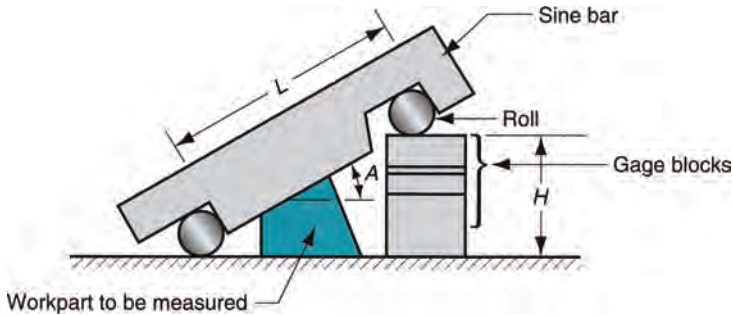


FIGURE 8.11 Setup for using a sine bar.

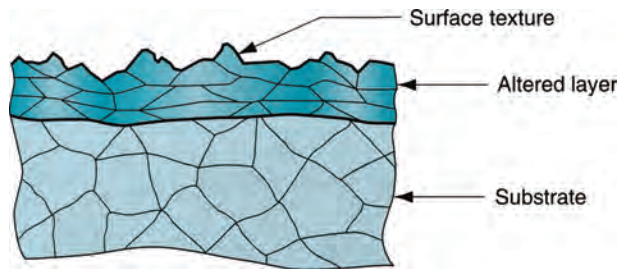


FIGURE 8.12 A magnified cross section of a typical metallic part surface.

REFERENCES

- American National Standards Institute, Inc. *Surface Texture*. ANSI B46.1-1978. New York: American Society of Mechanical Engineers; 1978.
- American National Standards Institute, Inc. *Dimensioning and Tolerancing*. ANSI Y14.5M-1982. New York: American Society of Mechanical Engineers; 1982.
- American National Standards Institute, Inc. *Surface Integrity*. ANSI B211.1-1986. Dearborn, Michigan: Society of Manufacturing Engineers; 1986.
- Bakerjian R, Mitchell P. *Tool and Manufacturing Engineers Handbook*. 4th ed., Vol. VI, *Design for Manufacturability*. Dearborn, Michigan: Society of Manufacturing Engineers; 1992.
- Brown & Sharpe. *Handbook of Metrology*. North Kingston, Rhode Island; 1992.
- Curtis M. *Handbook of Dimensional Measurement*. 4th ed. New York: Industrial Press; 2007.
- Drozda TJ, Wick C. *Tool and Manufacturing Engineers Handbook*. 4th ed., Vol. I, *Machining*. Dearborn, Michigan: Society of Manufacturing Engineers; 1983.
- Farago FT. *Handbook of Dimensional Measurement*. 3rd ed. New York: Industrial Press Inc.; 1994.
- Machining Data Handbook*, 3rd ed., Vol. II. Cincinnati, Ohio: Machinability Data Center; 1980, Ch. 18.
- Mummery, L. *Surface Texture Analysis—The Handbook*. Germany: Hommelwerke GmbH; 1990.
- Oberg E, Jones FD, Horton HL, Ryffel H. *Machinery's Handbook*. 26th ed. New York: Industrial Press; 2000.

- Schaffer GH. "The Many Faces of Surface Texture," Special Report 801. *American Machinist and Automated Manufacturing* 1988;61–68.
- Sheffield Measurement, a Cross & Trecker Company, *Surface Texture and Roundness Measurement Handbook*. Dayton, Ohio; 1991.
- Spitler D, Lantrip J, Nee J, Smith DA. *Fundamentals of Tool Design*. 5th ed. Dearborn, Michigan: Society of Manufacturing Engineers; 2003.
- Starrett S. Company. *Tools and Rules*. Athol, Massachusetts; 1992.
- Wick C, Veilleux RF. *Tool and Manufacturing Engineers Handbook*. 4th ed., Vol. IV, Dearborn, Michigan: Quality Control and Assembly. Society of Manufacturing Engineers; 1987, Section 1.
- Zecchino M. "Why Average Roughness Is Not Enough." *Advanced Materials & Processes* 2003, 25–28.

9

MASS PROPERTIES MEASUREMENT

DAVID TELLET (EDITOR)

- 9.1 Introduction
 - 9.1.1 Scope
 - 9.1.2 Other resources
 - 9.1.3 Terms
- 9.2 Mass and weight
 - 9.2.1 Measurement systems
- 9.3 Measurement methodology
 - 9.3.1 Steps in making a mass properties measurement
 - 9.3.2 Data and frames of reference
 - 9.3.3 Choosing a mass properties instrument
 - 9.3.4 Fixtures
- 9.4 Weight and mass measurement
 - 9.4.1 Types of scales
- 9.5 Center of gravity measurement
 - 9.5.1 Repositioning methods of CG measurement
 - 9.5.2 Multiple-point weighing method
 - 9.5.3 CG by MOI method
 - 9.5.4 Spin balance method
- 9.6 MOI measurement
 - 9.6.1 Hanging wire torsion pendulum
 - 9.6.2 Bifilar suspension
 - 9.6.3 Inverted torsion pendulum
 - 9.6.4 Combined CG and MOI measurement
- 9.7 POI measurement
 - 9.7.1 MOI method of measuring POI
 - 9.7.2 Spin balance method
- 9.8 Measuring large vehicles
 - 9.8.1 General process
 - 9.8.2 Weighing ships
 - 9.8.3 Weighing aircraft

9.9 Sources of uncertainty

9.9.1 Weight uncertainties

9.9.2 Factors affecting weight measurement

9.9.3 Sources of MOI uncertainty

References

9.1 INTRODUCTION

Mass properties—weight, center of gravity, and inertias—are critical performance parameters for any vehicle such as an aircraft, spacecraft, land vehicle, or ship as well as any subsystem or component. In spite of the excellent computer models that are available, the only way to ensure that the mass properties are accurate is to measure those properties. This leads to the questions that every engineer must consider in order to understand or develop a mass properties measurement plan.

- What should be measured?
- When should measurements be made?
- What measurement accuracy is required?
- What measurement equipment is available?

The mass properties of a moving object have a tremendous impact on how it performs during its intended operations and how it behaves during an emergency situation. For example, the weight and center of gravity (CG) of a land vehicle impacts its operability, economy, safety, and even its marketability. For marine vehicles, weight and CG location determines its draft, list, trim, handling, and its inherent ability to remain stable in a seaway. In the case of aircraft, missiles, and rockets, the vehicle's weight, CG, and moment of inertia (MOI) have a profound effect on the stability and controllability of the vehicle and its economy of operation.

Neglecting the mass properties of a vehicle has historically resulted in dire consequences, often with a significant loss of life. Planes have crashed due to instability and controllability problems. Cars have had a greater propensity to overturn during high-speed turns and have been the subject of expensive recalls and even model cancellations due to the neglect of mass properties. Ships have capsized due to insufficient stability when loads have shifted or when heavy seas have been encountered. In fact, entire product development programs have been canceled due to problems with mass properties. Therefore, the impact of mass properties is critical to the proper operation, economy, and safety of a moving object and this is why the accurate measurement of mass properties is an essential element of a successful product development program.

9.1.1 Scope

This article deals mainly with the mass properties measurement of individual objects and vehicles where precise mass properties measurement and management is necessary to meet strict engineering and performance requirements. The techniques, methods, tools,

and approach described here are primarily aimed at the application of mass properties measurement to vehicles and vehicle subsystems, rather than to production line or bulk measurements. Information on these other weight measurement methods and requirements can be found in National Institute of Standards and Technology (2011; <http://www.nist.gov>). The mass properties addressed herein are weight and mass, centers of gravity, moments of inertia, and products of inertia.

9.1.2 Other Resources

There are many organizations that can serve as valuable resources to the field of mass properties measurement. These include numerous governmental and independent professional societies that support the development of standards and practices for the measurement of physical properties. These organizations may offer valuable standards, recommended practices, training, educational resources, and publications. These organizations include

- *The Society of Allied Weight Engineers, Inc. (SAWE, Inc.):* An international organization based in the United States and dedicated to the promotion, practice, education, and innovation in the field of mass properties engineering. The SAWE's primary emphasis is on military and commercial aerospace, shipbuilding, land vehicle, and allied industries. Founded in 1939, the SAWE is a nonprofit organization. More information can be found at <http://www.sawe.org>.
- *National Institute of Standards and Technology (NIST):* A nonregulatory federal agency within the U.S. Department of Commerce dedicated to the promotion of U.S. innovation and industrial competitiveness by advancing measurement science. More information can be found at <http://www.nist.gov>.
- *International Bureau of Weights and Measures, or Bureau Internationale des Poids et Mesures (BIPM):* An international organization established in 1875 to ensure worldwide unification and coordination of measurement and the establishment of standards and scales for the measurement of physical quantities. The BIPM is based in France. More information can be found at <http://www.bipm.org>.
- *National Physical Laboratory:* The national measurement standards laboratory for the United Kingdom that was established in 1900 and is devoted to the development and application of physical measurement techniques. More information can be found at <http://www.npl.co.uk>.

9.1.3 Terms

As with any engineering field, mass properties engineering has many specific terms that are unique or that may not be familiar to all readers. A partial list follows

Bifilar Pendulum: A two wire suspension system to support an object so its moment of inertia can be determined by measuring the period of rotational oscillation when it is displaced through a small angle about a vertical axis.

Center of Gravity (Center of Mass; CG): The point at which the distributed mass of a body can be acted upon by a force without inducing any rotation of that body.

Displacement: (1) The weight of a ship or floating body in terms of the weight of the fluid displaced. Usually stated in terms of metric tons or long tons (2240 lbs).
 (2) The position of an object or CG of an object relative to a datum stated in linear units (i.e., inches or millimeters).

Fixture: A structure that supports and locates an object on the mass properties measurement assembly.

Mass: The measure of the quantity of matter in a body; a body's resistance to change in motion.

Moment of Inertia (MOI): The rotational inertia or resistance to change in direction or speed of rotation about a defined axis.

Longitudinal Center of Gravity (LCG): Normally measured from a datum and along the longitudinal axis of the object or vehicle.

Payload: The object or vehicle that is mounted in the mass properties measurement assembly for measurement. Also called unit under test (UUT) or test object.

Product of Inertia (POI): A measure of a body's dynamic (or coupled) imbalance resulting in a precession when rotating about an axis other than the body's principal axis.

Rangeability: The ratio between load capacity and accuracy for a measurement instrument such as a scale.

Tare: The weight (or other mass property) of a container or fixture in an empty state without any payload.

Transverse Center of Gravity (TCG): Normally measured from a datum and along the transverse axis of the object or vehicle.

Trifilar Pendulum: A suspension system using three wires to support an object to measure its MOI by timing rotational oscillations about a vertical axis.

Vertical Center of Gravity (VCG): Normally measured from a baseline up along the vertical axis of the object or vehicle.

Weight: The force that presses the object down on a scale due to the acceleration of gravity.

9.2 MASS AND WEIGHT

Before attempting to measure the mass properties of an object, it is vital to understand exactly what you will be measuring, and what you need to measure. The first thing that must be understood is the difference between mass and weight.

Mass is related to weight through Newton's second law:

$$W = Mg \quad (9.1)$$

where W is the weight of the object (gravity force); M is the mass of the object; and g is the acceleration of gravity.

Mass is the quantity of matter in an object, whereas weight is the force that presses the object down on a scale due to the acceleration of gravity. The mass of an object is a fixed quantity; its weight varies as a function of the acceleration of gravity. The mass properties

of an object are related to mass, not weight. For instance, mass properties do not change as a space vehicle leaves the attraction of Earth's gravity.

9.2.1 Measurement Systems

The two most prevalent measurement systems used internationally are the metric or SI system and the English or Imperial system. The engineer performing the measurement and the end user of the data must be made clearly aware of which system is being used. This is especially important in international collaboration projects or in any situation where mixed units might be used.

In 1999, the Mars Climate Orbiter crashed as a result of confusion over the system of units. The software program that controlled the thrusters was supplied with thrust data in pound-seconds but interpreted it as if it were Newton-seconds, resulting in an underestimation of the thruster impulse. This is only one of many thousands of errors that have occurred as a result of the conflict between Metric and English units (Boynton, 2001).

Many of these errors are the result of a misunderstanding regarding the difference between mass and weight. If you place an object on a scale in Europe, you will read its mass (generally expressed in kilograms). However, if you place an object on a scale in the United States, you will read a value equal to the force exerted by the acceleration of gravity (generally expressed in pound-force [lbf]). In reality, most scales (and all those used for commerce) are calibrated to read the weight of a given mass in a standard gravitational field, not the local gravitational field. Since you are really trying to use the scale to measure mass, when you weigh yourself on an American bathroom scale, it should read 6.22 slugs rather than 200 lbf.

Traditionally, a dimensionally inconsistent correction factor is used to convert from one set of units to the other. The expression $1 \text{ kg} = 2.205 \text{ lb}$ is not actually valid. It is like comparing apples to oranges. Mass does not equal force. This traditional conversion factor is based on the value of standard gravity.

The various measuring systems are fundamentally mass–length–time systems with force being a defined or derived term. The U.S. systems are fundamentally force–length–time systems with mass being defined or derived. Table 9.1 shows the two most commonly used systems of measurement.

If different names are used for weight and mass, then the problem of distinguishing between the two is minimized. The Metric SI system uses the word Newton for weight and the word kilogram for mass. The Newton is defined as the force required to accelerate a 1 kg mass by 1 m/s^2 . Since the standard value for g is 9.80665 m/s^2 , if an object weighs 9.80665 N, its mass is 1 kg.

In the English system, the aerospace industry has created a unit of mass called the slug. A 1 lb force is required to accelerate a 1 slug mass at 1 ft/s^2 . Since the standard value for g is 32.17405 ft/s^2 , if an object weighs 32.17405 lbf on Earth, then its mass is 1 slug.

TABLE 9.1 Dimensionally Correct Measuring Systems

System	Mass	Length	Time	Weight	g
SI (metric)	Kilogram (kg)	Meter (m)	Second (s)	Newton (N)	9.80665 m/s^2
US (foot)	Slug	Foot (ft)	Second (s)	Pound (lbf)	32.17405 ft/s^2

Unfortunately, not all systems of units adequately differentiate between mass and weight. In the United States, the word pound is commonly used for both mass and weight, resulting in endless confusion and errors in calculating mass properties and dynamic response. Officially, pound refers to mass (National Institute of Standards and Technology, 2011). However, the common usage of the word pound is the value you read on a scale, which is actually pound-force (lbf). If the term pound is used to describe a mass whose measured weight is 1 lb (force), this quantity must be divided by the acceleration of gravity in appropriate units to convert it to proper mass dimensions if it is to be used in mass properties calculations. Similarly, in metric countries the terms kilogram and gram are often used, incorrectly, to describe force as well as mass. To avoid confusion and uncertainty, an analysis of fundamental dimensions will confirm if correct units of measurement are being applied correctly to achieve desired results.

Throughout this article, both SI and English systems are used in the examples and in equations. Although SI is the preferred system, legacy projects in the United States still use English exclusively or some mix of the two systems (National Institute of Standards and Technology, 2011). The important thing is to understand the system being used and consistently use and report the correct units for the mass properties measurement being performed.

9.3 MEASUREMENT METHODOLOGY

9.3.1 Steps in Making a Mass Properties Measurement

As described in Boynton and Wiener (1998), there are nine steps required to measure the mass properties of an object:

1. Define the particular mass properties you need to measure and the required measurement accuracy.
2. Choose the correct type of measuring instrument. This choice will be driven by the availability of existing equipment, accuracy required, cost, and suitability for the measurement environment.
3. Define the coordinate system on the test object to be used as the mass properties reference axes. Any object has an infinite number of values for CG location, MOI, and POI, depending on where the reference axes are assigned. The axes may be related to the geometric centerline of the vehicle, a line of thrust, or may depend on the attachment interface to another stage of the vehicle, object, or structure.
4. Define the position of the test object on the mass properties measuring machine. There are an infinite number of ways a payload can be mounted on a mass properties machine. While the mass properties of the payload are fixed, the measured data will be dependent on its orientation relative to the measurement coordinate system (referring to basic position, not how accurate this position is). For example, a rocket can be mounted on the machine with its nose up or its nose down. The fins can be parallel to the X -axis of the machine or the Y -axis (or for that matter can be oriented at any angle). To avoid confusion, you need to make a drawing or sketch of the position of the payload on the machine so you can interpret the measured data correctly.

5. Determine the dimensional accuracy of the object being measured. This can be the limiting factor on accuracy. For example, you cannot measure the CG of a cylindrical object with an accuracy of 0.005 in. (0.127 mm) if the outer surface of the object has a runout (out of round) of 0.020 in. (0.508 mm).
6. Design the fixture required to mount the test object at a precise location relative to the measuring instrument. This will require a means of determining the location of the measurement axis of the instrument as well as a means of accurately supporting the test object on the instrument. Verify the position of the object on the instrument.
7. Take the mass properties measurement. Several orientations of the test object on the measurement instrument may be needed to meet specified requirements.
8. Reduce data to account for differences between test object configuration and that of the operational vehicle.
9. Report the mass properties data, both raw data and reduced data, including a sketch showing the orientation of the payload and definition of the measurement coordinate system.

9.3.2 Data and Frames of Reference

Whenever a measurement of the physical properties of an object is taken, it is important that the data be properly interpreted for use by someone who was not involved in the measurement. Therefore, the context of the physical measurement must be clearly and concisely communicated to the user of the information. This context is provided by data or frames of reference that are an established form of coordinate system, or set of axes, that serves to locate one specific point on an object relative to another. Typically, the definition of a specific unique point on a three-dimensional object requires location relative to three different planes and a common reference point.

There is no one correct datum or frame of reference for the measurement of physical properties. They are an adopted convention that can vary significantly from one profession to another, one technical discipline to another, one industry to another, and even one country to another.

9.3.2.1 Choosing the Frame of Reference Often there are two different vehicle frames of reference: the body frame, defined by the structure of the vehicle, and the inertial frame, defined by the mass properties of the vehicle. It is important to understand the difference between these two frames of reference (SAWE Members, 2007).

The body frame is a reference system that is related to the physical structure of the vehicle. This frame is easy to define for an ideal vehicle shape, but may be hard to locate on a real vehicle, because of loose manufacturing tolerances and other practical problems.

The inertial frame is a reference system defined by the principal axes of the vehicle. This can be crudely calculated, but it is necessary to make measurements of the real vehicle to accurately determine the location of this inertial frame relative to the body frame. Measurements are made on a mass properties instrument that determines CG location, moments of inertia, and (if necessary) products of inertia. These measurements define the inertial frame relative to the body frame within the tolerance limitations of both the structure and the measuring instruments.

9.3.2.2 Interpreting the Data The following are some general characteristics of mass properties data that must be observed when deciding on the orientation and coordinate system to be used for mass properties measurements.

Moment of inertia can only be positive, so there is never any uncertainty regarding sign. However, you should determine whether this magnitude should be expressed about the geometric centerline of the vehicle or about its CG, about an axis parallel to the geometric centerline or rotated so the data is about the principal axes. In most cases, there will not be a big difference in these three magnitudes. This can lead to confusion, since it will not be immediately obvious that the wrong data is being presented. Mass properties instruments are commercially available which report MOI and CG relative to the instrument centerline and also relative to the payload datum and coordinate system.

Center of gravity coordinates can be positive or negative. You should determine whether your positive axis agrees with the definition of axes used by the recipient of your data. Furthermore, CG distance can be expressed along a coordinate system defined by the geometry of the vehicle or along the principal axes. It is recommended that you provide a sketch that clearly shows the axes and their algebraic signs.

Product of inertia can also be positive or negative. Since this quantity is derived by multiplying the incremental masses by two different distances, the POI sign is even more prone to error than the sign of the CG data. What usually happens is not that the sign is wrong, but that the mass properties engineer and the recipient of his data are using different coordinate systems.

Moment of inertia is expressed about an axis. CG coordinates can be expressed as a distance along an axis or as an unbalance moment about an axis (CG along X corresponds to the CG unbalance moment about Y or Z). POI is relative to two axes (or it can be a tilt angle in a plane defined by two axes).

Six types of information are required to establish a mass properties reference system:

1. The location of the reference axes origin
2. The mathematical symbols or names used to define the reference axes
3. The zero point along each axis
4. The direction of positive values along each axis
5. The positive direction for rotation about each axis
6. A zero rotation angle reference about each axis

9.3.2.3 Standard Coordinate Systems

Aircraft Coordinate System The standard coordinate system defined by the geometric shape of an aircraft is a right-hand rule system as depicted in Figure 9.1. The X , or roll axis, is a line from the nose (or more commonly from a point forward of the nose) of the aircraft to the tail. The Y , or pitch axis, is a line from one wing tip to the other, and the Z , or yaw axis, is a vertical line through the CG of the aircraft. This standard convention is designated as A by Society of Allied Weight Engineers (2006). Measurements are positive when going toward the tail from the origin on the X -axis, to the right (when looking forward) on the Y -axis, and up on the Z -axis.

Spacecraft Coordinate System The geometric coordinate system for satellites, reentry vehicles, or any other space flight object that orbits the Earth is a right-hand rule system

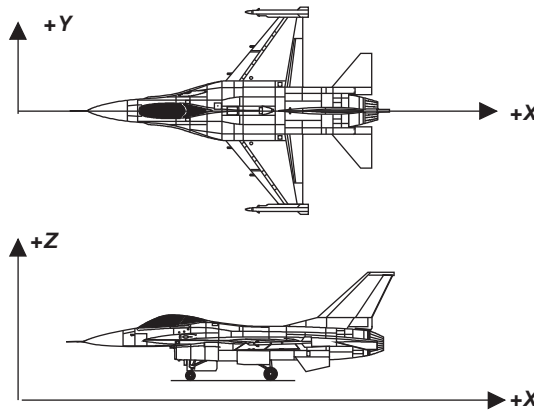


FIGURE 9.1 SAWE standard A, aircraft coordinate system.

as shown in Figure 9.2. In this standard convention commonly referred to as *S*, the positive *X*-axis points in the direction of motion during orbital flight, the *Z*-axis points toward Earth. The *Y*-axis runs transversely, perpendicular to the *Z*-axis. The origin of the *S* coordinate system can coincide with the nominal CG of the vehicle or may be located at a well defined hard point, a specific point on the vehicle which can be accurately measured and identified. If the vehicle has thrusters, their centerlines usually pass through the nominal CG and the intersection of thruster centerlines is the origin.

From the origin, points aft on the *X*-axis are considered positive, points to the right, or starboard, of the origin on the *Y*-axis are typically positive, and points down the *Z*-axis are positive.

Ship Coordinate System The forward standard coordinate system is a left hand rule system for marine vehicles as shown in Figure 9.3. This coordinate system is used with framing conventions originating at the forward perpendicular (FP) and ending at the stern.

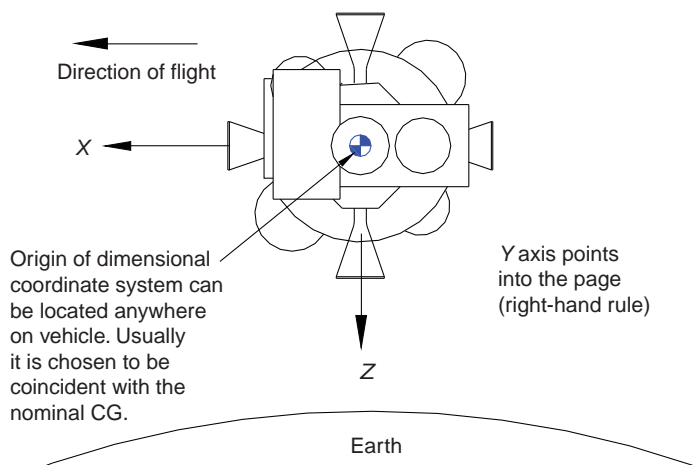


FIGURE 9.2 SAWE standard coordinate system for spacecraft (*S*).

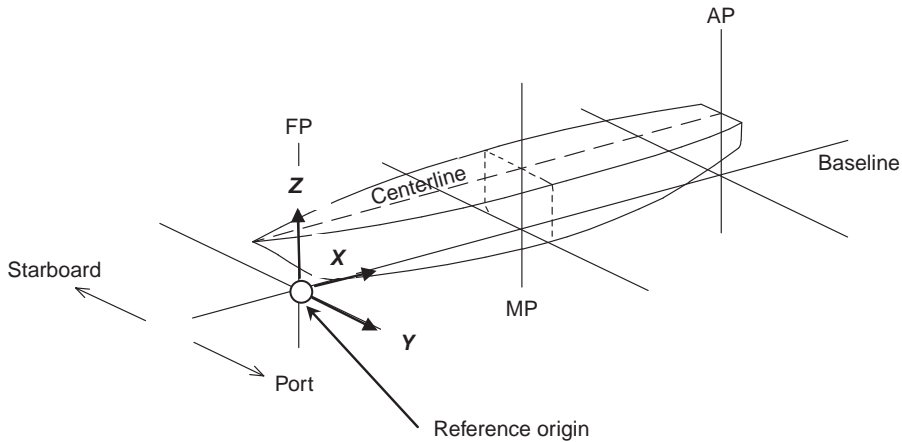


FIGURE 9.3 Standard ship coordinate system.

The X-axis is oriented at the baseline along the centerline of the ship. Longitudinal dimensions are measured along or parallel to this axis with the origin at the forward perpendicular. Locations aft of the origin are positive. The Y-axis runs transversely port and starboard. Transverse dimensions are measured along or parallel to this axis with the origin on centerline. Locations to port are positive. The Z-axis runs vertically and dimensions are measured along or parallel to this axis. Locations above the baseline are positive.

The weight moment of inertia of a marine vehicle is calculated relative to a reference origin about the longitudinal X-axis for roll, transverse Y-axis for pitch, and vertical Z-axis for yaw. The reference origin is located at the marine vehicle's center of gravity. A bow up rotation is positive for pitch, a bow to port rotation is positive for yaw, and a port side up rotation is positive for roll (Society of Allied Weight Engineers, 1996).

Other Coordinate Systems The reference scheme for wheeled vehicles is generally a right-hand rule datum as shown in Figure 9.4:

- Longitudinal station (X), forward of the datum is negative and aft of the datum is positive.
- Lateral buttline (Y), right-hand side of vehicle centerline is positive, looking forward.
- Vertical waterline (Z), up above water line 0.0 is positive.

The location of the datum varies between manufacturers and design firms. The reference system shown also may change for ground vehicle dynamic analysis.

9.3.3 Choosing a Mass Properties Instrument

A wide variety of mass properties measuring instruments is available. The choice of which one to use, in part, depends on what properties you want to measure, the accuracy required, the degree of automation required, and budgetary restrictions. If you are measuring center of gravity, you need to determine whether you need to measure along a single axis, or along more than one axis (some CG instruments are only capable of a single axis measurement). In addition, you need to choose the size of the instrument. This is usually

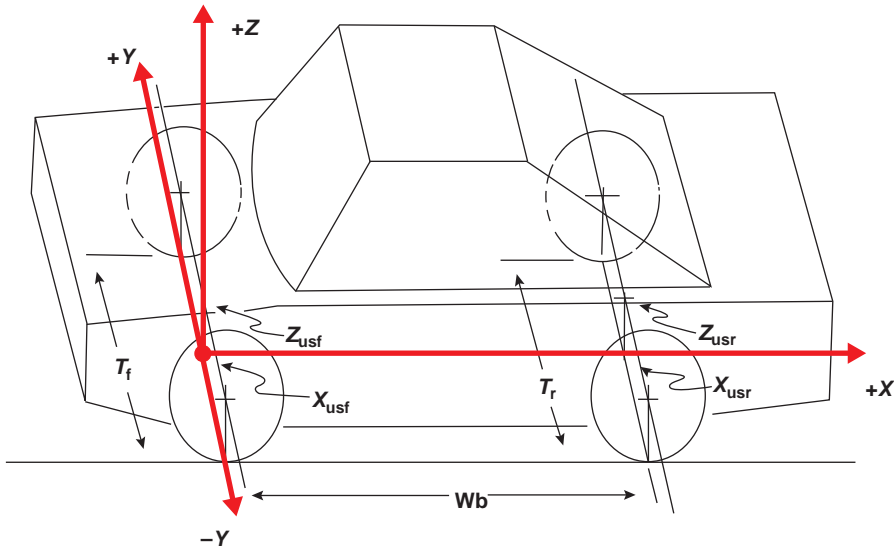


FIGURE 9.4 Standard wheeled vehicle coordinate system.

governed by the weight or physical size of the largest object you need to measure (Society of Allied Weight Engineers, 2006).

Since CG sensitivity decreases as the size of the instrument gets larger, the selection of an instrument may involve a trade-off between cost and current and future capacity needs. This can result in the selection of too large an instrument, resulting in limited accuracy for the present requirements.

9.3.3.1 Properties to Measure There are instruments available that measure only one mass property, and many instruments that can measure two or more mass properties. Using a combined function instrument often eliminates the effort, cost, and risk involved in moving the test item to another instrument or fixture.

The requirements for what properties to measure will help define the instruments needed for the measurement:

Need to Measure Both CG and MOI: In order to measure MOI, you will need an instrument with a gas or oil-bearing rotary table.

Need to Measure Both CG and Weight: Multiple-point weighing instruments have this capability. These instruments are available with load-cell technology or with the more accurate force-cell technology.

Need to Measure Both CG and POI: If you have a requirement to measure dynamic imbalance (POI) as well as CG, then a spin balance machine would be a good solution, particularly if your goal is to balance the test object for minimum POI and CG offset about a particular axis. Some spin balance machines are available with a separate static CG feature. This improves CG measurement accuracy over what would normally be available with a spin balance machine.

Need Maximum CG Accuracy: Instruments with a gas-bearing rotary table and force restoration moment measuring technology are the most accurate.

Need an Explosion Proof Design: Most types of mass properties measurement equipment are available with this option.

Need to Measure CG of Objects Weighing More than 25,000 lbf: Three-point reaction force machines have been made to measure objects as heavy as the space shuttle. Gas-bearing rotary table machines are currently limited to about 25,000 lb (11,340 kg) because of practical problems and cost involved in building gas bearings larger than this.

9.3.3.2 Instruments and Methods A summary of instruments and their principles and characteristics is shown in Table 9.2.

9.3.4 Fixtures

The major source of error in most mass properties measurements is the inability to accurately define the position of the object being measured relative to the measurement axis of the instrument. This is usually accomplished using a precision fixture that supports and locates the object. As described in Society of Allied Weight Engineers (2006), mass properties fixtures must perform three basic functions

1. The fixture must support the payload in a repeatable manner relative to the mass properties instrument.
2. The fixture must provide a means to precisely relate the payload coordinate system to the mass properties instrument coordinate system.
3. The fixture must secure the payload to the mass properties instrument rigidly to minimize all modes of deflection during measurement.

9.3.4.1 Fixture Interfaces There are four basic types of object-to-fixture interfaces

1. *Attachment Point Interface* where the fixture emulates the actual interface between the object and a mating part. This is suitable for rocket/missile stages.
2. *Hard Point Interface* where a system of hard points or rings built into the object being measured is used for inspection, alignment and assembly reference.
3. *Adjustable Interface* where the object has no well-defined hard points and a sophisticated (and time-consuming) method must be used to determine the position of the axes. The fixture is then adjusted to move the object so its axes are coincident with the machine axes.
4. *Calculated Interface* where the object is placed in approximately the correct position, a measurement is made with the object in this position, electronic probes or other means sense the position of the object relative to the machine, and the data are then corrected mathematically so it is expressed relative to the object axes.

Cylinders and other objects having an axis of revolution, such as cones, ellipsoids, and stepped cylinders, can usually be supported on a V-block when measured horizontally, as shown in Figure 9.5. Some objects that are not cylindrical can be mounted in rings so they can be treated as cylinders.

More information on fixtures and fixture design guidelines may be found in Society of Allied Weight Engineers (2006).

TABLE 9.2 Summary of Instruments' Principles and Characteristics

Measurement System Function	Typical Weight Capacities	Operating Principle and Features	Characteristics (R = Range/Sensitivity)
Mass only	1 kg to hundreds of tons	Single purpose. Calibrated in mass or weight units	Wide range of capacities, styles, accuracy and prices. Spring scales: low accuracy and cost Strain gage type: $R = 1/2000$ – $1/10,000$, medium cost, capacities 500 g–50,000 kg Force restoration: up to $R = 1/12,000,000$, highest cost, capacities to 6,000 kg
CG (moment) only	120–500 g	Single sensitive axis, force restoration transducer	High sensitivity for small to medium parts. Requires separate weight measurement to determine CG offset distance
CG (moment) only	1–15 kg	Manually operated rotary table LVDT transducer	Rugged. Not well suited for tall parts. Requires separate weight measurement to determine CG offset distance
Mass and CG	5–10,000 kg	Multi-point weight and CG table	Medium to high capacity, modest accuracy, fast operation. Measures both weight and two axes CG location simultaneously. Well suited to high volume measurements
MOI only	100 g–120 kg	Torsion pendulum with low tare collets for mounting payloads	Manually operated small instruments have very low tare and high accuracy (to 0.01%); Measures MOI about instrument centerline only. Tolerates small CG offset moment
MOI only	250–6,000 kg	Gas-bearing table with torsion pendulum	Large manually operated instruments have good accuracy (to 0.1%) and relatively low cost. Can report only MOI about instrument centerline. Tolerates moderate CG offset moment

(continued)

TABLE 9.2 (Continued)

Measurement System Function	Typical Weight Capacities	Operating Principle and Features	Characteristics (<i>R</i> = Range/Sensitivity)
CG and MOI	5–10,000 kg	Gas-bearing rotary table force restoration transducer for CG; torsion pendulum for MOI	Wide range of capacities available High accuracy. Typical <i>R</i> for CG = 1/30,000 One instrument covers wide range of payloads. Requires separate weight measurement to determine CG offset distance
CG, MOI, POI, static and dynamic balancing	5–6,000 kg	Gas-bearing rotary table. Spin balance 2-plane force measurement for POI. Dynamic CG torsion pendulum for MOI. Force restoration transducer for static CG	Spin speeds from 20 to 800 RPM for POI and dynamic CG measurement. Static and dynamic balancing. Static CG and MOI comparable to above instrument Requires separate weight measurement to determine CG offset distance

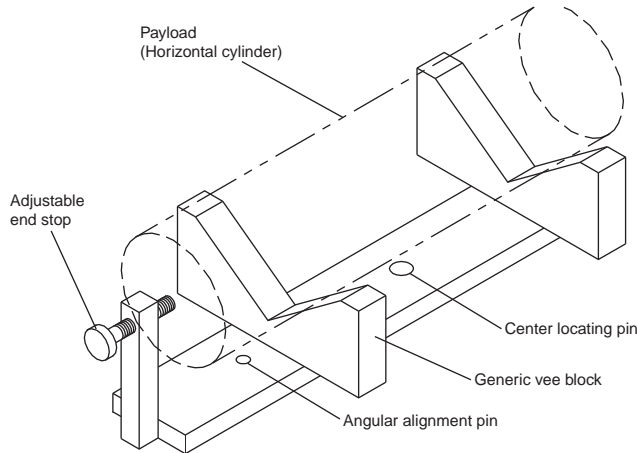


FIGURE 9.5 Generic V-block fixture.

9.3.4.2 Defining the Instrument Axes High accuracy mass properties machines have a mounting table that rotates, making it easy to determine the measurement axes with great precision. The measurement axis is simply the center of rotation of the object-mounting table. Often the object can be dial indicated to align the object with this axis.

CG instruments that use the two-, three-, or four-point weighing method do not have a rotating mounting plate and, therefore, have no well-defined measurement axis. This constitutes a major source of error and is one of the reasons why the three-point reaction force method of CG measurement is less accurate than the rotating table method.

9.3.4.3 Work Reversal Method Whether a fixture is generic or custom designed for a unique payload, its relationship to the instrument coordinate system must be accurately established. If the fixture has been properly designed, its location will be highly repeatable on the mass properties instrument. It is often more cost effective to let the instrument determine the location of the fixture datum relative to the instrument than it is to attempt manufacturing to extremely close tolerances. One of the simplest and most effective means to do this is called work reversal.

Figure 9.6 shows how work reversal is used to determine the true CG offset of a cylinder from its centerline, as well as the V-block centerline offset from the instrument centerline. It is important to note that the part used for this test must be cylindrical and well within the allowable CG tolerance. The CG location of the payload, or a mass model, is measured. The part is then rolled 180° in the fixture and remeasured. Half the difference between the two measurements is the CG offset from the cylinder centerline, while half the sum of the two measurements is the fixture centerline offset from the instrument zero.

The same method can be used to determine the end stop location using a mass model with a known length. The average CG location, as measured from both ends, must be located at a distance equal to half the length of the part.

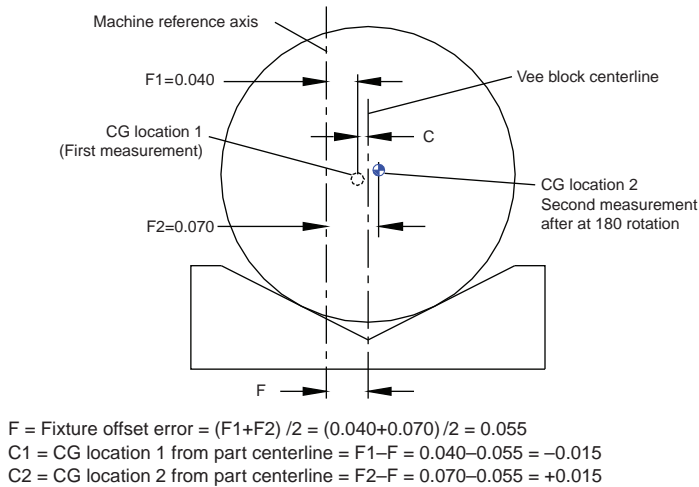


FIGURE 9.6 Finding V-block offset by work reversal.

9.4 WEIGHT AND MASS MEASUREMENT

Weight and mass measurement can be as simple as stepping on a bathroom scale or as complicated as using multiple load cells to weigh an aircraft. In most cases, the method and instruments must be chosen carefully to achieve the needed accuracy. The sources of error or uncertainty must also be understood.

9.4.1 Types of Scales

Although spring and balance scales are still used, nearly all precision scales currently in use have electronic digital displays and computer interface capability. The most common scales still use strain gage load cells with typical accuracy of one part in 2000 (0.05%). These scales lend themselves to computer interfacing at relatively low cost. Pulsed DC power supplies and linearization circuits have allowed accuracies to one part in 5000 (0.02%) at slightly higher cost (Society of Allied Weight Engineers, 2006).

The next level of price with improved rangeability comes with new ceramic capacitive strain gages. These scales can be made with accuracies up to 0.002%. Transducer stiffness is comparable with strain gage beam cells. These transducers have one drawback: they can be damaged if a hard object is dropped on the scale, since the ceramic spring is brittle and cannot withstand shock. Some newer scales using this technology incorporate spring shock dampers to eliminate this problem.

The biggest innovation has been the application of force restoration technology to weight measurement. This technology has been in use since the late 1950s in both electronic and pneumatic process control transducers. The newest generation of electronic force rebalance transducers can achieve accuracies on the order of one part in 10,000,000 (0.00001%) in laboratory balances.

In the more common bench scale ranges up to 25 lb (11.34 kg), the accuracy can approach 0.0001%. In the larger sizes, 75–13,000 lb (34–5,897 kg), accuracies are typically 0.004–0.002%. At this time, the load range available is from micrograms to 13,000 lb (5,897 kg).

These scales are highly programmable to accommodate many weighing conditions: stability (i.e., animal weighing), parts counting on weight basis, selectable units of measurement, and so on. They are fully compatible with computer interfacing. The major disadvantages are price, and slow response time. The slow response time is due to the use of a closed loop rebalance circuit. The slow response time also limits them from being used for dynamic measurements.

9.4.1.1 Force Restoration Principle In a scale that is designed with the force restoration principle, when a load is applied, the transducer deflects. A current driven restoring force is applied through a closed loop control system until the unloaded geometry is restored. The applied current is then related to the applied force. Since the loaded geometry after the restoring force is applied is the same as the unloaded geometry, the transducer is inherently linear like the time honored balance beam scale. This is unlike the strain gage load cell that relies on the deformation of the sensitive spring element to generate an output. High accuracy mass properties measuring instruments for static CG and moment measurement use this force restoration technology.

9.4.1.2 Corner Loading Error When a scale manufacturer quotes the accuracy of his scale, usually he is referring to the accuracy when an object is placed on the platform so that its CG is in the center of the scale. If you place the object off-center, then a moment will be created which tends to tip the platform of the scale. On many scales this will introduce an error. Depending on the internal mechanism in the scale, this error can be as large as 0.5%. High quality scales use parallel beam flexing and other compensating mechanisms, so that this effect can be as small as 0.001%. You can test the scale you are using by first placing a test weight in the center of the scale and measuring its weight. Then you move the weight to each of the corners of the scale and remeasure the weight. You may be surprised at how large the change is. This type of error is the major limiting factor when using two or more scales to determine CG. Placing an object on two scales simultaneously introduces side forces on the scales. Since the CG position is determined by small differences in weight measurement, huge CG errors can result from this side load (corner loading) effect.

9.5 CENTER OF GRAVITY MEASUREMENT

9.5.1 Repositioning Methods of CG Measurement

A basic method of CG measurement is the free-pivot method where the test object is balanced on a pivot and allowed to tilt. The test object is moved relative to the pivot of the instrument until a balance is obtained. Some means is then required to measure the final position of the object. This method of CG measurement is inexpensive, very time-consuming, and generally the least accurate of all methods.

9.5.1.1 Broom Handle Method The simplest method of measuring CG is to balance the object on a round rod. This method only works if the object has a low profile with a rigid surface that will not be indented if its entire weight is supported on a narrow rod, and the surface of the object is flat and smooth.

Since the total CG of the test object lies above the pivot axis, the object can tip in either direction when the pivot point is near its CG. This deadband results in an error in

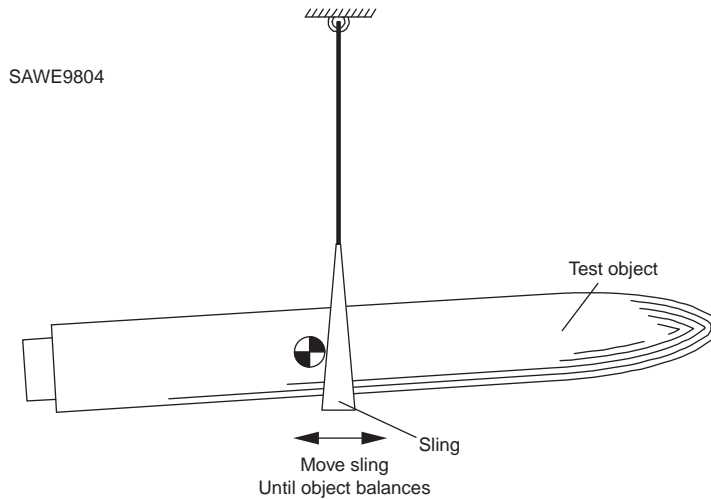


FIGURE 9.7 Hanging pivot method of CG measurement.

CG location. The magnitude of this error is proportional to the amount of angular tilt of the test object, which is permitted. Reducing this tilt will decrease the magnitude of this error. However, the small amount of travel makes the instrument extremely tedious to operate; there is no advance warning that the balance point is being reached and extremely fine adjustments of the test part position must be added to prevent overshooting this very narrow point.

9.5.1.2 Hanging Pivot Method If an object is suspended from a knife-edge, the CG of the object will lie below the center of pivot of the knife-edge. In this method, the position of the test object is moved relative to this pivot point until a balance condition is achieved (see Figure 9.7).

Hanging systems such as this one where the test object CG is always below the pivot point have the disadvantage that their sensitivity is low. Shifting the lateral CG of the test object results in only a small change in the level condition of the instrument. The amount of tilt, which results from a CG offset, is a function of the CG height difference between the object and the knife-edge. Reducing this distance increases sensitivity. However, it also increases measurement time, since the object rocks back and forth at a slower rate when sensitivity is increased. Each time the item is moved, the operator must wait for the system to settle again. In this respect, this method is like using the old beam balance scales.

The sensitivity of the instrument is directly dependent on the accuracy of the means used to detect the level of the instrument. Even the smallest error in the position of the bubble level relative to the structure of the instrument results in a large error in measured CG location.

Since there is no readout to indicate how far to move the object, the process is strictly trial and error. The initial cost of this method is the lowest of any type of CG measuring technique, but often the labor required to make a measurement more than offsets the cost, causing this method to be in fact, the most expensive of all techniques.

Once you have obtained an accurate balance, you must relate the final position of the object to the pivot axis of the knife-edge using something such as a transit and a coordinate measuring machine (Boynton and Wiener, 1998).

9.5.2 Multiple-Point Weighing Method

The multiple-point weighing method can be used for the combined measurement of weight and CG. In this method, a test platform is supported by three (or more) load cells as shown in Figure 9.8. The CG location is calculated from the relationship between force measurements at these points. In the past, the accuracy of this method has been limited by the dynamic range of load cells, so that these instruments were not suitable for projectile and missile measurements. The introduction of force rebalance technology to CG measurement has reduced force measurement uncertainty by a factor of 30. When this technology is applied to the multiple-point weighing method, accuracy improvement is great enough so that this method now becomes acceptable for many applications (Society of Allied Weight Engineers, 2006). Weight is the sum of the transducer readings

$$W = A + B + C \quad (9.2)$$

where A , B , and C are force readings on the three force transducers.

To determine CG, one takes moments about A where X and Y are the CG measurement coordinates. If all the transducers outputs are set to zero when fixturing is in place, Equations (9.3)–(9.7) are used to determine the CG location of the test part (see Figure 9.9). In practice, tare readings are subtracted from the part measurements and the values represent the net A , B , and C forces required to support the part weight and CG offset moment.

$$\sum M_x = (B + C)L - WX = 0 \quad (9.3)$$

$$\sum M_y = \frac{CD}{2} - \frac{BD}{2} - WY = 0 \quad (9.4)$$

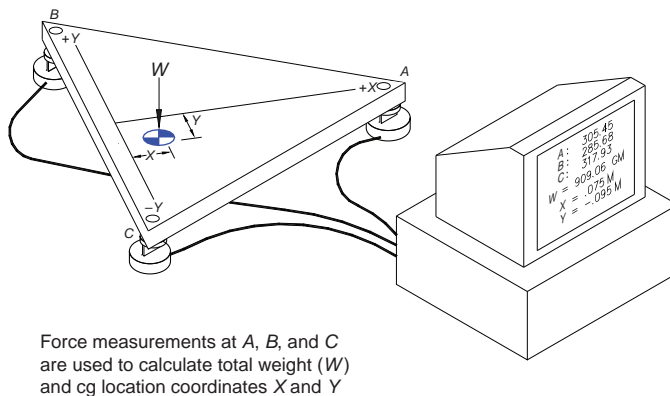


FIGURE 9.8 Multipoint weight and center of gravity principle.

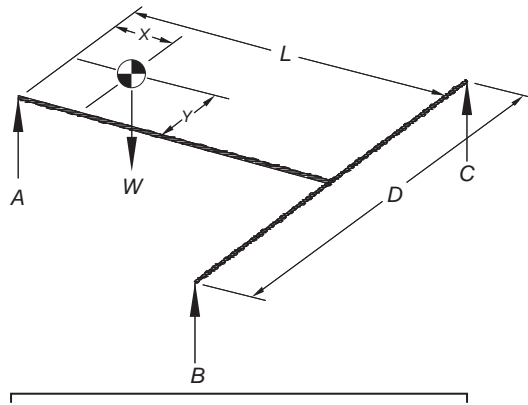


FIGURE 9.9 Free-body diagram of the multipoint weight and center of gravity principle.

$$= \frac{D}{2}(C - B) - WY \quad (9.5)$$

$$X = \frac{(B + C)L}{W} \quad (9.6)$$

$$Y = \frac{(C - B)D}{2W} \quad (9.7)$$

Each force transducer supports an equal load when the payload CG is centered. However, with this type of instrument, each force transducer should be capable of supporting the full weight of the payload to allow for CG offset. Since sensitivity and accuracy of force transducers are inversely related to their capacity, sensitivity of this weight–CG configuration is quite limited. An alternative configuration, the piggyback design, shown in Figure 9.10, uses one large high sensitivity weight platform to support the payload and two high sensitivity force transducers to measure CG offset moments. The payload CG is

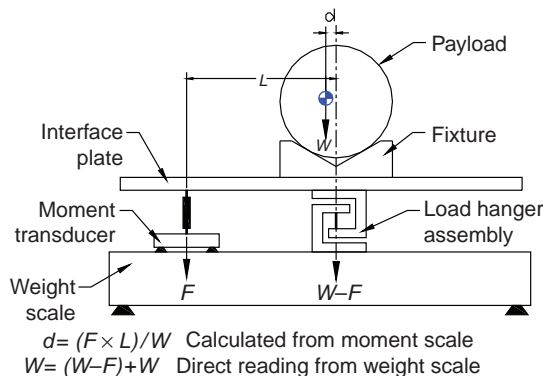


FIGURE 9.10 Weight CG “piggyback” arrangement.

supported over a flexible hanger support, which transfers the load to the weight platform. If the CG is not directly over the support, the unbalance moment is supported by the sensitive moment transducers, which are located at a known distance from the support. This configuration combines high CG sensitivity with high load capacity. It is best suited for a dedicated product in a high production environment.

The disadvantage of these methods is that leveling errors cannot be readily detected or corrected, and the datum is not as well defined as on a rotary table.

9.5.3 CG by MOI Method

In the MOI method for finding an object's CG, the test object is mounted on an inverted torsion pendulum (moment of inertia instrument, see Figure 9.13), and successive moment of inertia measurements made for at least three positions of the test object. The CG location can then be calculated from the small change in the MOI, which results from moving the principal axis of the object.

Center of gravity is determined on a torsion pendulum by making use of the parallel axis theorem. If the moment of inertia of the object about axis *A–A* through its CG is I_A , then the moment of inertia through axis *B–B* is

$$I_B = I_A + d^2M \quad (9.8)$$

where M is the mass of the object and d is the distance between axis *A–A* and axis *B–B*. Note that the minimum measured moment of inertia of an object occurs when the axis of measurement coincides with the CG of the object.

In practice, the CG (along one axis only) can be determined by mounting the object in a V-block fixture on an MOI instrument and taking several MOI measurements with different object positions. The position that results in the smallest measured MOI is one where the CG is coincident with the axis of measurement. This can be a tedious procedure and is prone to position measurement errors. A better method is to measure MOI at three known positions and calculate CG from this data using methods described in Boynton (1977) and Boynton and Bell (1981).

If the displacements between the three measurement positions are made small, then the sensitivity of this method is abysmal. Accuracy of better than 0.1 in. (2.54 mm) is difficult to obtain. If the displacements are made large relative to the radius of gyration of the test object, then the accuracy improves from a theoretical standpoint. However, torsion pendulums do not operate successfully with large CG offsets, due to the gravity pendulum error, so that the increased measurement error partially offsets the gain in sensitivity, and the accuracy of measurement is still worse than other methods.

9.5.4 Spin Balance Method

In this method, the test object is rotated and force transducers sense the reactions on the bearings that support the part during rotation (see Figure 9.11). These forces are due to both gravity and centrifugal force (the higher the spin speed, the less significant the gravity force is). The CG location of the part may then be separated from the dynamic unbalance of the part using calculations that involve the magnitude of the bearing forces and their phase relationship. Spin balance machines rotate the test

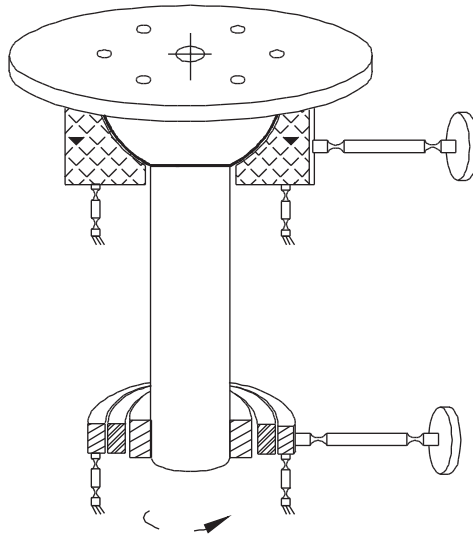


FIGURE 9.11 The spin balance method of CG measurement is expensive and has limited accuracy.

item at speeds ranging from 20 to 10,000 RPM, and measure the reaction forces acting against the bearings in the machine due to dynamic unbalance (a combination of CG offset and product of inertia).

At first analysis, it might appear that it does not make any difference whether CG is measured in a static or a dynamic mode. However, there are a number of considerations that make the spin method of CG measurement unsatisfactory. If one needs to measure the longitudinal CG of a long test object (e.g., 10 m long rocket), then static measurement is the only practical way. Spinning such a test object would require tremendous power and generate high winds in the test laboratory. If you are measuring the CG of a partially filled fuel tank, then the fuel will ride up the sides of the tank if you spin it, because of the centrifugal force. This results in an erroneous CG measurement. If the test object has extended solar panels, then centrifugal forces may damage or deflect them. For objects with a large CG offset, the force limits of the transducers will be exceeded if you spin the part. If a part has a very large product of inertia unbalance, CG measurement will be more accurate in the static mode, since the small CG forces do not have to be separated from the larger POI forces. Finally, and perhaps the most important consideration, when an irregularly shaped object is spun, aerodynamic forces cause large variations in the measured CG offset, severely limiting measurement accuracy.

9.6 MOI MEASUREMENT

The moment of inertia is the rotational inertia of an object about a defined axis. MOI for a rotational body is analogous to mass for a body in linear motion. The moment of inertia can be measured by hanging an object from a wire, twisting it to start it oscillating, and then timing the period of oscillation. If more accuracy is desired, an inverted torsion pendulum instrument is required.

9.6.1 Hanging Wire Torsion Pendulum

Although hanging wire pendulums are usually not accurate enough for satellite and missile measurements, they are useful for measuring the MOI of large objects such as an aircraft.

This method consists of hanging an object from a wire, twisting it to start it oscillating, and then timing the period of oscillation. Although it might appear to be a simple device, the structure required to support the upper end of the wire can be very expensive, and some accurate means is required to time the period of oscillation. One problem with the hanging wire method is that the object swings from side to side and rocks up and down rather than rotating smoothly about an axis, making it difficult to acquire accurate time period data.

It is essential that the center of gravity of the object be aligned horizontally with the wire. Otherwise, the moment the object is released, there will be a couple generated and the motion of the pendulum will be sideways as well as torsional. In addition, the payload will tip to bring the CG directly under the wire.

A single hanging (steel) wire has a torsional stiffness (k) that is proportional to the fourth power of the diameter

$$k = \frac{1,178,000d^4}{L} \text{ in.-lb/radian} \quad (9.9)$$

where L is the length of wire and d is the diameter (in.) of wire.

The equation of motion for this pendulum is

$$I = Ct^2 = \frac{kt^2}{4\pi^2} \quad (9.10)$$

where t is the time for one complete period of oscillation and C is a calibration constant.

However, the actual period of oscillation will not follow this formula because the wire stretches under load, and there will be both a swinging pendulum effect and a rocking pendulum. Typical measurement uncertainty with this method is about 3%.

9.6.2 Bifilar Suspension

A frequently used method of determining the inertia of an object about one axis is the bifilar pendulum method. This is a relatively inexpensive measurement method and is easy to set up for many regular shaped objects. This method avoids some of the problems associated with a single wire pendulum such as the need to make numerous wire diameter measurements.

The general arrangement of the bifilar suspension method is shown in Figure 9.12. Since both suspension wires must be spaced equally from the center of gravity of the test object, and since both wires must be of equal length, this method may be difficult to set up for irregular shaped objects.

The equation governing this inertia is

$$I = \frac{Wd^2t^2}{16\pi^2L} \quad (9.11)$$

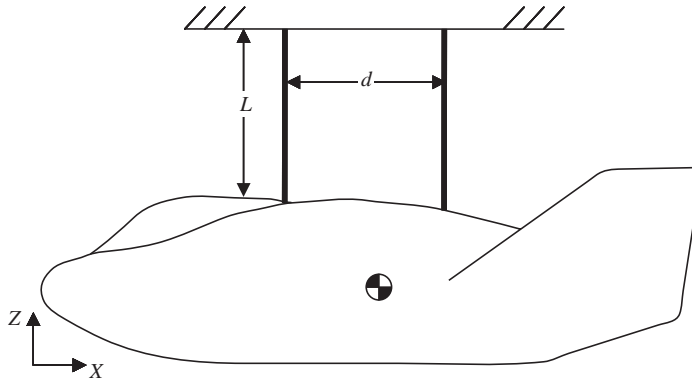


FIGURE 9.12 MOI measurement (in this case yaw) using bifilar suspension method.

where W is the weight of the object, d is the distance between the suspension wires, t is the time in seconds, and L is the length of the wires. To get the inertia in mass units, I must be multiplied by gravitational acceleration, g .

Typical measurement uncertainty for the bifilar method is about 2%, with experienced individuals routinely achieving 1% by selecting cord spacing and lengths and averaging numerous measurements.

9.6.3 Inverted Torsion Pendulum

Modern moment of inertia instruments consist of an inverted torsion pendulum that oscillates in a rotational sense and a means of measuring the exact period of oscillation of the torsion pendulum. Instead of hanging from a torsion rod or wire, the test object rests on a precision rotary table attached to the top of the instrument. Low friction bearings support the table and payload while constraining the motion of this torsion member to pure rotation (see Figure 9.13). Air bearings provide the best performance.

The measurement of the moment of inertia of the test part is based on the change in the natural frequency of oscillation of the torsion pendulum resulting from the addition of the test part mass. This change in natural frequency is compared with the change in natural frequency that occurs when a calibration mass of known moment of inertia is placed on the instrument.

In this process, the object is secured to the table with its CG aligned with the axis of the bearing. The part is rotated and released. It will then oscillate about the fixed axis of the instrument and the total time for one complete oscillation can be displayed on a digital period counter. The total combined moment of inertia of the test object, its fixture, and the instrument itself can be calculated from the formula

$$I_x = Ct_x^2 \quad (9.12)$$

where I_x is equal to the total moment of inertia, C is the calibration constant of the instrument (a function of its torsional stiffness), and t_x is the period of oscillation in seconds. Then the test object is removed from the instrument, and the tare moment of inertia of the instrument and the fixture is determined by measuring the oscillation

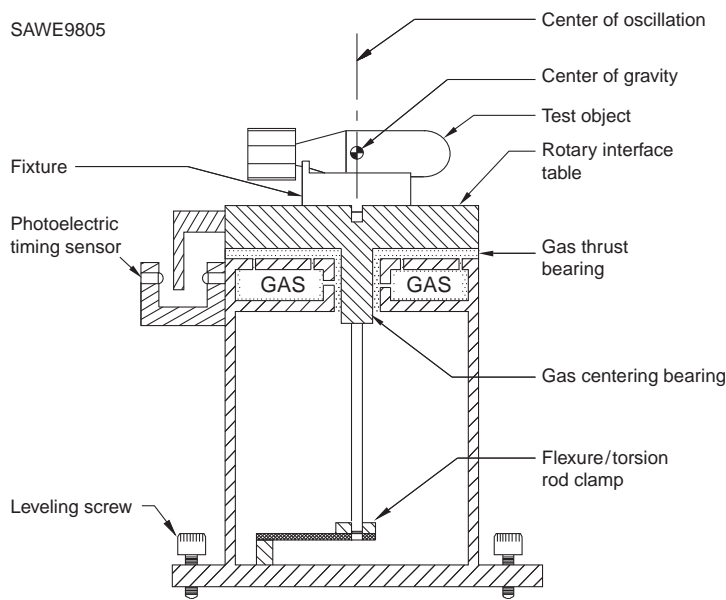


FIGURE 9.13 MOI measurement using inverted torsion pendulum.

time without the test object.

$$I_o = Ct_o^2 \tag{9.13}$$

The moment of inertia of the test object is then the difference between the total inertia and tare inertia.

$$I = I_x - I_o = \text{net moment of inertia of object} \tag{9.14}$$

In order to establish the value of the calibration constant, C , of the instrument, MOI calibration standards are measured. MOI calibration standards are precision weights of simple geometry, known mass, and known physical dimensions.

The calibration procedure is identical with the procedure for measuring the moment of inertia of an object of unknown MOI, except that in the computation, the inertia is a known quantity and the value of the calibration constant is the unknown, which must be solved for.

Because the weight of the object is supported by the air bearing, these instruments are linear over a wide range of test part weight and moment of inertia. Only a single calibration measurement is required to establish the value of the calibration constant used for all measurements.

9.6.4 Combined CG and MOI Measurement

For combined CG and MOI measurements, a spherical gas-bearing instrument, as shown in Figure 9.14, is used. In this type of instrument, the test object (and fixture) is supported by a spherical gas bearing (Society of Allied Weight Engineers, 2006). An overturning moment restraint is provided by a hollow tube, which extends from the base of the

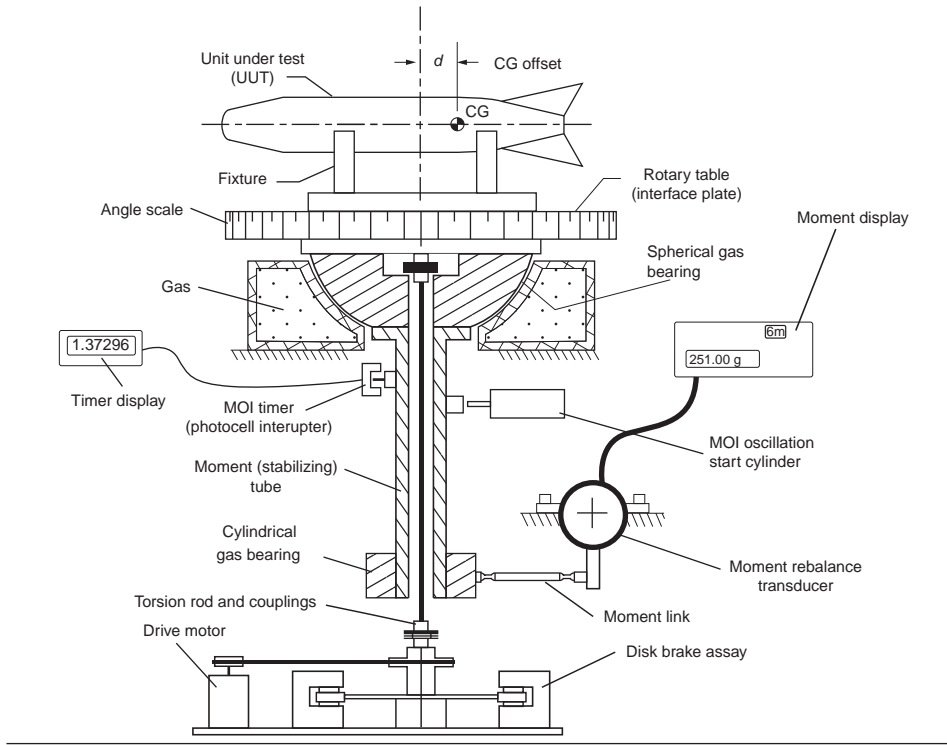


FIGURE 9.14 Basic elements of a spherical gas bearing type of instrument.

spherical bearing. The lower end of this hollow tube is attached to a second (cylindrical) gas bearing, which is connected through a moment rebalance transducer to the rigid instrument base structure. For maximum accuracy, the overturning moment produced by a displacement of the test object CG from the center of rotation of the table is sensed using this force restoration transducer. Measuring this moment and dividing by the test object weight yields the CG displacement from the center.

A torsion rod extends downward from the upper surface of the bearing (the rotary table) to a clamping mechanism at the bottom of the rod. When this clamp is released, the spherical bearing is free to turn about the vertical axis for CG measurement and when clamped, the machine is converted to a torsion pendulum for MOI measurement.

When the rotary table concept is implemented with a spherical gas bearing, the resulting instrument is capable of higher accuracy than any other method. Instrument CG measurement uncertainty is typically ± 0.001 in. (0.0254 mm), and moment of inertia measurement uncertainty is 0.1%.

9.7 POI MEASUREMENT

The product of inertia is a measure of an object's dynamic imbalance and is used to determine rotational stabilities of the object.

9.7.1 MOI Method of Measuring POI

This POI measuring method uses a torsion pendulum to determine POI by making use of the mathematical relationship between POI and MOI of an object. Special fixtures must be constructed to move the object to a number of positions while keeping both the object and the fixture CG near the center of oscillation. The moment of inertia of the object is measured in each orientation. The tare MOI of the fixtures must then be measured and subtracted from the measurements with the object. The net MOI of the object in the different orientations is then used to determine the POI of the object. The calculations are quite complex, so a computer is used.

The axes of minimum and maximum MOI correspond to the axes where the POI is zero. The POI is a maximum at an angle of 45° from these principal axes.

If the test part was fixtured so that it could be rotated through an angle C about a horizontal axis (i.e., the Z-axis) and MOI measured about numerous axes in the X-Y plane, including the X and Y axes, the MOI would be found to vary sinusoidally. If the angle C ranges over 180° , the maximum and minimum values of MOI can be seen in a plot of MOI versus C (Figure 9.15).

The axes about which the maximum and minimum MOIs are measured are the principal axes. For all other axes, the moment of inertia I_{Axy} , about an axis (A) in the X-Y plane at an angle C from the +X-axis, and the product of inertia P_{xy} are related through the equation

$$I_{Axy} = I_{yy}\sin^2 C + I_{xx}\cos^2 C - P_{xy}\sin^2 C \quad (9.15)$$

Solving this equation for P_{xy} forms the basis for the MOI method of POI determination

$$P_{xy} = \frac{I_{yy}\sin^2 C + I_{xx}\cos^2 C - I_{Axy}}{\sin^2 C} \quad (9.16)$$

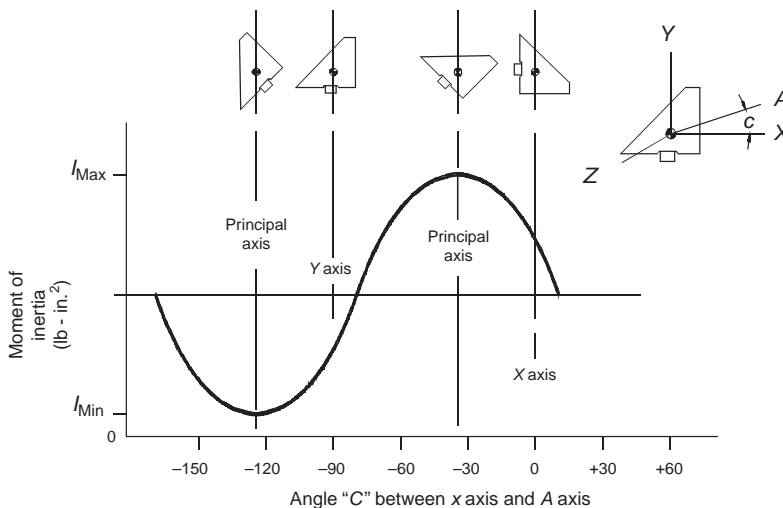


FIGURE 9.15 Minimum and maximum POI values located on principal axes.

POI in a single plane requires three MOI measurements about axes that go through the CG. These measurements are

1. MOI about the yaw axis (I_{xx})
2. MOI about the pitch axis (I_{yy})
3. MOI about an axis at 45° between pitch and yaw (I_{aa}). Note: Any angle C may be used but 45° provides the most accuracy.

These three measurements permit calculating the POI in the pitch/yaw plane using the equation:

$$P_{xy} = \frac{(I_{yy} - I_{xx})}{2 - I_{aa}} \quad (9.17)$$

For the general case, the total number of MOI measurements needed for POI calculations is nine; three in each of three mutually perpendicular planes. The intersections of these planes must be at the CG, so the MOI about each of these axes will be common to two planes, thus reducing the total number of measurements to six: three about the X , Y , and Z axes, and three about axes at 45° between the X - Y , Y - Z , and Z - X axes. If vacuum data are required, the same six MOI measurements must also be repeated in a helium atmosphere (Boynton et al., 1991).

9.7.2 Spin Balance Method

Product of inertia is generally measured using a spin balance machine. In this type of machine, the object is rotated at a fixed speed, and the reaction forces against the upper and the lower spindle bearings are measured. Product of inertia is then calculated automatically by the machine's computer, using formulas that involve the vertical spacing between the upper and the lower bearings, and the height of the payload CG above the mounting surface of the machine. The CG offset of the part may be separated mathematically from the dynamic unbalance of the part.

When the test object spins, there are two forces acting through the CG of the object: gravity forces acting downward and centrifugal forces acting horizontally (the higher the spin speed, the less significant the gravity force is). The magnitude of the downward gravity force is equal to the weight of the object in pounds (M_1). The magnitude of the horizontal centrifugal force is

$$F_1 = M_1 \times R_1 \times S^2 \quad (9.18)$$

where M_1 is the mass of unbalance in slugs, R_1 is the radius of CG in feet, and S is the speed in radians per second.

Converting the mass into weight and the speed into RPM

$$F_1 = \frac{W_1 \times R_1 \times (\text{RPM})^2}{35,207} \quad (9.19)$$

where W_1 is the weight of unbalance mass in lbs, R_1 is the radius of CG of unbalance in inches, RPM is the speed of rotation in RPM, and 35,207 is the constant to transform units.

The horizontal forces applied to the bearings of the balancing machine will depend on the geometry of the balancing machine and the type of unbalance. If neither a CG offset nor a POI unbalance is present in the rotating test object, then the horizontal forces on the bearings will be zero. A POI (with no CG offset) results in equal forces being applied 180° out of phase on the two bearings (the couple due to the product of inertia is offset by an equal couple on the bearings of the balancing machine). A CG offset will cause a different force to be applied to each bearing. In order to evaluate the accuracy of a balancing machine in measuring CG, it is necessary to know the relative values of POI and CG offset. For a flywheel, POI is usually small, and a balancing machine will be able to accurately measure CG offset. For a tall rocket, for example, the reverse is true.

If the goal of the measurement is to ballast the test object for minimum POI and CG offset about a particular axis, then the measurement becomes much more accurate. As the unbalance is reduced in successive iterations, the residual unbalance can be measured on a more sensitive scale. As the magnitude of POI is reduced, the CG sensitivity improves and as the CG offset is reduced, POI sensitivity improves.

To calculate the CG offset from the measured bearing forces, we can make use of the fact that the sum of the moments around each transducer must be zero (since the system is stable). Independent calculations can be made at the X and Y axes, and the resultant of these two calculations determined to yield the magnitude and angle of the CG offset.

Two transducers are required to separate POI from CG offset. In order to separate the unbalance due to POI from the unbalance due to CG offset, it is necessary to use a force transducer on each of the spindle bearings in the balancing machine. Then both CG offset and product of inertia are obtained in a single measurement.

9.8 MEASURING LARGE VEHICLES

9.8.1 General Process

Measuring mass properties of complex vehicles such as ships and aircraft is more complicated than measuring an individual component. This increased complexity is a result not only of the sheer size of the objects but also from the numerous potential configurations of the vehicles. Therefore, the task of measuring large complex vehicles is twofold: first, measuring the vehicle in its current configuration; second, determining the vehicle's standard operating condition. Once the mass properties of the vehicle in a standard operating condition are known, it is possible to calculate the mass properties of any conceivable configuration. Depending on the quantity of variable items on the vehicle, the vehicle may be placed in the standard operating condition prior to the measurement or a comprehensive survey of all variable items may be performed. The resultant answer is more accurate if the vehicle is measured in the standard operating condition, but in many cases such an effort is impractical. For most ships this is the case.

In most cases, only the weight and center of gravity of large vehicles can be measured directly. Other mass properties are normally calculated from component databases or derived from performance testing.

9.8.2 Weighing Ships

Ships are too large to be weighed in the conventional sense. The determination of a ship's weight and center of gravity involves calculations using observed measurements and

calculated volumetric properties (hydrostatics). The primary method for measuring the weight and CG of a ship is the inclining experiment. This experiment provides the displacement of the ship (the weight of the ship in terms of weight of water displaced) and the vertical, longitudinal, and transverse centers of gravity. A ship's MOI and POI cannot be measured and are calculated values.

9.8.2.1 The Inclining Experiment Archimedes' principle states that the weight of a floating object equals the weight of the fluid displaced by that object. This principle is used in the inclining experiment. The ship's drafts (the distance from the keel to the waterline) are recorded, and water samples are taken to determine the water density. Once this information is collected the ship's weight can be calculated by consulting the ship's hydrostatics table, which provides calculated displaced volume for the ship at various drafts. These volumes can be mathematically adjusted for the ship's trim and list and longitudinal flexing of the hull (hogging or sagging).

Once the displaced volume of the ship is known, the weight (displacement, Δ) is determined by multiplying this volume by the density of the water as determined from the samples. The longitudinal center of gravity is determined by locating the center of gravity of the displaced water (also known as the center of buoyancy) as the ship will find equilibrium such that this center is vertically aligned with the center of gravity of the ship.

The vertical center of gravity of a ship cannot be determined by merely reading the ship's drafts because the ship's static attitude is not a function of the vertical center of gravity. In order to determine the vertical center of gravity, the ship is inclined slightly by moving a known weight across the deck of the ship. Moving this weight creates a heeling moment, which induces a list on the ship. By measuring the angle of list (θ), the weight of the movable weight (w), and the distance the weight is moved (d), the distance between the ship's metacenter (KM), and vertical center of gravity (KG) can be determined (see Figure 9.16). The measure of initial stability is called the metacentric height (GM) and can be calculated by

$$GM = \frac{wd}{\Delta \tan \theta} \quad (9.20)$$

The distance from the ship's keel to its vertical center of gravity (KG) can be calculated from the equation

$$KG = KM - GM \quad (9.21)$$

where KM is a property of the ship's geometry which is usually listed for each draft in the ship's hydrostatics table.

Inclining Procedure The basic procedure for an inclining experiment as outlined below has not changed greatly in the past century except for improved instrumentation and data processing:

Inventory or Deadweight Survey: This survey is performed either before or after the inclining experiment. This survey consists of inspecting every compartment of the ship for weights that are not part of the lightship condition (ship without

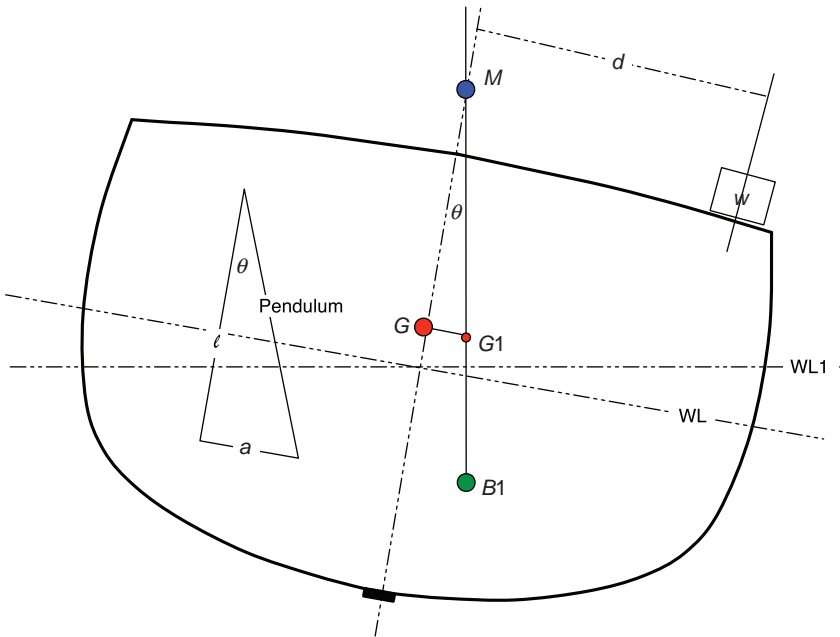


FIGURE 9.16 Diagram of ship inclining experiment method to measure vertical center of gravity.

consumable loads) and recording their weight and CG. The survey also includes recording missing lightship items.

Draft Readings: In calm water, the ship's draft (the distance from the keel to the waterline) is recorded in several locations (often six; port and starboard at three locations along the ship's longitudinal axis) and water samples are taken to determine the water density.

Weight Movements: As shown in Equation 9.20, a measure of the ship's stability, GM , can be calculated by moving a known weight a known distance athwartships, and recording the resultant list angle. In actuality, a series of weight movements are done and angle measurements taken for each movement. The angle of inclination is normally less than 3° .

Tangent Readings: For each weight movement, the angle of heel is recorded through pendulums or inclinometers. The tangents of the angles are plotted against the heeling moment, and the slope of the fitted line is used to determine the GM :

$$GM = \frac{\text{Slope}}{\Delta} \quad (9.22)$$

Correction to Standard Loads: From the draft readings and the weight movements, the ship's weight and CG is known for the as-inclined condition. The dead-weight survey data are used to calculate the lightship weight and CG, and then standard loads are added to determine the weight and CG of each standard condition. These standard conditions are tracked as part of the normal ship weight control process.

9.8.2.2 Sources of Uncertainty in Ship Measurement The weight of a ship is not directly weighed; therefore, the uncertainty of the result is dependent on more factors than a conventional weighing. These sources of uncertainty can be broken down into factors affecting the accuracy of the draft readings and deadweight survey. Sources of uncertainty in the former include building tolerances, measuring the ship's drafts, uncertainties in the water density profile, and uncertainties in the ship's hydrostatic calculations. Sources of uncertainty in the deadweight survey are largely related to data recording and estimates of the weights of the variable items, but also include uncertainties due to the build tolerances and calculated capacities of tanks. Liquid loads on a ship can be a significant portion of the total weight; thus the uncertainty of the loads is often dominated by the uncertainty of the liquids.

Greater detail on the uncertainties associated with an inclining experiment or a dunnage survey can be found in Shakshober and Montgomery (1967), Tellet (2005), and Cimino and Tellet (2007).

9.8.3 Weighing Aircraft

Unlike ships, the mass properties of an aircraft are determined directly by the use of calibrated scales and dimensional readings of the complete vehicle. The two main mass properties of interest are the weight of the aircraft and its horizontal center of gravity. These measurements should be performed on all new aircraft and after each major upgrade or overhaul that impacts the vehicle's configuration. This information is critical to the safe and efficient performance of the aircraft. There are many cases in which the mass properties of the plane have been neglected and damage to the plane has occurred during loading, or even the worse the plane suffered a loss of control during takeoff and crashed with serious consequences.

9.8.3.1 Weighing Preparation Proper premeasurement preparation is essential to obtaining an accurate and precise weight and CG for the aircraft. The craft should be as close as possible to its basic configuration with no fuel or load items such as bombs, munitions, crew members, or equipment not having a fixed position onboard the aircraft. All liquid reservoirs and tanks should be empty or filled to normal capacity. The hydraulic fluid reservoir and all other reservoirs containing contents required for the normal operation of the aircraft should be full. All waste tanks should be empty. An aircraft should never be weighed with partially filled tanks or reservoirs.

All fuel should be drained in accordance with manufacturer's instructions. Any fuel remaining in the system after draining should be considered part of the empty aircraft weight. If it is not feasible to drain the fuel, the fuel tanks should be topped off and the weight of the fuel and its moment subtracted from those of the weighed aircraft. Engine lubricating oil should also be treated similar to fuel.

All permanent ballast must be properly installed and any temporary ballast removed from the aircraft prior to weighing.

A complete inventory of all onboard equipment should be conducted before weighing identifying each piece of onboard equipment and its weight and location.

The aircraft to be weighed should be in a clean condition. All dirt, grease, and moisture should be eliminated and sufficient time should be provided for the aircraft to dry prior to weighing. The aircraft should also be positioned in a closed hangar with no blowers or ventilation systems impinging on any surface of the aircraft. The slope of the floor should

be as level as possible, not exceeding 0.25 in./ft (21 mm/m). Further information on the preparations required prior to weighing an aircraft can be found in United States Air Force (2008) and Federal Aviation Administration (2007).

9.8.3.2 Weighing Equipment and Procedures There are two types of scales used to weigh an aircraft. In the first type, scales are onto which the aircraft is rolled so the weight can be read at each wheel. In some cases, scales are portable and can be used in any hangar; and in other cases, large scales are permanently installed flush with the floor. In both instances, the aircraft is rolled onto the scale surfaces, and weight of the aircraft is read directly on an electronic readout. This type of weighing completely eliminates the need for jacks and the risk of side loads which can be present. The second type employs electronic load cells that are attached to axle jacks or wing jacks. When the aircraft is raised and leveled, the weight is measured by the load cells and transmitted electronically to a readout device. Other accessories that are used during weighing are a spirit level, leveling bars, plumb bobs and chalk lines, steel measuring tapes, and hydrometers for measuring the density of fuel or oil. Figure 9.17 shows the arrangement of an aircraft weighing using three load cells.

The basic weighing procedure using platform scales is as follows:

1. Zero the scale.
2. Tow aircraft onto the scales.
3. Level the aircraft.
4. Read the scales and make dimensional measurements for CG calculations.
5. Remove the aircraft from the scales.

The basic weighing procedure to be followed when electronic load cells are employed is as follows:

1. Level the aircraft.
2. Position the jacks below jack pads or designated lifting points.
3. Remove all chocks and release brakes.
4. Jack the aircraft in accordance with jacking instructions using all jacks simultaneously.

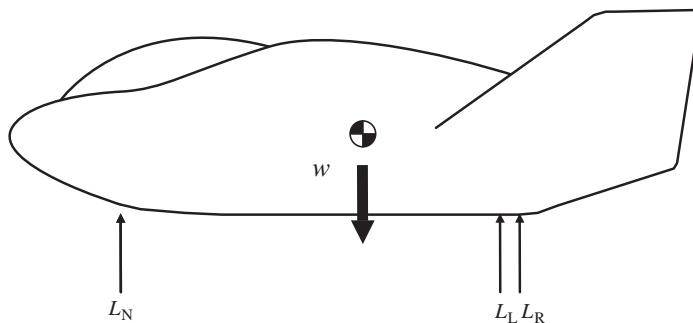


FIGURE 9.17 Weighing an aircraft using three load cells.

5. Level the aircraft.
6. Read the scales for weight and make dimensional measurements for CG determination.
7. Lower the aircraft and remove the jacks and load cells from the aircraft.

More detailed information on the process for weighing an aircraft may be found in United States Air Force (2008).

9.8.3.3 CG Determination The horizontal CG of an aircraft is determined by dividing the total horizontal moment by the total weight of the vehicle. When the plane is at a level attitude, a plumb line is dropped from a datum point (an easily identifiable physical location usually specified by the manufacturer from which measurements can be made) to the floor below. A chalk line is then drawn through the point parallel to the longitudinal axis of the aircraft. A lateral line is then drawn between the actual weighing points (main wheels) and the point is marked along the longitudinal line at the weighing point for the other (nose or tail) weighing point. These lines and marks allow accurate measurements between the datum and the weighing points to determine their moment arms. The CG is then determined by adding the weight and moment of each weighing point to arrive at the total weight and moment. Dividing the total moment by the total weight results in the CG relative to the datum.

The location of the CG may be expressed in terms of distance from the datum point or as a percentage of the leading edge mean aerodynamic cord (MAC) which is located a specified distance from the datum by the manufacturer. For a more detailed description of determining the CG of an aircraft refer to United States Air Force (2008) and Federal Aviation Administration (2007).

9.8.3.4 Sources of Uncertainty in Aircraft Weighing The main sources of uncertainty in the weighing of an aircraft are inaccuracies in the inventory of the equipment onboard, inaccuracies in the scales being used, or human error in the performance of the weighing activity. The only other major source of uncertainty involves the aircraft and its interaction with its surroundings. Every precaution should be taken to ensure that the aircraft is not acted upon by any force that will induce any loading, horizontal or vertical, on the aircraft being measured. This is the reason for the closed hangar and turning off blowers or ventilation fans described in the preparation for weighing. The scale used should be calibrated, exercised, and zeroed in strict accordance with the manufacturer's instructions to ensure its accuracy and repeatability. It is good engineering practice to perform the weighing at least twice to reduce the chance for error.

9.9 SOURCES OF UNCERTAINTY

As in all engineering fields, there are uncertainties in the measurement of mass properties that must be considered. The previous sections have touched on the relative accuracies of specific instruments or methods. Following are some specific sources of uncertainty that should be considered in any mass properties measurement.

9.9.1 Weight Uncertainties

The mass of an object is fixed and is the same whether the object is on the Earth or in outer space. Weight, on the other hand, is a force that depends on several factors that are related to the location of the scale. With the advent of force restoration technology, it is possible to measure an object at different locations on the Earth and observe significantly different values for weight.

The force a mass exerts on a scale is affected by four factors:

1. The gravitational mass attraction to the Earth at a particular location, which is in part related to the altitude and latitude of that location.
2. The gravitational mass attraction to the sun and moon at the particular location, which may reach 0.003% of the acceleration of Earth's gravity at certain dates during the year when the sun and moon align.
3. The centrifugal force due to the rotation of the Earth, which varies from zero at the North Pole to a maximum value at the equator.
4. The buoyancy of the object as it floats in a sea of air. This can be compensated for by determining the enclosed volume and calculating the weight of the displaced air (the density of which can vary due to the weather).

These factors combine to result in a change in the weight of an object of almost 1% over the surface of the Earth, and about 0.2% over the contiguous United States. If a mass weighs 100 lb (45.36 kg) at one location in the United States, it could weigh 99.8 lb (45.35 kg) somewhere else in the United States. To get around this problem, mass is used (rather than weight) to standardize and calibrate scales. To calibrate a scale, a standard calibration mass is placed on the scale. The scale is then adjusted until it reads the appropriate standard weight. The standard weight is the weight the mass would have at standard gravity (32.17405 ft/s^2 or 9.80665 m/s^2).

The traditional "Scales of Justice" balance beam compares one mass against another mass; therefore, the measurement does not vary with changes in gravitational field strength.

A problem occurs when a scale is calibrated at one location and then moved to another location to weigh an object. For large scale capacities, it is often not possible to bring a calibration weight to the new site, either because this weight is not available, or because of the problems of shipping a calibration mass weighing many thousands of pounds. Therefore, it is necessary to correct for the change in the acceleration of gravity between the site where the scale was calibrated and the site where the object is being measured. The National Geodetic Information Center in Rockville, MD (<http://www.ngs.noaa.gov/GRAV-D/>) has data on the weight correction required for many locations on Earth. If these data are not available, then another method to determine the correction is to calibrate a small scale at the first site and ship this scale to the new site, together with its calibration weight. The small calibration weight is then remeasured. If, for example, the measured value of this small weight is 0.05% high, then the acceleration of gravity is 0.05% higher at this new location, and the measurement of the large object must be divided by 1.0005 to correct for this change in the acceleration of gravity.

The following sections discuss in detail the factors affecting weight measurement (Boynton, 2001).

9.9.2 Factors Affecting Weight Measurement

Measured weight is affected by factors from gravity variations as well as influences from local external factors such as environment and test setup.

9.9.2.1 Gravity Effects

Latitude Correction: The most significant variable in determining the acceleration of gravity is the latitude. The acceleration of gravity varies from 9.780 m/s^2 (32.087 ft/s^2) at the equator to 9.832 m/s^2 (32.257 ft/s^2) at the poles (a difference of 0.53%). The value of g is the smallest at the equator (due to the centrifugal force and the bulging of the Earth).

The radius of the Earth is approximately 22 km (13.67 miles) more at the equator than at the North or South Pole, due to the bulge in the Earth, which resulted from centrifugal forces while the Earth was cooling. In addition, the centrifugal force due to the Earth's rotation counteracts the centripetal force due to the attraction of gravity.

The force is

$$F_{\text{cen}} = \frac{m_2 v^2}{R} \quad (9.23)$$

where m_2 is the mass of object being weighed, R is the distance to Earth's rotational axis, and v is the speed (zero at poles; 463 m/s at equator)..

The local value for g at sea level can be calculated using the formula

$$g = 9.80613(1 - 0.0026325 \cos 2L) \quad (9.24)$$

where L is the latitude in degrees and g is in m/s^2 .

Altitude Correction: For locations on the surface of the Earth, the gravitational attraction is inversely proportional to the square of the distance to the mass center of the Earth. Therefore, gravity decreases as you increase altitude (Boynton, 2001).

At sea level,

$$g = \frac{Gm_1}{r^2} \quad (9.25)$$

where G is the universal gravitational constant, $6.67259 \times 10^{-11} \text{ nm}^2/\text{kg}^2$, m_1 is the current accepted mass of Earth in $\text{kg} = 5.9736 \times 10^{24} \text{ kg}$, and r is the mean radius from the center of Earth's mass = 6,378,100 m.

At an altitude h

$$\begin{aligned} g_h &= \frac{Gm_1}{(r+h)^2} \\ &= \frac{Gm_1}{(r^2 + 2rh + h^2)} \\ \frac{g}{g_h} &= \frac{(r^2 + 2rh + h^2)}{r^2} \\ &= 1 + \frac{2h}{r} + \frac{h^2}{r^2} \end{aligned} \quad (9.26)$$

$\frac{h^2}{r^2}$ is extremely small, so we assume $\frac{g}{g_h} = 1 + \frac{2h}{r}$ and thus

$$g_h = \frac{g}{\left(1 + \frac{2h}{r}\right)} \quad (9.27)$$

In addition, the centrifugal force increases as the radius increases, resulting in a further decrease in the acceleration of gravity. This centrifugal effect depends on the latitude. The decrease in g due to rotation is zero at the poles and reaches a maximum of about 1.567×10^{-4} m/s per kilometer of altitude at the equator.

The altitude correction is a relatively small number. If the altitude is increased by 3000 m, then the measured weight decreases by 0.11%. However, high accuracy scales can detect changes in measured weight as small as that which results from an increase in elevation of only 40 m, so the scale should be calibrated at the same altitude as the measurement.

Tidal Variation: An object on the surface of the Earth is attracted to every celestial body. Most of these masses are too far away to have any significance on weight, but the sun and the moon do have a significant effect. If you have a scale with an accuracy of 0.003% or better, you will notice that the weight of an object varies as a function of the time of day. This effect is most pronounced during spring and fall when the sun and moon align.

Correcting for Variations in the Acceleration of Gravity

Method 1: Recalibrate the scale at a new location. Once you have performed a calibration using a mass which is traceable to NIST at a specific location, you have automatically corrected for the acceleration of gravity at the location. No mathematical manipulation of the data will be necessary to correct for latitude or longitude.

Method 2: Correct for the acceleration of gravity by using a small scale and calibration weight. If you have a means of calibrating your scale at one location, but you then need to move your scale to another location and you are unable to bring a large calibration mass to this location, you can accurately determine the correction necessary by the following procedure, which uses an additional small scale and small calibration mass.

1. Calibrate the small scale at the first location. From this point on you may adjust the zero of this scale but not the span.
2. When you arrive at the second location, place the small calibration weight on the small scale and note the measured weight. If the acceleration of gravity is lower at the new location, then the measured weight of the small calibration mass will be low.
3. Use your large scale to measure the unknown object. Then multiply the large scale reading by a correction factor equal to the true mass of the small calibration weight divided by the small scale reading in step 2.

Method 3: Correct for acceleration of gravity by knowing the altitude and latitude of the location where the scale was calibrated and where the final measurement will

take place. In some instances, you will have no means of calibrating a scale. It may have been calibrated at the factory only. In this case, you will need to know the acceleration of gravity at the location where the scale was calibrated and also at the location where you will be making your measurements.

Although anomalies in the Earth's crust will affect local gravity, gravity is mainly a function of latitude and altitude. Gravity at a particular location can be calculated with a precision of about 0.002% using the following formula

$$M = \frac{(g_c)(\text{Measured mass})}{g_m} \quad (9.28)$$

where M is the true mass, g_c is the acceleration of gravity where scale (or load cell) was calibrated, and g_m is the acceleration of gravity where the measurement was made.

Method 4: Correction for acceleration of gravity by using information available from the National Geodetic Information Center. Values for the acceleration of gravity at specific locations can be purchased from the National Geophysical Data Center. The true mass can be calculated by using the formula given in Method 3.

9.9.2.2 Other Effects

Buoyancy: If an object is immersed in a fluid or gas, the object experiences a buoyant force equal to the weight of the fluid or gas it displaces (Archimedes' principle). This explains why a balloon rises. When an object is weighed in air, there is an upward force introduced which reduces the measured weight. For large lightweight objects, this effect can be quite significant. Air has an average density of about 1.20 kg/m^3 (0.0749 lb/ft^3). If an object has a volume of 1 m^3 , then 1.2 kg must be added to the measured mass of the object to yield the true mass. This correction is particularly important for space vehicles whose mass is measured in the Earth's atmosphere.

To complicate this analysis, the scale was originally calibrated using a standard mass. This mass has volume and also experiences an upward force due to buoyancy. Therefore, for very accurate measurements, the calibration adjustment of the scale must be corrected for the buoyancy of this weight. This effect is very small. Usually, a calibration mass is made of solid metal and has a density of 8000 kg/m^3 . Therefore, its buoyancy error is about 1 part in 6666 (or 0.015%).

Method to correct for buoyancy, using calculated air density

1. Calibrate the scale and adjust its output to read the exact value of the calibration weight.
2. Determine the density of the calibration weight (brass = 8570 kg/m^3 ; stainless steel = 7870 kg/m^3 ; cast iron = 7100 kg/m^3).
3. Measure the weight of the object.
4. Determine the volume of the air displaced by the object (often not an easy task).
Note: If openings allow air to enter a space, then this space is not part of the volume of air displaced.
5. Determine the approximate density of the object by dividing the measured weight by the volume of air displaced.

6. Determine the density of the air by using the following formula

$$\rho = \frac{(0.020582h + 0.348444p - 0.00252t \times h)}{(t + 273.15)} \quad (9.29)$$

where p is the atmospheric pressure in mbar (hPa), h is the relative humidity in %, t is the temperature in degrees Celsius, and ρ is the air density in kg/m^3 .

7. Perform the buoyancy correction using the formula

$$\text{True mass} = \text{measured mass} \times \left[\frac{1 - (\rho/s)}{1 - (\rho/u)} \right] \quad (9.30)$$

where s is the density of calibration weight in kg/m^3 , u is the density of unknown in kg/m^3 , and ρ is the density of air in kg/m^3 .

One can also correct for buoyancy effects by weighing two objects of identical mass but different volumes. The difference in measured weight will be

$$w_1 - w_2 = \delta(v_2 - v_1) \quad (9.31)$$

$$\rho = \frac{w_1 - w_2}{v_2 - v_1} \quad (9.32)$$

This value for ρ can then be used in the formula in step 7 mentioned in Section 9.2.2.

Condensation: In areas with high humidity, sudden changes in temperature can produce condensation on the object being measured, which adds to the weight of the object. This most commonly occurs when an object is brought from an air conditioned space to a non air conditioned one. Condensation can be minimized by allowing an object's temperature to equalize before weighing it.

Electrostatic Attraction: It is remarkable how dramatic this effect can be if you are weighing a large lightweight object such as a Mylar decoy reentry cone. If the object is covered by a clear plastic draft shield during measurement, the attraction between object and shield can be as much as 2% of the weight of the object.

Magnetic Attraction: If the object you are weighing contains permanent magnets, there will be an attraction to any magnetic material near or on the scale. There will also be a small attraction force to the magnetic north of the Earth. It may be necessary to demagnetize the object before weighing it. Generally, you can detect magnetic errors by repositioning the object on the weighing pan of the scale. If the readings are very sensitive to position on the pan, then the problem may be magnetic attraction (but it might also be the corner loading error of the scale). Using a compass is also a good way to test for magnetic attraction.

Drafts or Air Currents: Generally, this effect will be quite obvious, since drafts will introduce a random variation in the readings. However, there are some instances where drafts can produce a relatively steady downward or upward force. These effects can be minimized by making sure that the object is at the same temperature as the surrounding air, by avoiding direct sunlight or bright lights, and by turning off any source of airflow in the immediate test area.

9.9.3 Sources of MOI Uncertainty

9.9.3.1 Coincident CG It is important that the center of gravity of the test part and fixture be positioned as close as possible to the rotational axis of the torsion pendulum. Otherwise, measurement error will increase for a number of reasons.

One source of error is the so-called gravity pendulum effect that occurs because the axis of rotation of the torsion pendulum can never be made exactly vertical. If a part is oscillated about an axis which does not fall on its center of gravity, then the force of gravity acting through the center of gravity will tend to bring the center of gravity to its low point (i.e., the direction in which the axis of rotation of the torsion pendulum is tilted), resulting in a change in the effective calibration constant of the instrument. This effect can be minimized by leveling the table and repositioning the test part, so that its center of gravity lies close to the rotational axis of the instrument.

A second source of error with offset center of gravity is the classical axis translation error. Mathematically, this increase is equal to the test part mass times the square of the offset distance. Since this increase can be exactly calculated, it can be subtracted from the measured moment of inertia; therefore, it does not constitute an uncertainty. For small offsets (less than 1% of the radius of gyration), this effect is negligible. For larger offsets, instruments are available which measure both CG and MOI. The software then automatically reports the MOI about the center of oscillation and about the CG. To utilize this feature, the test part must be weighed.

9.9.3.2 Effect of Air Mass For large lightweight objects, the measured mass properties are often different from the calculated values. In particular, measured moment of inertia can be 10–20% larger than calculated. Inexact analysis (e.g., neglecting to input the MOIs of small parts into the model) often contributes up to 5% or more to this discrepancy. A second consideration is that air has significant mass and alters the mass properties in two ways.

The first way is when air is trapped inside the payload. This will increase its mass by an amount equal to the unoccupied volume in the payload times the density of air (0.0749 lb/ft^3 [1.2 kg/m^3]). The other air effect is when air is dragged or pushed along by any protrusions on the outer surface of the payload. This can dramatically increase moment of inertia. For example, the roll moment of inertia of a missile flying in air is much larger than the roll MOI of the missile in a vacuum.

How you handle this difference depends on whether the payload operates in the vacuum of space or in air. If the payload flies in a vacuum, then measured values must be decreased to eliminate the effect of air mass. The best way of doing this is to make a second measurement in helium and then extrapolate the value in vacuum (Boynton et al., 1991).

9.9.3.3 Minimum MOI that can be Measured Moment of inertia instruments have amazing dynamic ranges. An MOI instrument that is designed to measure objects weighing up to 3000 lb (1361 kg) can often detect the change in MOI due to the addition of an object weighing 0.1 lb (0.05 kg). However, accuracy is reduced when the MOI of an object is smaller than the tare MOI of the instrument. The error is primarily due to the inherent 0.1% limitation accuracy of the instrument.

As an extreme example, if a payload has a MOI of 10 lb-in^2 and the instrument has a tare MOI of $10,000 \text{ lb-in}^2$, then a 0.1% uncertainty in tare is 10 lb-in^2 . When the payload

is mounted, the uncertainty in the gross measurement ($10,010 \text{ lb-in}^2$) is 0.1% of the gross or 10.01 lb-in^2 . The sum of these ($10 + 10.01 = 20.01$) is the total uncertainty in the net payload MOI so the final answer would be payload MOI = $10 \pm 20.01 \text{ lb-in}^2$, which is a more than 200% uncertainty. For this reason, the ratio of gross to net should be kept to a minimum. A ratio greater than two dramatically reduces the accuracy of the measurement.

Another source of uncertainty is thermal expansion and contraction of the instrument and fixture during the time between tare and object MOI measurement. Reducing short-term temperature change can increase the usable range of an instrument. Improved accuracy can be achieved by simply shutting off the heating or air conditioning system during the interval of time between tare and object measurement.

9.9.3.4 Damping Air-bearing MOI instruments themselves have very small losses, and the effect of this damping can generally be ignored. For payloads which introduce significant damping due to air turbulence while oscillating, the actual period of oscillation is greater than the undamped natural period by an amount determined by the damping ratio, z . If the torsion pendulum is being used as an instrument to measure moment of inertia, then the measured moment of inertia will be greater than the true value. This error can be eliminated if the following equation is used in place of Equation 9.12. The quantity z^2 is the error.

$$I = Ct^2(1 - z^2) \quad (9.33)$$

In order to make use of this equation, the value of the damping ratio, z , must be determined. This is accomplished by noting the rate at which the amplitude of oscillation decays. If we define the logarithmic decrement as the natural logarithm of the ratio of any two successive amplitudes, then the log decrement, d , of the starting amplitude, a_0 , as compared to the peak amplitude, a_n , after n cycles have elapsed is given by the equation:

$$d = \frac{\ln(a_0/a_n)}{n} \quad (9.34)$$

For small values of z , the logarithmic decrement, d , can be related to z by the following relationship

$$d = 2\pi z \quad (9.35)$$

If we now count the number of oscillations of our torsion pendulum, n , for a decay in peak amplitude of 10:1, we may combine the above equations and solve for the error resulting from damping. It should be noted that this effect is in addition to the increase in MOI due to entrained air.

Percentage of error due to damping = $100z^2$

$$\text{Percentage of error} = \frac{100(\ln 10^2)}{(2\pi n)^2} \quad (9.36)$$

$$\text{Percentage of error} = \frac{13.41}{n^2} \quad (9.37)$$

REFERENCES

- Boynton R. Techniques for improving center of gravity measurement accuracy. 36th Annual Conference of the Society of Allied Weight Engineers; San Diego, California; May 9–12, 1977; SAWE Paper No. 1169.
- Boynton R. Precise measurement of mass. 60th Annual Conference of the Society of Allied Weight Engineers; Arlington, Texas; May 19–23, 2001; SAWE Paper No. 3147.
- Boynton R, Bell R. Measuring moment of inertia through test part CG when CG location is unknown. 40th Annual Conference of the Society of Allied Weight Engineers; Dayton, Ohio; May 4–7, 1981; SAWE Paper No. 1440.
- Boynton R, Bell R, Wiener K. Using helium to predict the mass properties of an object in the vacuum of space. 50th Annual Conference of the Society of Allied Weight Engineers; San Diego, California; May 20–22, 1991; SAWE Paper No. 2024.
- Boynton R, Wiener K. Mass properties measurement handbook. 57th Annual Conference of the Society of Allied Weight Engineers; Wichita, Kansas; May 18–20, 1998; SAWE Paper No. 2444.
- Cimino D, Tellet D, editors. *Marine Vehicle Weight Engineering*. SAWE, Inc.; 2007.
- Federal Aviation Administration. *Weighing the Aircraft and Determining the Empty Weight Center of Gravity*. FAA Aircraft Weight and Balance Handbook FAA-H-8083-1A; Federal Aviation Administration; 2007.
- National Institute of Standards and Technology. *Specifications, Tolerances, and Other Technical Requirements for Weighing and Measuring Devices*. NIST Handbook 44. NIST; 2011.
- SAWE Members. *Weight Engineers Handbook*. Society of Allied Weight Engineers; 2007.
- Shakshober MC, Montgomery JB. Analysis of the Inclining Experiment. *Presentation to Hampton Roads Section of the Society of Naval Architects and Marine Engineers*; 1967 Feb.
- Society of Allied Weight Engineers. Recommended Practice 13, Standard Coordinated System for Reporting Mass Properties of Surface Ships and Submarines; 1996 Jun.
- Society of Allied Weight Engineers. Recommended Practice 16, Measurement of Missile and Spacecraft Mass Properties; 2006 May.
- Tellet D. Inclining experiment sensitivity analysis using excel simulation tools. 64th Annual Conference of the Society of Allied Weight Engineers; Annapolis, Maryland; 2005 May; SAWE Paper No. 3367.
- United States Air Force. Technical Manual 1-1B-50 Weight and Balance; 2008 April; Basic Technical Order for USAF Aircraft.

10

FORCE MEASUREMENT

PATRICK COLLINS

- 10.1 Introduction
 - 10.2 Force transducers
 - 10.3 Universal testing machines
 - 10.4 The strain gauge sensor
 - 10.4.1 Strain gauge circuit compensation
 - 10.5 Resonant element transducers
 - 10.6 Surface acoustic wave transducers
 - 10.7 Dynamometers
 - 10.8 Optical force transducers
 - 10.9 Magneto-elastic transducers
 - 10.10 Force balance transducers
 - 10.11 Force transducer characteristics
 - 10.11.1 Capacity
 - 10.11.2 Output
 - 10.11.3 Repeatability
 - 10.11.4 Creep
 - 10.11.5 Temperature coefficient
 - 10.11.6 Accuracy
 - 10.12 Calibration
 - 10.12.1 Uncertainty
 - 10.13 Conclusion
- Glossary of terms
- References

10.1 INTRODUCTION

Force is one of the ubiquitous physical phenomena. From the strong nuclear forces holding subatomic particles together to the weak forces holding the Earth in orbit around the sun and the solar system around the galactic center, forces are an influence in everything we see and do.

Aristotle (1934) is credited with an approach that considered the natural phenomena with a combination of observation and philosophical reasoning. He postulated that the relative proportion of “Earth,” the heaviest element, and “Fire,” the lightest element, determined the natural motion of a body. Earth tended toward the center of the universe, whereas Fire would like to rise upwards away from the center. Terrestrial objects move upwards or downwards toward their natural place in accordance with their composition of the four elements. His observations were constrained by the technology available at the time. For example, a vacuum was thought to be impossible, but if it did theoretically exist, then motion in a vacuum would be infinitely fast. The ideal speed of an object moving with terrestrial motion (upwards or downwards) was directly proportional to its weight.

Given the inability to observe a system where friction played only a negligible part, it was thought that the natural state of things was to be at rest. Forced or unnatural motion required the constant application of force to sustain the motion. Some philosophical thinking was applied to a situation such as the flight of a spear through the air. After an initial impetus, the constant force was thought to be provided by the air as the spear passed through it. The motion of heavenly bodies was considered to be different. The sun, moon, and stars were embedded in crystal spheres that rotated with an unchanging circular motion, with the planets embedded in spheres within spheres to account for their apparently erratic motion.

While we now think that the Aristotelian physics contains some shortcomings, but from a purely practical point of view, it has some resonance. Pushing a cart loaded with bags of rice to the market place requires the expenditure of a constant force to keep the cart moving.

In the seventeenth century, Galileo Galilei (1953) experimented with cannon balls and inclined slopes to theorize that acceleration due to gravity did not depend on the weight of the object and that the force of friction was responsible for slowing down the motion which would otherwise remain constant.

With the publication of Sir Isaac Newton’s “*Principia Mathematica*” (Newton, 1685) in 1687, a mathematical explanation of the laws of motion was proposed that would last for almost 300 years. Newton realized that the same force of gravity was responsible for items falling toward the Earth, and the motion of the moon around the Earth. Newton provided the means to calculate and predict the motion of bodies both terrestrial and celestial based of the forces experienced between the objects. Newton’s reasoning was so compelling that when inconsistencies in the orbit of the planet Mercury were observed that could not be fully explained by the Newtonian model, astrophysicists predicted the presence of an unobserved new planet that would account for the discrepancies. No such planet was found, but Albert Einstein (1916) sensationally added a correction to his theory of general relativity that precisely accounted for the observed motion.

So, force is a measurement of the interaction between objects. It is a vector quantity with both magnitude and direction. If a body is at equilibrium, the sum of forces acting

upon the body is zero. The derived SI unit of force is named in honor of the aforementioned Sir Isaac Newton with 1 N being the force required to accelerate a mass of 1 kg by 1 m/s/s.

In general, force is calculated using the following formula:

$$F = ma \quad (10.1)$$

where F is force in Newton, m is the mass of the object in kilogram, and a is its acceleration in meters per second per second.

Because it is not practical to accelerate a mass by 1 m/s² to produce a standard force of 1 N, we tend to use a more readily available and predictable acceleration factor to produce a known force on a transducer, that is, the acceleration due to gravity. This modifies Equation (10.1) to become

$$F = mg \quad (10.2)$$

Local gravity is relatively easy to determine and so when a known mass is applied in any geographical location, the force in Newton is easily determined also. Using such masses, force transducers can be calibrated so that wherever in the world they are used, they will report the correct force.

10.2 FORCE TRANSDUCERS

A transducer is a device that senses a physical parameter and reports it quantitatively as an output of a different type. In the case of a force transducer, the output is most commonly electrical, meaning a change in resistance, capacitance, or voltage, for example. These changes are usually quite small, so associated instrumentation is required to interpret the output quantity and convert it to a display for the operator or user to see.

Force transducers take many forms, with some of the more common forms listed in Table 10.1, along with their approximate ranges and accuracies.

The majority of force measurement devices used in industry fall into the first category, with the others satisfying more specialist applications. This chapter will accordingly emphasize the strain gauge load cell devices, with briefer descriptions of the other transducer types.

In the seventeenth century, British physicist Hooke (1678) stated “*Ut tensio, sic vis*,” roughly translated: “as the extension, so the force,” implying that there is a linear relationship between extension and applied force, and although originally derived for springs, can be applied to most elastic materials. Put another way, when a material is placed under tension, applied force (the stress) is proportional to the amount by which the material deforms (the strain), provided that the elastic limit is not exceeded. This observation is defined by the following equation:

$$F = -kx \quad (10.3)$$

TABLE 10.1 Common Force Transducers along with Their Typical Uncertainties and Temperature Characteristics (Institute of Measurement and Control, 1996)

Device Type	Typical Range of Rated Capacities	Typical Total Uncertainty (% fso)	Temperature Sensitivity and Operating Range (% fso per °C)
Strain gauge load cells			
Semiconductor gauges	1 N to 10 kN	0.2–1%	0.2–0.5% (–40°C to +80°C)
Thin film gauges	0.1 N to 100 N	0.02–1%	0.02% (–40°C to +80°C)
Foil gauges	5 N to 50 MN	0.02–1%	0.0015% (–40°C to +80°C)
Piezoelectric crystal	1.5 mN to 120 MN	0.3–1%	0.02% (–190°C to +200°C)
Hydraulic	500 N to 5 MN	0.25–5%	0.02–0.1% (5–40°C)
Pneumatic	10 N to 500 kN	0.1–2%	0.02–0.1% (5–40°C)
LVDT, capacitive, tuning-fork, vibrating wire, optical	10 mN to 1 MN	0.02–2%	0.02–0.05% (–40°C to +80°C)
Magneto-elastic	2 kN to 50 MN	0.5–2%	0.03–0.05% (–40°C to +80°C)
Gyroscopic	50–250 N	0.001%	0.0001% (–10°C to +40°C)
Force balance	0.25–20 N	1 part in 10 ⁶	0.0001% (–10°C to +40°C)

where F is the restorative force, x is the displacement, and k is a force (or spring) constant. Put another way

$$E = \text{stress/strain} \quad (10.4)$$

where stress is the force applied to the specimen, and strain is the relative elongation of the specimen from its original length.

Here, E is also known as Young's Modulus (1807), represented as the gradient of the stress/strain curve within the initial linear region. This is an extremely useful value to determine, because it gives a clear indication of how much the material will deform under applied tensile loading. A bungee rope, for instance, must have a low enough Young's Modulus to guarantee the jumper will achieve a long, thrilling jump without the rope jarring when the slack is taken up, but a high enough Young's Modulus to ensure the rope is sufficiently stiff to bring the jumper springing back skywards prior to hitting the ground.

A common stress/strain curve is shown in Figure 10.1. The trace shown in Figure 10.1 is typical of an Aluminum metal plate specimen and shows the linear region, where Hooke's law is obeyed, followed by a plastic deformation phase indicated by a steady decline in gradient toward the ultimate tensile strength (UTS) point, and then specimen break.

Most force transducers consist of a sensing element and an elastic element. The sensing element is the strain gauge, vibrating element or optical sensor, and so on. The elastic element is the bulk medium upon which the sensing element is placed and is designed to operate within the linear region of the stress/strain curve. This ensures that the load cell returns to its zero load position with minimal error.

It is called the elastic element because it is designed so that when a force is applied, it deforms such that the sensing element can report a measurement, and after the applied force

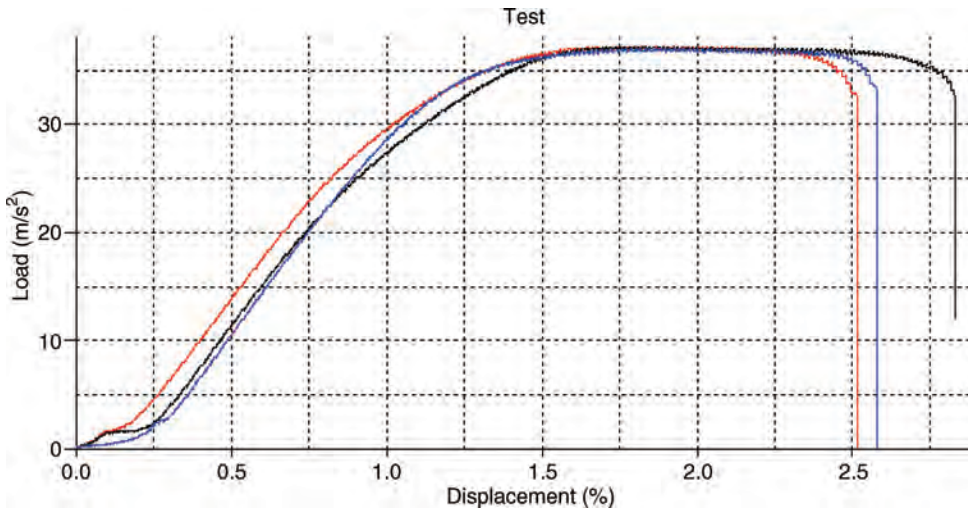


FIGURE 10.1 Typical stress/strain curve for aluminum plate metal (Courtesy of Mecmesin Ltd).

is removed, the element returns to its original dimensions. They can be designed to allow measurement in tension and compression or in compression only. Static torque sensors are also constructed using the same principles, although the designs are more complex. Load cells come in many shapes and sizes, but materials tend to remain the same. Steel and aluminum are the most common and are used for their properties as well as their relative abundance. Choice of material depends largely upon the design criteria: aluminum has better thermal stability, whereas stainless steel has better mechanical stability. Figure 10.2 shows some of the common forms (Institute of Measurement and Control, 1996).

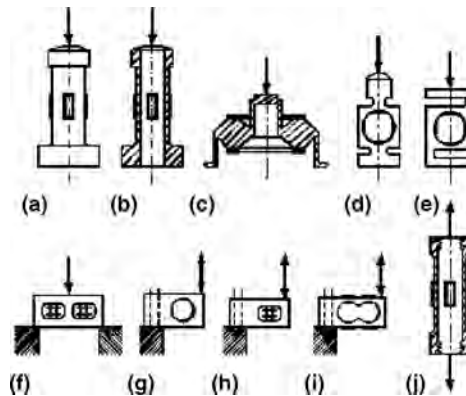


FIGURE 10.2 Load cell types and their common approximate load ranges. (a) Compression on cylinder 50 kN to 50 MN. (b) Compression on cylinder (hollow) 10 kN to 50 MN. (c) Torodial ring 1 kN to 5 MN. (d) Ring 1 kN to 1 MN. (e) S-beam (bending or shear) 200 N to 50 kN. (f) Double-ended beam (simplified) 500 N to 50 kN. (h) Shear beam 1–500 kN. (i) Double-bending beam 100–10 N. (j) Tension cylinder 50 kN to 50 MN.

10.3 UNIVERSAL TESTING MACHINES

Load cells are commonly used in materials and product testing applications. This type of testing requires a method of controlling the sometimes complicated test routines. Universal test machines often fulfill this purpose and are available from many manufacturers.

Universal testers provide a means to hold a sample and place it under a compressive or tensile load. They can be simple motorized machines, whereby the speed and direction are controlled using manual controls, hand operated machines where the operator uses a wheel or pump control to increase the force on the sample, right up to high end computer controlled test stands that can run sophisticated multistage test programs. They are available with hydraulic actuation or electromechanical control using electric motors driving a ball screw arrangement, for a wide range of forces from 500 N up to 30 MN. They will perform complex calculations such as Young's modulus and create and print reports. When used in conjunction with other instrumentation, such as extensometers, even more sophistication is possible. With the ever growing library of international test methods, universal testers are required to provide more functionality and more calculations. Figure 10.3 shows just a few of the range of available universal test machines.

Portable force measurement is catered for by a plethora of hand-held instrumentation, from simple gram gauges to advanced force gauges that allow data storage and transmission as well as peak and trough calculations. Some will allow simple control functions for operating some test stands so that no computer is necessary.

Force measurement in this form is used in all industries, from medical and pharmaceutical to food, automotive, and electronics sectors. Both the finished products and the raw materials they are made of are tested for either quality control or to determine basic material properties. Figure 10.4 shows some examples of force measurement applications in industry, including replacement joint testing, PCB component pull off testing and food testing.

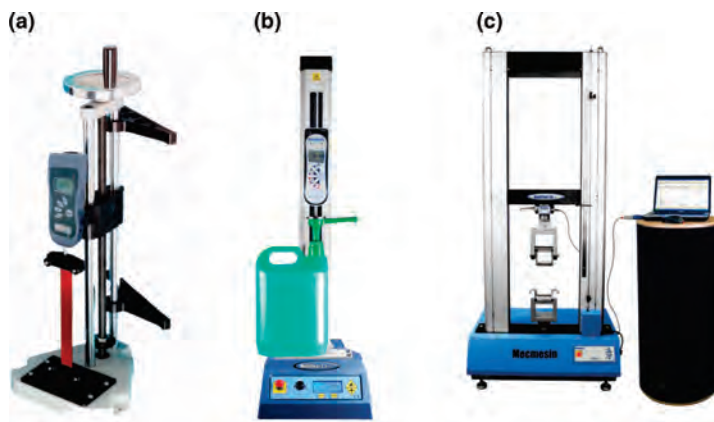


FIGURE 10.3 (a) Manual, (b) motorized, and (c) computerized universal test machines (Courtesy of Mecmesin Ltd).



FIGURE 10.4 Some applications of universal test machines and force measurement (Courtesy of Mecmesin Ltd).

10.4 THE STRAIN GAUGE SENSOR

Although the then Lord Kelvin first reported on the relationship between strain on a metal and its electrical resistance in 1856 (Thomson and William, 1856), it was not until 1936 that a practical implementation of a bonded strain gauge (Simmons, 1942) was realized, by the eccentric Edward E Simmons, Jr. The invention, which consisted of wires glued to the surface of the object whose stress or strain was to be measured, had an enormous impact on the engineering world and was of great importance during the World War II (Dietrich and Jane, 1986).

In 1952, with the advent of etching technology for electrical components, the first strain gauges were developed using etched metal foil instead of wire (Jackson, 1990). The benefits of this technique are that the resistance of the strain gauge can be more accurately and repeatably achieved, and also a much finer pattern can be achieved thus increasing the length of resistor in a smaller space allowing for smaller sensors. This revolutionized the strain gauge industry, and now most strain gauges are of the metal foil type (Figure 10.5). They are manufactured by photo-etching the resistors from a metal coating on a substrate. The substrate has a dual purpose; that of a backing for the strain gauge and as a resistive barrier between the gauge and the item whose strain we want to measure. The gauge on its substrate is adhesively bonded to the component so that strain on the component is transferred efficiently to the gauge. The bonding process is a challenging one given the tolerances that need to be achieved and this is a specialist task, which is highly valued.

As is well known (Young and Freedman), electrical resistance per unit length, L , of a metal wire is inversely proportional to its cross-sectional area:

$$R = \rho L/A \quad (10.5)$$

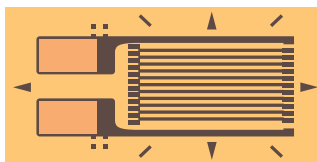


FIGURE 10.5 A typical strain gauge, complete with mounting guide marks.

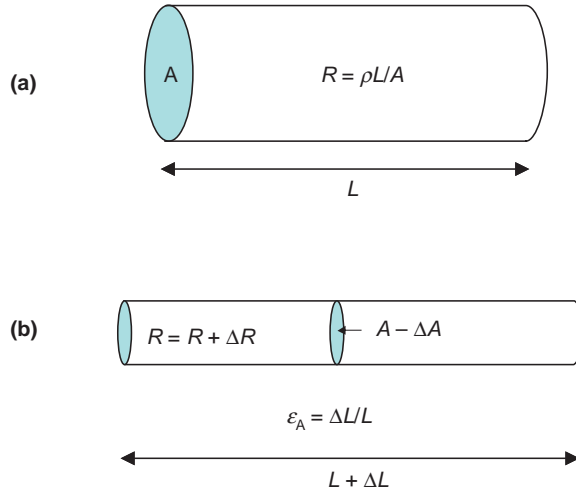


FIGURE 10.6 (a) Metal wire resistor and (b) metal wire resistor under strain.

where R is the resistance, ρ is the resistivity of the material, and A is the cross-sectional area.

Figure 10.6 shows the most basic form of strain gauge, which is a wire of known length, L , and cross-sectional area, A . When an external force is applied to the wire, which causes strain, ϵ (i.e., it is stretched or compressed), its cross-section is also deformed and so a corresponding change in the wire's resistance, R , can be measured.

By differentiating Equation (10.1) and dividing by R , we get

$$\frac{dR}{R} = \frac{d\rho}{\rho} + \frac{dL}{L} - \frac{dA}{A} \quad (10.6)$$

We define axial strain by $\epsilon_a = dL/L$ and transverse strain by $\epsilon_T = dr/r = -\nu\epsilon_a = -\nu(dL/L)$ (Charles Wheatstone, 1843), where ν is the Poisson ratio (the ratio between axial and transverse strain) for the wire material.

The change of radius of the wire under strain is

$$dr = r \left(1 - \nu \frac{dL}{L} \right) \quad (10.7)$$

And so the rate of change of the cross-sectional area is given by

$$\frac{dA}{A} = 1 - (1 + \epsilon_T)^2 = 2\epsilon_T + \epsilon_T^2 \approx 2\epsilon_T \quad (10.8)$$

And when this is substituted back into Equation 10.1, and dividing by axial strain, we get

$$S = \frac{dR/R}{\epsilon_A} = \frac{d\rho/\rho}{\epsilon_A} + (1 + 2\nu) \quad (10.9)$$

S is the sensitivity of the resistor material to strain—the change of resistance versus strain. This is directly analogous to the sensitivity of the strain gauge to strain and when applied; thus, it is called the Gauge factor; a figure of merit for strain gauges. The material that is used to make the strain gauge dictates the gauge factor, the most common being Constantan (copper–nickel alloy), Nichrome (nickel–chromium alloy), and Platinum (normally alloyed with tungsten) (Charles Wheatstone, 1843). For wire and foil strain gauges made from Constantan or Nichrome, the gauge factor is approximately 2, whereas for Pt-W resistors it is approximately 4. Constantan is a common material for strain gauges as it exhibits a linear response over a large range of strain, and it is relatively stable with changing temperature.

Because the change in resistance is very small, a suitably sensitive circuit is required that can measure it, and the most common circuit configuration used in this type of application is illustrated in Figure 10.7, commonly known as a Wheatstone Bridge (Langdon, 1985).

In order to be balanced, $R_1R_3 = R_2R_4$. The simplest form of strain gauge bridge is made by placing a single strain gauge into the circuit in place of R_1 with R_2 , R_3 , and R_4 chosen to balance the bridge—often equal. It can be shown (Charles Wheatstone, 1843) that the system sensitivity in its simplest form ($R_1 = R_g$) is

$$S_s = \frac{r}{1+r} S_g \sqrt{p_g R_g} \quad (10.10)$$

where p_g is the power dissipation of the gauge, which limits the maximum supply voltage, and r is the ratio R_1/R_2 . Of course, where R_2 , R_3 , and R_4 are chosen to be equal to R_g then $r = 1$ although a value of r in the range 3–5 gives circuit efficiencies of between 75 and 83% while allowing a reasonable supply voltage.

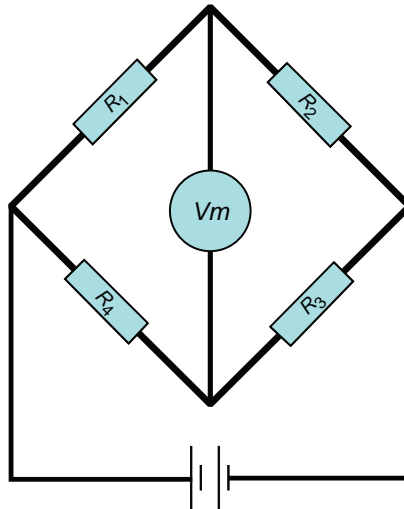


FIGURE 10.7 Classic representation of a Wheatstone Bridge circuit.

Sensitivity varies with different configurations, with the most sensitive being the use of four gauges in the bridge—normally with the gauges in positions R_1 and R_3 oriented to measure strain in one dimension (tension or compression) and R_2 and R_4 in the other—which roughly doubles the sensitivity over the simplest form. It must be remembered that although sensitivity can be increased by attaching more gauges, the cost will increase accordingly due to the extra work involved in bonding the gauges.

So, the Wheatstone bridge, when properly balanced, is well suited to sensing very small changes in the balance of the bridge, and the output of a strain gauge circuit is measured in mV/V , where V is the excitation voltage. Most commonly, strain gauge circuits are designed to give an output of 2, 3, or 4 mV/V . The disadvantage here is the need for amplification of the output, which also increases any noise in the system.

10.4.1 Strain Gauge Circuit Compensation

To an extent, the Wheatstone bridge design lends itself to self-compensation for environmental factors such as temperature by virtue of the balanced nature of the resistors—if all resistors are the same then they will be affected by temperature to the same extent and the net effect is theoretically zero. For more demanding transducer applications, compensation is required for temperature and sometimes zero-balance, where, however well balanced the gauges are, their characteristics cannot be identical and so the balance of the bridge is not exact when under zero load. These things can be compensated for by careful tuning of compensation resistors placed at the vertices of the bridge where the output is read, as shown in Figure 10.8. Placing a well-tuned resistor with high thermal conductivity (commonly copper) at the vertex of R_4 and R_3 can compensate for zero drift due to

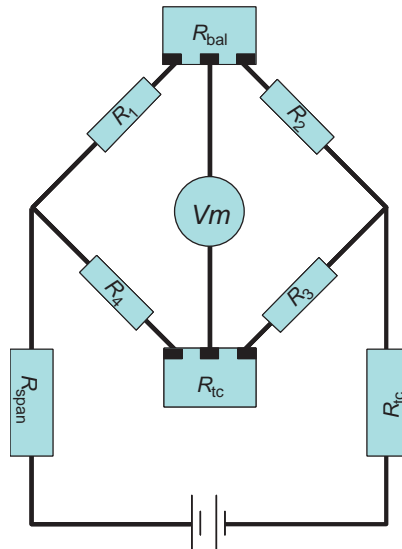


FIGURE 10.8 Strain gauge bridge with compensation for zero balance (R_{bal}), zero temperature compensation (R_{tc}), span adjust (R_{span}), and span temperature compensation (R_{tc}).

temperature, whereas a temperature stable resistor, commonly Constantan, at the vertex of R_1 and R_2 can compensate against bridge imbalance. Span effects can be compensated for by placing resistors either side of the voltage supply.

10.5 RESONANT ELEMENT TRANSDUCERS

Sometimes also known as tuning fork sensors, these devices are used in a wide variety of applications from measuring pressure, to liquid density and from accelerometers to torque sensors (Langdon, 1985). The principle of operation is similar to a guitar string, which, when plucked, resonates with a tone which is relative to the tension on it. In other words, the resonant frequency of the string is a function of the measurand—tension, the string material property, and its thickness and length. The frequency then increases with increasing axial force.

Similarly in a resonant element force transducer, the structure of the device is a vibrating element, which when excited, vibrates with a natural resonant frequency. When a stress or strain is applied to the device the frequency changes proportionally.

The fundamental frequency of a vibrating beam under tension, T is governed by the equation:

$$T \times \frac{d^2 Y}{dX^2} - E \times I \times \frac{d^4 Y}{dX^4} = \rho \times A \times \frac{d^2 Y}{dt^2} \quad (10.11)$$

where I is the second moment of area: $(b \times h^3)/12$, E is the Young's modulus of the material, r is the material density, A is the cross-sectional area of the beam ($b \times h$), b and h are breadth and height of the beam, respectively, and t is the time.

Then, assuming simple harmonic motion and applying the boundary conditions, the natural frequency f_{01} , when axial force $T=0$, is given by

$$f_{T1} = \frac{a^2}{2\pi \times l^2} \times \sqrt{\frac{E \times I}{\rho \times A}} \quad (10.12)$$

where a is the value determined by mode number of vibration $n_v \approx 0.5 \times \pi \times (1 + 2n)$, and l is the beam length.

When the applied tension changes, the natural frequency of the beam under axial force f_{T1} becomes

$$f_{T1} = f_{01} \times \sqrt{1 + \frac{1}{a^3} \times th \times \frac{a}{2} \times \left(a \times th \times \frac{a}{2} - 2 \right) \times \frac{l^2 \times T}{2E \times I}} \quad (10.13)$$

These equations can be used to determine the design of a tuning fork sensor with respect to the motion of the device under strain, the nominal frequency of the device and frequency change at full scale.

An electronic closed loop circuit such as the one illustrated in Figure 10.9 (Cheshmehdoost and Jones, 1993) tracks the output of the device and, by also phase shifting the output by 90° , brings the input and output into phase. This keeps the sensor in continuous oscillation, and as the transducer is loaded the resonant frequency is maintained.

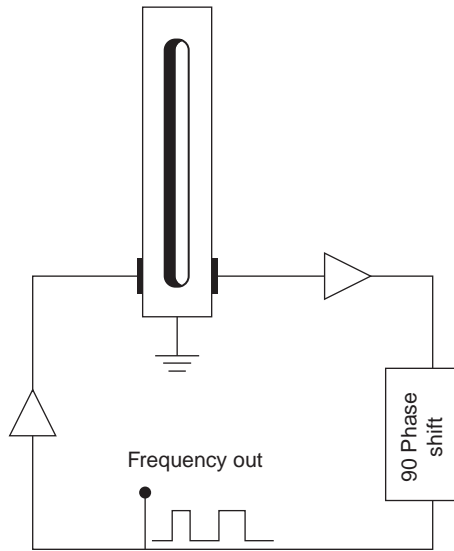


FIGURE 10.9 Illustrating the closed loop circuit that maintains relative oscillation.

Examples of three common forms of resonant element transducer: single, double beam, and triple beam are illustrated in Figure 10.10 (Eernisse et al., 1996):

They are manufactured by conventional means, usually by laser cutting the resonant element from steel and then photo-etching down to the required thickness (Yan et al., 2003). The associated excitation and readout piezoelectric components can then be screen printed on to the element. These are all well-understood manufacturing methods and so volume production can be low, which is attractive commercially.

Figure 10.11 shows a photograph of a triple beam tuning fork design, as represented in Figure 10.10c (Yan et al., 2003).

Because the device has a frequency output, it requires no analogue to digital electronics to read it out. Advantages that are claimed for this type of sensor are

- long-term stability ($\pm 3.5 \times 10^{-5}$ over 12 months),
- easy interface to digital electronics,
- relatively easy manufacturing and cost-effective in quantity.

Their disadvantages include a relatively small operating range and design considerations to retrofit into existing applications.

Tuning fork transducers have mostly been used in torque measurement applications as shown in Figure 10.12 and described in (Intiang, 2007) and in precision scales, but they are becoming more widely used, with a 50 N load cell based on tuning fork technology having been recently evaluated using the 500 N force standard machine at the National Metrology Institute of Japan (Toshiyuki et al., 2006). Although they show some good characteristics such as low hysteresis and good long-term stability, they are large when compared with strain gauge load cells and have only comparable repeatability.

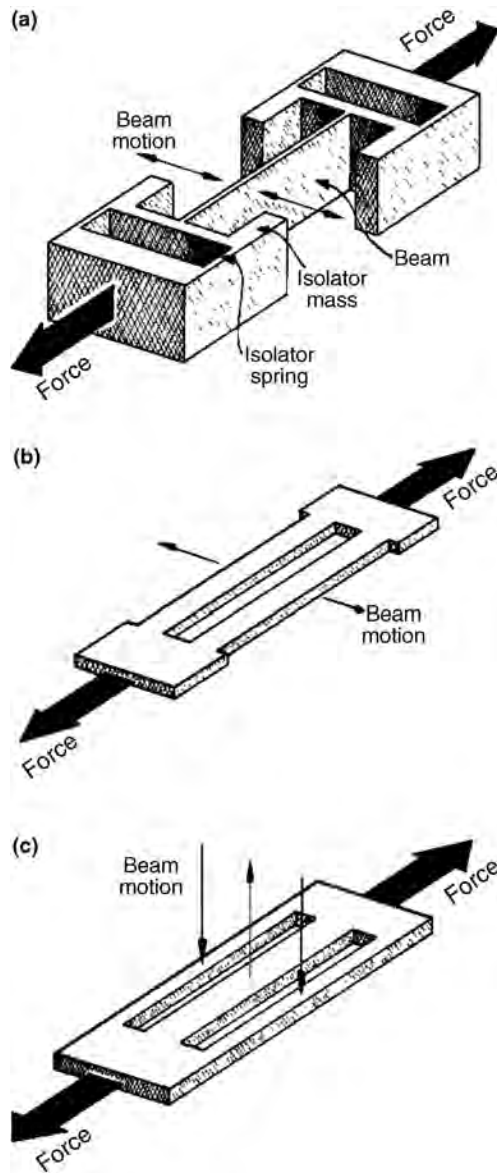


FIGURE 10.10 Three different types of tuning fork sensor: (a) single, (b) double, and (c) triple beam.



FIGURE 10.11 A typical triple beam tuning fork sensor.

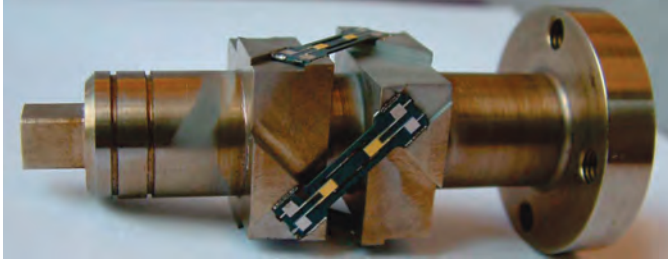


FIGURE 10.12 Resonant sensors mounted in a torque measurement arrangement (Forcesenssys Ltd., 2007).

10.6 SURFACE ACOUSTIC WAVE TRANSDUCERS

The existence of surface waves, or Rayleigh waves, was first proposed theoretically in 1885 by, as their name suggests, the then Lord Rayleigh (John William Strutt) (Rayleigh, 1885) who predicted their presence during earthquakes. However, it would be some 35 years before the development of seismological measurements, which confirmed his theories.

When an earthquake occurs, its energy is propagated in waves of three main types: The primary wave, or P-wave; the secondary wave, or S-wave; and surface waves in the form of Love and Rayleigh waves. The P-wave is analogous to an acoustic, or compression wave. It travels faster than the other two wave types and so is normally detected first. The S-wave is a sinusoid with associated amplitude and frequency. A surface wave is a hybrid of P- and S-waves, which propagates at the interface between two materials, and whose amplitude decreases exponentially with depth beneath—in the context of an earthquake—the surface of the Earth. Rayleigh waves are those surface waves that propagate with particle motion perpendicular in the vertical direction to the direction of propagation. Love waves have particle motion in the transverse direction and perpendicular to the direction of propagation. Diagrams illustrating these wave types can be seen in Figure 10.13 (Ikelle and Amundsen, 2005).

In 1965, the practical implementation of a surface acoustic wave (SAW) on an elastic substrate (quartz) was realized by Voltmer and White (1965) and this paved the way to the subsequent extensive use of SAW devices in telecommunications as band-pass filters and delay components (Campbell, 1989).

SAW devices such as the one in Figure 10.14 consist of an excitation construction made up of “fingers,” spaced at half the resonant wavelength of the desired wave, on a piezoelectric substrate. There are then receivers placed on the substrate, which detects the arrival of SAW’s and the output frequency is then transmitted as a quasi-digital signal to the processing electronics. The frequency of the SAW is given by (Hauptmann, 1991)

$$f = \frac{(2n\pi - \varphi_{el}) \times v_{SAW}}{L} \quad (10.14)$$

where φ_{el} = electrical phase shift of the amplifier circuit, n = constant determined by electrode design, v_{SAW} = velocity of the SAW, and L = distance between transmitter and receiver structures.

This means that frequency is directly dependent on wave velocity and L , which enables development of SAW devices to measure a range of parameters including

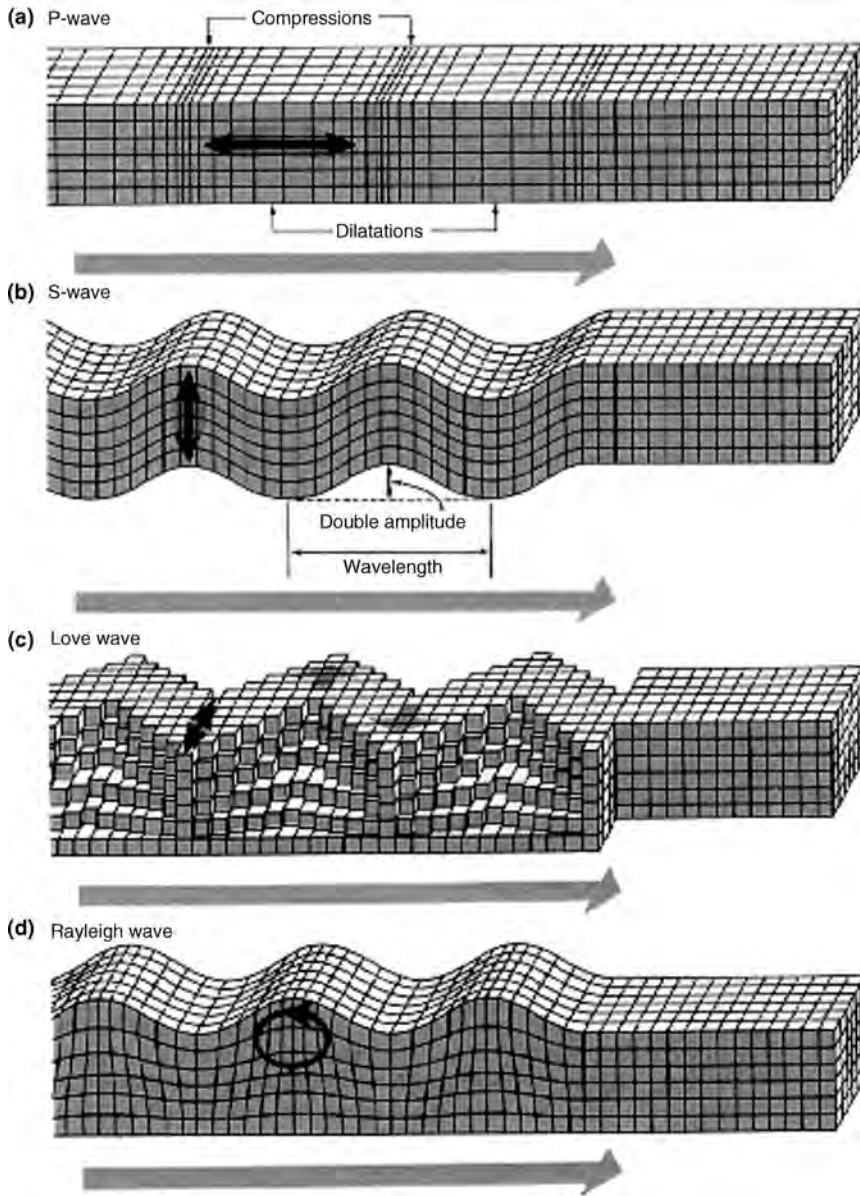


FIGURE 10.13 Illustrating the different wave types observed during earthquakes.

acoustic, temperature, and dimensional changes (and hence force and torque) to name but a few.

Since then the use of SAW transducers has broadened into other applications, which include temperature (Wold et al., 1991) and pressure measurement (Cullen and Reeder, 1975; Cullen and Montress, 1980), and of course, force measurement—in particular torque measurement.

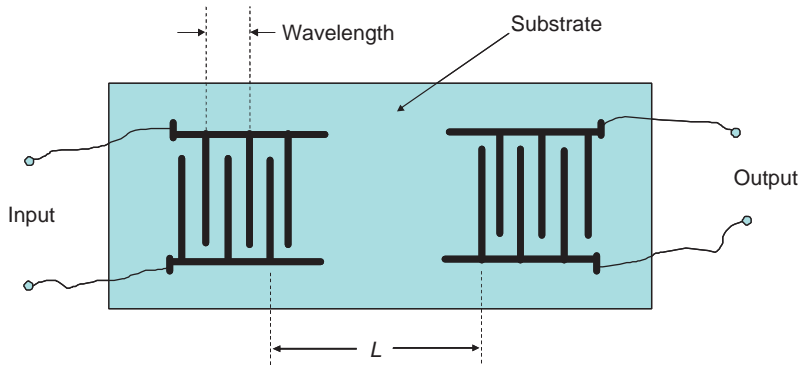


FIGURE 10.14 The layout of a typical SAW transducer.

In 1990, the first noncontact torque transducer based on a SAW device was patented (Lonsdale and Lonsdale, 1990) and since then the noncontact torque measurement industry has been served mostly by this technology. Applications include Formula 1 vehicle drive train monitoring as well as uses in torque monitoring for regenerative braking systems (Kalinin et al., 2010).

If the SAW device is rigidly mounted to a flat spot on a shaft, and the shaft experiences a torque, this torque will stress the sensor and turn it into a wireless passive lightweight torque sensor. As the shaft is rotated one way, the SAW torque sensor is placed in tension. As the shaft is rotated the other way, it is placed in compression. For practical applications, two SAW torque sensors are utilized such that their center lines are at right angles (Figure 10.15) (Lonsdale and Lonsdale, 1990). With this system, when one sensor is in compression, the other sensor is in tension. Since both sensors are exposed to the same temperature, the combination of the two signals will act to minimize temperature drift effects.

When compared with other torque sensors, including resistive strain gauges, optical transducers, and torsion bars, SAW torque sensors offer lower cost, higher reliability, and wireless operation.

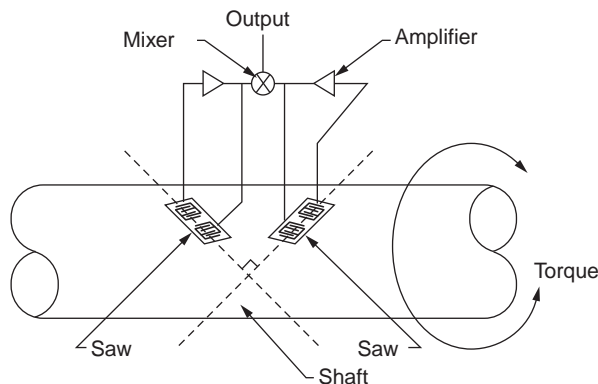


FIGURE 10.15 Configuration of SAW devices on a shaft to measure torque.

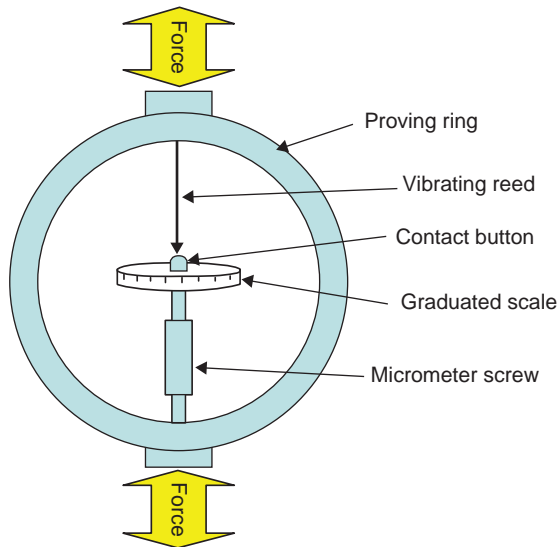


FIGURE 10.16 Schematic of a proving ring with manually adjusted micrometer screw.

10.7 DYNAMOMETERS

Strictly speaking, a dynamometer is any force transducer that uses the elastic nature of a material to determine the force causing deformation. This includes load cells, pressure transducers, and so on. A particular type of dynamometer that is encountered in the field of force measurement is a ring dynamometer, or “proving ring.”

The common form, shown in Figure 10.16, utilizes a manually adjusted micrometer readout to measure the diameter of the ring under tension and compression. In this configuration, a vibrating reed is set in motion by tapping it with a pencil. The micrometer is then adjusted to move the contact button until it is just in contact with the reed, which will make an audible buzzing noise. The graduated scale is then read off, which will indicate the force applied to the ring. This method was developed in 1946 (Bruce et al., 1946), and remains to this day as a proven and reliable method of measuring force with relatively low uncertainty. If a digital readout is required, then the micrometer can be replaced by strain gauges or a LVDT—a displacement transducer.

Proving rings are used commonly to measure soil compaction but also as calibrating instruments.

10.8 OPTICAL FORCE TRANSDUCERS

There are two main types of optical strain gauge, but both are spectroscopic in nature. One utilizes a change in the arms of an interferometer arrangement of fiber optics. In another, a diffraction grating is incorporated into the fiber with the spacing of the grating being altered by strain upon the fiber through the substrate material. This is otherwise known as a fiber Bragg grating sensor or FBG sensor, a diagram of which can be seen in Figure 10.17.

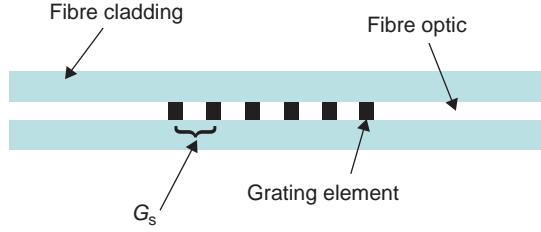


FIGURE 10.17 Representation of a FBG sensor.

The Bragg condition for constructive interference is a product of the theory of X-ray diffraction by crystals. It is named after father and son physicists, Sir William and Lawrence Bragg who jointly won the Nobel Prize for physics in 1915, and is utilized in fiber optic sensors that measure, among other things, strain. The condition states (Born and Wolf, 1999)

$$2d \times \sin\psi = n\lambda \quad (10.15)$$

where λ is the wavelength of light, n is an integer, ψ is the angle of incidence with respect to the crystal surface, and $2d$ is the path length difference between two incident light rays.

When this is applied to fiber optic, the isolated wavelength, that is, the one of interest is the one reflected by the grating elements and the equation slightly changes (Yun-Jiang, 1997)

$$\lambda_B = 2\eta \times G_s \quad (10.16)$$

where λ_B is the “Bragg” wavelength, η is the refractive index of the optical fiber, and G_s is the spacing of the grating elements in the fiber.

In reality, there are thousands of elements in a FBG sensor, which is manufactured by “writing” the elements into a length of Germanium doped optical fiber. The “writing” is done by exposing the fiber to a spatial pattern of ultraviolet light in the region of 244–248 nm. This fabrication process is based on the photorefractive effect, which was observed in Ge-doped optical fibers in 1978 by Hill et al. (1978).

When a FBG sensor is subjected to a strain, the spacing between the grating elements is increased, and the Bragg wavelength is altered according to the equation (Manfred Kreuzer)

$$\Delta\lambda = \lambda_B(k \times \varepsilon + \alpha_\delta \times \Delta T) \quad (10.17)$$

where $\Delta\lambda$ is the change in Bragg wavelength, $k = 1 - p$, (p is the photo-elastic coefficient of the fiber), α_δ is the refractive index temperature coefficient, and ΔT is the change in temperature.

And so, provided temperature is compensated for (fiber-optic sensors are highly temperature dependent and so are used as temperature sensors also) the change in wavelength can be used as a measurement of the strain on the fiber.

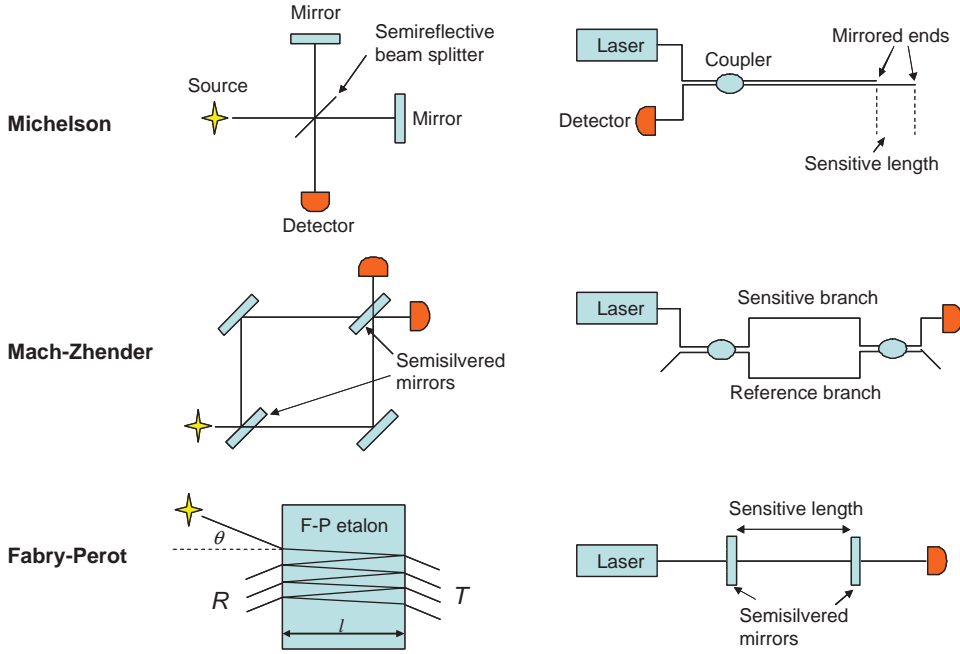


FIGURE 10.18 The different interferometer types and their fiber-optic counterparts.

Fiber-optic interferometers such as those illustrated in Figure 10.18 constitute the most sensitive of the fiber-optic force sensors and may be based on the many available designs including Michelson, Mach-Zehnder, and Fabry-Perot (Tatti et al., 1997).

In an interferometer, interference between two beams of light is described by the following equation:

$$\Delta\Phi = \frac{2\pi}{\lambda} n \times \Delta l \quad (10.18)$$

where $\Delta\Phi$ is the phase change of the resultant signal, n is the refractive index, λ is the wavelength, and Δl is the relative path change of the two beams.

In the case of a fiber-optic strain sensor, the path length is introduced when the paths of the interferometer are subjected to a differential strain. When this happens, two changes occur: its length changes and also the refractive index.

$$\Delta l = \varepsilon_1 \times l_0 \quad (10.19)$$

$$n = n_0 \left\{ 1 - \frac{1}{4} n_0^2 [2 \times P_{12} \varepsilon_1 + (P_{11} + P_{12})(\varepsilon_2 + \varepsilon_3)] \right\} \quad (10.20)$$

where l_0 and n_0 are the initial (unstrained) length and refractive index, respectively, P_{xy} are coefficients that depend upon the glass material, and ε_1 , ε_2 , and ε_3 are strain components acting on the fiber.

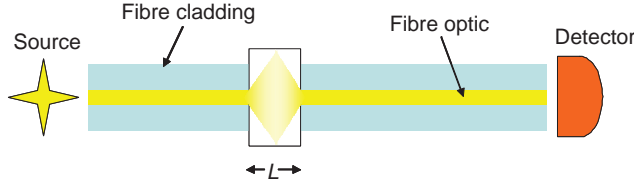


FIGURE 10.19 Representation of a fiber-optic intensity monitoring force sensor.

These, when substituted into Equation (10.18), and if it is assumed that $\varepsilon_2 = \varepsilon_3 = -\nu_f \varepsilon_1$, gives the phase change due to a strain ε_1 on the fiber.

$$\Delta\Phi(\varepsilon_1) = \frac{2\pi}{\lambda} l_0 \varepsilon_1 n_0 \left\{ 1 - \frac{1}{2} n_0^2 [P_{12} \varepsilon_1 - \nu_f \varepsilon_1 (P_{11} + P_{12})] \right\} \quad (10.21)$$

where ν_f is Poissons ratio for the fiber.

The Intensity at the detector of the resultant signal is given by

$$I = \alpha [1 + V \cos \Delta\Phi(\varepsilon_1)] \quad (10.22)$$

where $\alpha = E^2$ (E is the optical electric field) and V is the fringe visibility.

Hence, intensity at the detector can be related to strain directly.

Other fiber-optic sensors include intensity monitoring sensors (Gholamzadeh and Nabovati, 2008) (Figure 10.19) where a cavity length is altered by the applied load.

An increase in cavity length induces an intensity drop and a vice versa. A detector monitors the output intensity and this is then converted to a strain. Although simple in their design, they suffer from inefficiencies due to the coupling, alignment and bending.

Fiber-optic force sensors can be very sensitive and are also immune to electromagnetic interference, and because of this find their applications in challenging environments, such as the measurement of strain in overhead power cables.

10.9 MAGNETO-ELASTIC TRANSDUCERS

In 1842, Joule (1847) discovered that the properties (by which he meant its dimensions) of ferromagnetic material are coupled to its magnetic field flux. It is known as the Joule effect, but more commonly called Magnetostriction. He observed that if the material magnetic flux was increased, its length changed and as its length changed, so did its transverse dimension, thus keeping the material volume (approximately) constant. There is a reciprocal effect, known as the Villary effect, whereby subjecting a magnetic material to stress or strain will alter its magnetic flux. There are other magnetoelastic phenomena, such as the Matteucci effect, in which a material exhibits a change in an induced helical magnetic flux when subjected to a torque. Both of these principles are used in force transducers.

Nearly, all ferromagnetic materials will exhibit some magnetostrictive behavior, although much current research involves amorphous alloys (Hauptmann, 1991)—the so-called giant magnetoelastic materials, such as Terfenol-D—an alloy of terbium, iron, and dysprosium because they give such high values of magnetostriction. An illustration of the effect is shown in Figure 10.20 (Ekreem et al., 2007).

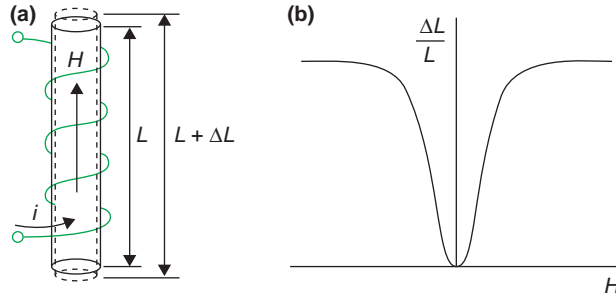


FIGURE 10.20 The Joule effect (a) The changes in shape in response to the magnetic field H . (b) The relationship between $\Delta L/L$ and H .

In Figure 10.20, a rod of a magnetic material with length, L , is surrounded by a coil of wire carrying an electrical current so that a magnetic field, H , is produced along the rod. When current is flowing, the length of the rod will increase by a small amount ΔL . The strain $\Delta L/L$ is called the magnetostriction, denoted by λ , and normally measured in ppm. The graph in Figure 10.20b shows that as the magnetic field, H , increases, λ also increases but not indefinitely. There is a saturation point, normally denoted λ_{sat} whereby no matter how intense the magnetic field, the magnetostriction does not change. Also to be noted is that the rod exhibits an increase of length regardless of the sign of the magnetic field increase.

Magneto-elastic force sensors are used in many industrial applications, including as torque sensors in the automotive industry (Dan Mihai, 2011), as sensors and transducers in railway engines (Bienkowski et al., 2010) and in civil engineering applications (Ausanio et al., 2005) to name but a few.

10.10 FORCE BALANCE TRANSDUCERS

The force balance utilizes one of the oldest forms of weight measurement, the beam balance. If we have ever been weighed by our family doctor, we have seen the beam balance in action. In the balance, an unknown mass is balanced against a known mass, with the distance from the fulcrum giving the value of the weight given by the known mass.

In a modern force balance, the deflection caused by the applied, unknown, mass is used to regulate a current input to an actuator that acts to balance the applied force. This feedback system works to retain the balance in equilibrium, and the current applied to the actuator is directly related to the force applied by the unknown input.

Force balances are widely used as accelerometers in seismographs, aircraft and drilling platform stabilization systems, and so on (Dan Mihai, 2011).

10.11 FORCE TRANSDUCER CHARACTERISTICS

Any force transducer will be supplied with a quoted specification, which places quantitative values on various aspects of the characteristics of a force measurement system. Table 10.2 lists the typical values in a 1000 N (220 lbf) strain gauge load cell.

So what does all this mean? Taking these items in turn.

TABLE 10.2 Typical Load Cell Characteristics

Capacity	1000 N
Output	2 mV/V
Creep	$\pm 0.01\%$ FS
Non linearity	$\pm 0.05\%$ FS
Hysteresis	$\pm 0.05\%$ FS
Repeatability	$\pm 0.05\%$ FS
Temperature coefficient	$\pm 0.025\%$ FS
Temperature range	40°C to +60°C
Accuracy	$\pm 0.1\%$
Combined uncertainty	$\pm 0.02\%$

10.11.1 Capacity

This is the maximum force that can be measured reliably using this particular load cell. However, commercially available load cells may have up to 200% overload capacity whereby the load cell can withstand an overload without damage, although it is advisable that load cells that have been overloaded be recalibrated. As with most transducers, it is inadvisable to use load cells within the first or last 10% of their design range. Table 10.1 shows that there is considerable crossover between device types and their operable ranges, but some will be much better suited to some applications than others, and so care should be taken to specify the device to the application accordingly.

10.11.2 Output

The output of a strain gauge load cell is expressed as mV/V, with the specified output being that at the rated capacity. So, with a specified output of 2 mV/V, the actual output voltage from a strain gauge load cell with a 10 V DC excitation (supply) voltage will be 20 mV.

Figure 10.21 illustrates the various concepts of nonlinearity and hysteresis as well as rated output and capacity. Load cells by their spring-like nature are linear in their

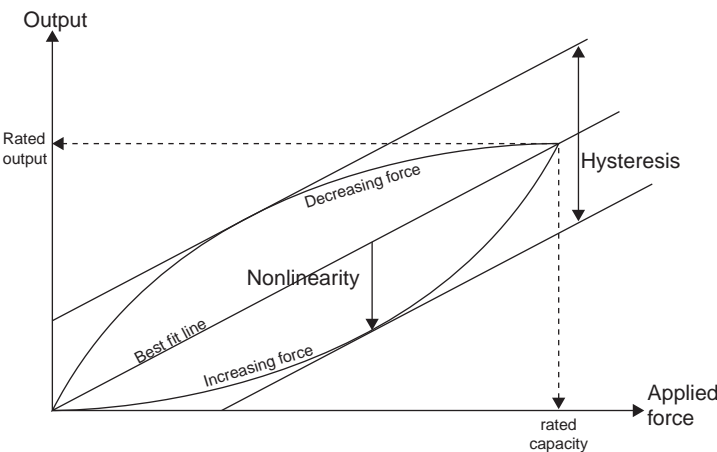


FIGURE 10.21 Illustration of a typical load cell response curve.

response but with manufacturing inconsistencies and engineering tolerances, there may be some deviation from true linear behavior. When adjusting a load cell prior to calibration, a straight-line fit is made to the device output, and the maximum deviation from this line is calculated and reported as the device nonlinearity. Hysteresis is the maximum difference in output with increasing and decreasing input. Often, load cell manufacturers or suppliers will specify a *combined error*, which is a combination of nonlinearity and hysteresis.

10.11.3 Repeatability

Repeatability is the standard deviation of output at any given repeated input, while keeping the measurement conditions the same. In a normal calibration, the repeatability is calculated for all points in the calibration and the specified repeatability is normally the maximum standard deviation. Repeatability should not be confused with reproducibility, which is the measurement of the output variation when the measurement conditions are different each time. This should not be confused with *reproducibility*, which is sometimes also quoted, but is the standard deviation of output at a given input but under differing conditions.

10.11.4 Creep

This is the extent of output change over time after a step change input is given, normally measured over a 30 s.

10.11.5 Temperature Coefficient

This is a measurement of the load cell susceptibility to temperature change, normally stated as %FSD per degree Celsius.

10.11.6 Accuracy

This is the statement of closeness of the output to the input quantity, taking into account uncertainty of measurement. This value is determined during calibration. Accuracy should not be confused with *precision* or *resolution*, these being closeness of repeated output values and the minimum observable output change, respectively.

10.12 CALIBRATION

The importance of having traceable calibration records for your force transducer is not to be underestimated. All calibrated load cells can be traced back to primary mass standards, which are in turn directly traceable to the International Prototype kilogram (IPK), Figure 10.22.

The IPK and its six sister copies are stored at the International Bureau of Weights and Measures in an environmentally monitored vault in Sèvres on the outskirts of Paris. Official copies of the IPK, which are measured against the IPK every 50 years, are held by other nations and constitute the national standard kg. Masses used in calibration laboratories to calibrate other masses are calibrated against the national standard. Those masses are called primary standards and are used to calibrate the so-called secondary standards,



FIGURE 10.22 A photograph of the IPK in Paris.

which may be either masses or very accurate force transducers. The key to any acceptable calibration is traceability to the primary standard. That is to say that a documentation trail can be followed back to the national primary standard.

Of course, it is not enough to know that your mass is the equivalent of a kilogram when measuring force: Equation 10.2 states $F = mg$, where g is the acceleration due to gravity. So, to measure static forces on Earth, we need to know what the acceleration due to gravity is at our location. Because g decreases radially from the center of the Earth according to the inverse square law, the acceleration due to gravity at the Earth's surface very much depends on the altitude and latitude of your measurement position according to the equation (Boynnton, 2001)

$$g = 9.80613 (1 - 0.0026325 \cos 2L) (1 - 3.92 \cdot 10^{-7} H) \quad (10.23)$$

where L is the latitude of the measurement position in degrees and H is the altitude, or height.

The force applied by an adjusted 1 kg mass measured with a force transducer in a laboratory in Mexico local g is 9.77954 ms^{-2} , would be approximately 0.4% different from an identical measurement made in Oslo local g is 9.825 ms^{-2} .

Load cells are calibrated using a force application system, which can be as simple as a calibrated mass hung by a hook, normally for smaller force ranges for health and safety reasons, or a so-called weight stack, which may be manually or computer controlled,

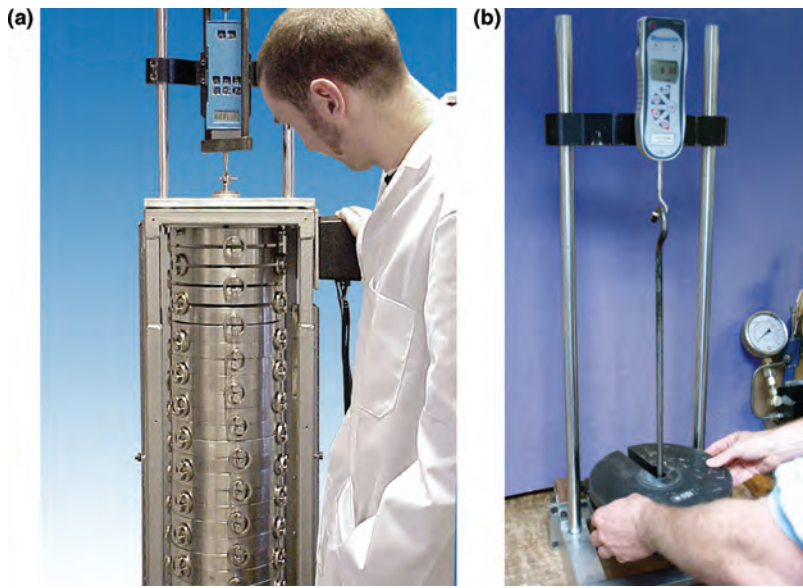


FIGURE 10.23 (a) A small 2500 N weight stack and (b) a basic frame for hanging calibrated masses onto a force gauge (Courtesy of Mecmesin Ltd).

where a set of calibrated masses can be applied sequentially to measure a load cell's output at regular intervals across its range. Weight stacks can range from small (<1 kN) through to the very large (>1 MN). The largest weight stack in Europe has a capacity of 1.2 MN and is at the National Physical Laboratory in London, England (Figure 10.23). The largest weight stack in the world resides in the United States and is rated at 4.45 MN (1000,000 lbf)!

10.12.1 Uncertainty

There are many texts (International Organization for Standardization, 1995; M3003, M2007; European co-operation for Accreditation, 1999) that cover uncertainty of measurement in much more detail than is given here, but we will give a brief overview.

In all measurements, there is never absolute certainty of the result. However, we can estimate the level of uncertainty in a measurement in order to communicate the likelihood of a result being accurate. We can also design our measurement systems in such a way as to minimize uncertainty. This is especially important in calibration, because whoever is providing the calibration service is making a statement about a device's accuracy that a customer will use in their analysis of the measurements they make. Given that some applications are safety critical it is vitally important that the customer is provided with accurate estimations of calibration measurement uncertainty. Using masses calibrated against primary standards ensures that the uncertainties related to the masses are minimized, but there are many other contributing factors.

Temperature: Because load cells (and the systems that hold them) are made of metal (aluminum or steel mostly), temperature will have an effect on performance. Temperature stability during calibration is therefore vital.

Humidity: Humidity affects the buoyancy of the air, thereby affecting the applied force, and so should be kept as constant as possible during calibration.

Gravity: The local value for gravitational acceleration (g) should be taken into account when calibrating and using masses. If a calibrated mass is taken elsewhere to be used to verify a measurement, then a simple correction factor should be calculated for the new location depending on its value for g . An example follows that illustrates this.

For example, say we want to calibrate a device at our laboratory using a mass that applies a force of 1000 N. Previously, g at our location has been calculated to be 9.81 ms^{-2} . A mass is then manufactured and adjusted to our specification and it is delivered with its certificate confirming traceability to the primary standard, so that if an instrument is then calibrated in our lab using the mass, we will know that exactly 1000 N (with a small uncertainty) is being applied to our device. The instrument is then sold to a customer in a different location with a local g value of 9.8120000 and 12 months later the customer requires a recalibration at their premises. The following options are presented: We could purchase a mass that is adjusted to the g value at the customer premises, which could be costly and of limited use, or we could use our own adjusted masses and apply a correction factor taking into account the g value where the customer is. The correction factor would be evaluated as follows.

A mass is calibrated and adjusted to apply a specific force according to Equation 10.2, so if that mass is transported elsewhere and applied to an instrument previously calibrated in the laboratory, the resultant measurement will read

$$F' = m(g + \Delta g) \quad (10.24)$$

Taking m out of the equation

$$F' = F \left(1 + \frac{\Delta g}{g} \right) \quad (10.25)$$

And so

$$F' \left(1 + \frac{\Delta g}{g} \right)^{-1} = F \quad (10.26)$$

Substituting the g values from the example, a correction factor of 0.9997962 should be used to correct the displayed force. Without using g correction, at 1000 N the error would be approximately 0.2 N.

For the 1.2 MN calibration stack in Figure 10.24, in order to reduce uncertainties associated with the force application it is constructed on top of 40 m piles for stability and each mass in the stack is trimmed to its own value of g depending upon its altitude in the stack!

In minimizing uncertainties, providing the correct mechanical linkage is also vital, that is, the means to connect your load cell to the applied force. Because strain gauges are positioned very carefully on an elastic element of a specific shape, which is designed to measure force in a specific direction, and which as a whole is calibrated to measure axially, any off axis loading (i.e., side-loading) will result in an erroneous

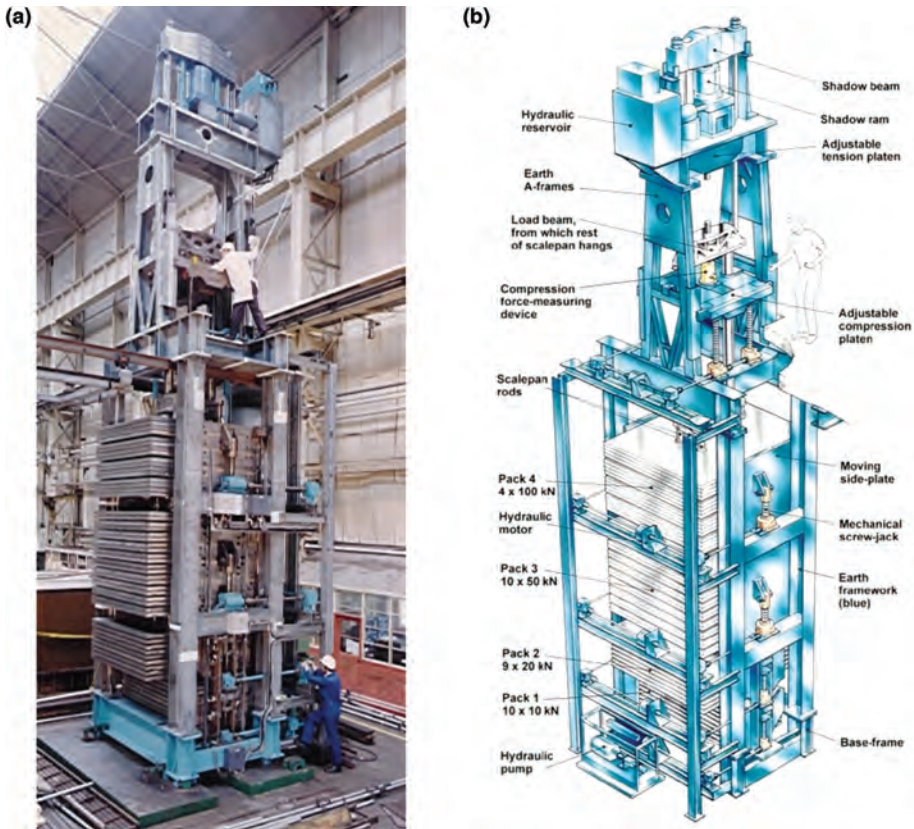


FIGURE 10.24 (a) Photograph of the construction of the 1.2 MN weight stack at NPL in London. (b) a schematic of the design. (Both courtesy of NPL, Teddington, London, UK.)

measurement. For this reason, it is vital that when using the load cell that it is mounted correctly, as incorrect fixturing can cause off-axis loading that will introduce errors, sometimes significant. Axial alignment of force vectors is easier in tension than in compression, because in tension, provided that no part of the system is rigidly fixed, the system will have a tendency to self align. In compression though, it is easy to end up in a situation where you are trying to “push a rope” so to speak, and in doing so can introduce significant alignment errors, and so a reliable method of linking the system components needs to be devised.

10.12.1.1 Other Equipment If other equipment is used in the calibration, for example voltmeters, thermometers, and so on, then the uncertainties of these instruments should be considered in the uncertainty budget.

10.12.1.2 Operator Inconsistency Finally, in a laboratory with more than one operator, results of calibrations can vary depending on the operator. Laboratory procedures may be in place that minimize the uncertainties introduced when different operators perform the same function, but there may be random errors introduced that should be considered.

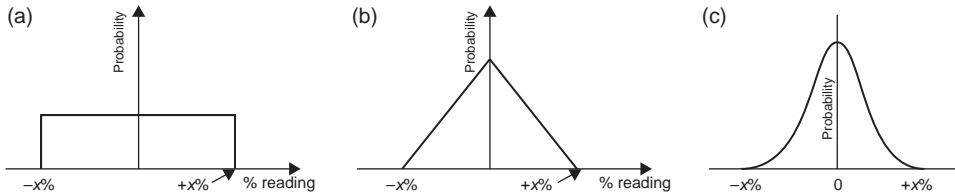


FIGURE 10.25 (a) Rectangular, (b) triangular, and (c) normal probability distribution.

The list could go on, but the contributions made become smaller until they are no longer significant—it is also important to retain a sense of proportion! While it may be commendable to calculate the uncertainty contribution of a given item to the eighth decimal place, this is hardly significant to a device with a required accuracy of 0.5%.

When preparing your uncertainty budget, it is important to note that, because uncertainties are not precisely known they normally quoted as $\pm x\%$ along with a confidence limit. We must, therefore, understand the probability distribution associated with it. Taking the contributing factor of resolution as an example: If, say, a force gauge is reading a nominal value of 1.55 N but the last significant digit of the display is fluctuating between 1.54 and 1.56 N then, because we only know the upper and lower bounds of the error, the true value could be anywhere between these limits with an equal probability of any one true value in that range. If we plotted the probability distribution it would look like the diagram in Figure 10.25a).

The uncertainty of this type of contributing factor has what we call a rectangular distribution. A combination of two rectangle distributions gives, by the process of convolution, a triangle distribution as shown in Figure 10.25b. Triangular distributions are rare but could apply to the process of interpolation between two known values for example, and this is unlikely in the case of an uncertainty budget of a measurement process. If many probability distributions are combined or where random factors are involved, then the result will be a normal distribution—Figure 10.25c. Normal distributions are assigned to uncertainties given in calibration certificates, for example, the “calibration of adjusted masses” entry in Table 10.3. In force measurement, the majority of uncertainty budget contributors have rectangle or normal distributions.

Once all contributions have been accounted for and their values calculated, a divisor is applied, which is dependent upon the probability distribution function. Uncertainties are combined by quadrature addition, that is

$$\text{Combined}_U = \sqrt{\sum u^2} \quad (10.27)$$

In order to quote a confidence level of 95%, the combined uncertainty is multiplied by a factor of 2. If a value, $k = 3$ were to be used, the confidence level would rise to 99.7% but this is rarely used in force measurement. A typical uncertainty analysis table for a dead-weight frame such as the ones in Figure 10.19 is shown in Table 10.3.

Looking at the table, we can see the contributors that have been considered in the overall uncertainty budget along with their probability distribution functions. There was one contributor in this particular analysis that was not included, which was the effect of temperature on the deadweight frame. Because of thermal fluctuations, the metal frame will

TABLE 10.3 A Typical Uncertainty Budget for a Deadweight Calibration System

Source of Uncertainty	Value $\pm\%$	Distribution	Divisor	c_i	u_{iT}
Static frame alignment	0.001	Rectangle	1.73	1	0.00058
Frame flex under load	0.0001	Rectangle	1.73	1	0.000058
Calibration of adjusted masses	0.01	Normal	2	1	0.005
Drift of adjusted masses	0.01	Rectangle	1.73	1	0.0058
Random	0.001	Normal	2	1	0.0005
Combined uncertainty normal		Normal			0.011
Expanded uncertainty		Normal ($k = 2$)			0.022

(Courtesy of Mecmesin Ltd)

undergo a dimensional change due to expansion and contraction. After analysis it was seen that the overall effect on the application of mass was approximately 0.0000001% and so was not included.

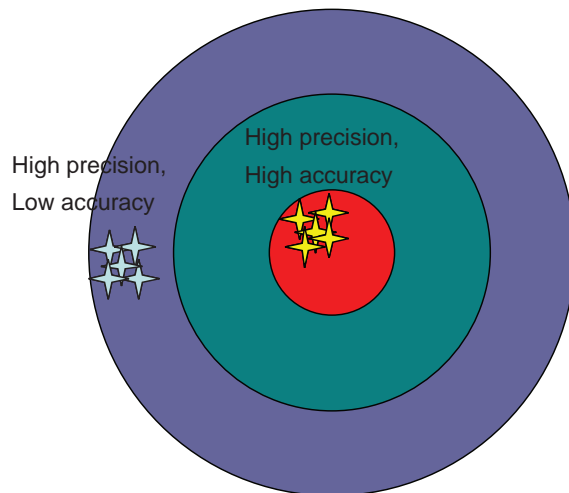
10.13 CONCLUSION

There are many ways to measure force and torque, and this chapter has given a review of but a few of the more common methods, along with an overview of calibration and traceability considerations. Which method is used will depend on many factors including range, application, environment, life requirement of the sensor, instrumentation availability, and, of course, cost.

GLOSSARY OF TERMS

Accuracy

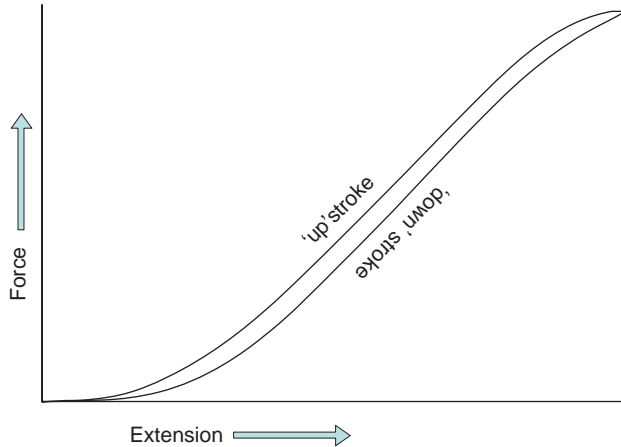
This is an expression of how close a measurement is to the actual value. Not to be confused with precision.



Analogue Output	An analogue output is a continuous signal, from a transducer, given usually as a voltage or a current. The output from a load cell is an analogue one but it is put through an analogue to digital converter so that it can be manipulated and stored in a useful form for computers.
Axial Strain	Axial, meaning “on axis.” This is the extent of sample deformation in the direction in which the force is applied.
Band Pass	In filtering, a filter that allows a given range of frequencies to pass unhindered, while frequencies above and below the filter band limits are blocked.
(Load Cell) Bridge Resistance	The nominal electrical resistance of the load cell circuit measured at the excitation connections of the load cell with zero load applied to the load cell and the output connections open circuit. Expressed in ohms (Ω).
Calibration	The comparison of transducer output against input quantities that are traceable to international standards. Calibration would normally be carried out by a recognized authority and should result in the issue of a calibration certificate. If there is any change in transducer performance or damage to the transducer or if there is any unusual behavior exhibited by the transducer, it should be recalibrated.
Capacity (of Load Cell)	The maximum axial load a load cell is designed to measure within its specifications.
Compression	The application of a force, the end result of which is to reduce the sample height. Applying pressure to a sample in order to deform or flatten it.
Creep	When a change in load is applied to a specimen, the new stress and/or strain reading can drift by a small amount over time. This is termed creep.
Deflection	The degree to which a structural element is displaced under a load. In Mecmesin systems, this is measured from the crosshead position as standard, but where more accurate measurement of deflection is required an extensometer should be used.
Deformation	A change in the dimensions of a material under stress or strain.
Displacement	The total driven movement of a test stand. Linear displacement is usually measured in millimeters or inches, whereas rotary displacement is usually measured in revolutions (revs) or degrees. Note that displacement may differ from specimen deformation.
Drift	The change in transducer output when zero input is given. Not to be confused with creep, which is the change in transducer output under a load.
Ductility	The amount of plastic deformation exhibited by a material before it ruptures.

Eccentricity	The difference between the direction of load application to the true axial direction.
Elastic/Elasticity	A material is said to be elastic if it deforms under stress (e.g., external forces), but then returns to its original shape when the stress is removed. The amount of deformation is called the strain. Elasticity is the ability of a material to return to its original dimensions subsequent to the application of stress or strain. The linear portion of the stress/strain curve of a Hookean material is called the elastic region because if applied force is removed within this region, the material will return to its original shape and size.
Elastic Limit	This is the stress value during a tensile or compression test after which the sample suffers permanent deformation. For Hookean materials, it is roughly equal to the limit of proportionality (LOP).
Energy	The work done on a sample during a tensile or compression test. It is calculated as the area under the stress/strain curve and is measured in Joules, the SI derived unit for energy.
Equilibrium	Achieved when the sum of all forces acting upon a body is zero.
Extension	The amount by which an object is increased in length.
Extensometer	An instrument used to measure extension.
Extrinsic	Extrinsic properties are characteristics of a test specimen as a whole, rather than the material from which it is made. Examples of extrinsic properties are mass and volume. Opposite of intrinsic, examples of which include temperature, density, and melting point.
Force	The classic definition of force is any action that alters or tends to alter a body's state of rest or of uniform motion in a straight line. Force has both magnitude and direction, making it a vector quantity. Newton's second law states that an object with a constant mass will accelerate in proportion to the net force acting upon and in inverse proportion to its mass. Equivalently, the net force on an object equals the rate at which its momentum changes. Force is measured in newton, which is an SI derived unit. One Newton is the force required to accelerate 1 kg to 1 m/s^2 . Forces acting on three-dimensional objects may also cause them to rotate or deform, or result in a change in pressure and/or volume in some cases. The tendency of a force to cause changes in rotational speed about an axis is called torque.
Frequency	The number of times an event occurs per unit time. Rate. Measured in hertz (Hz), kilohertz (kHz), or megahertz (MHz).

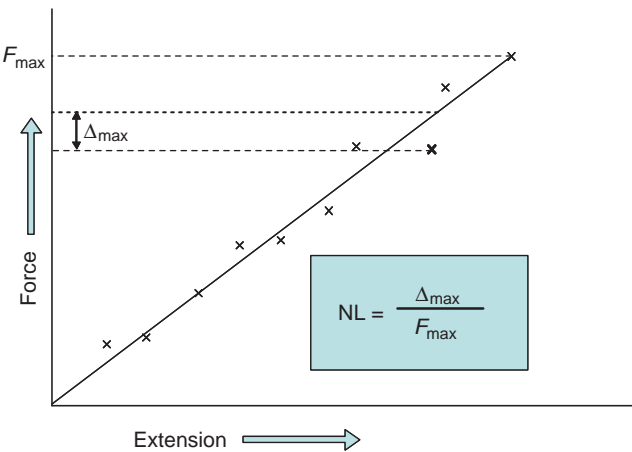
Frequency Response	This is the ability of a system output to replicate the input. Most electrical systems have limited frequency response (bandwidth), which allows the elimination of noise from the measurement. It is important for a system to have a bandwidth sufficient to capture significant events of the test, for example, details at sample break. Mecmesin Emperor driven systems have a sampling frequency of 2000 Hz, which means that a reading is taken every 0.5 ms.
Friction	Static friction is the force required to start one surface sliding against another. Dynamic friction is the force required to keep one surface sliding over another.
Full Scale (FSD)	This is the capacity of a sensor, in this context, the load measurement system or load cell. FSD stands for full-scale deflection and is derived from old analogue displays, which used a needle against a background scale.
Gauge Factor	A figure of merit for strain gauges. A sensitivity ratio that describes the relative ability of a strain gauge to measure strain.
Gauge Length	Used to calculate elongation. This is often stated in industry standards as the distance between the grips on a universal testing machine.
Grating	An optical device that allows spectroscopic analysis of incident light.
Gravity	The relatively weak force that is measured between two masses.
Grips/Fixtures	A mechanical device that grasps and holds the test specimen. Mecmesin grips include mechanical, pneumatic, vice grips, textile attachments, and specialist fixtures.
Hooke's Law	Assuming perfectly elastic behavior, stress is linearly related to strain: Robert Hooke, in the seventeenth century, stated "As the extension, so the force." $F = -kx$, where x is the displacement, F is the restoring force required and k is the force, or spring constant. Materials that exhibit near ideal elastic behavior are sometimes called "Hookean" materials.
Hysteresis	When a material is cycled from zero stress to a peak stress and then back to zero again, the stress/strain curve follows a closed loop. This is because the material does not have perfect elasticity and is slightly stretched on the "up" stroke. The maximum difference between stress on the upward and downward strokes of the cycle is quoted as the hysteresis, normally in terms of % FSD.



Intrinsic	The fundamental properties of a material. Melting point, density, heat capacity, and so on.
Joule	The SI derived unit of work, energy, and heat. It is the work done when a force of 1 N moves through a distance of 1 m. The area under a stress/strain curve gives the measurement of work done on a sample during a test.
Kilogram Force (Kgf)	The “kilogram force” is 9.80665 N. The practice of expressing force in kgf arose from the historical confusion that existed between “mass” and that particular force with which we are all familiar and which we refer to as “weight.” Although not adopted as a unit by the International System of Units (SI), kgf may be defined as “the force that produces an acceleration of 9.80665 m/s^2 (this is the internationally adopted value of the acceleration due to gravity) on a mass of one kilogram.” A kgf may sometimes be written as “kp” and referred to as a “kilopond.”
Kilopascal (kPa)	The SI derived unit of pressure and stress is the Pascal (force per unit area: $1 \text{ Pa} = 1 \text{ N/m}^2$). The “kilopascal” is 1000 N/m^2 or 1 mN/mm^2 .
lbf	The true imperial unit of force is the poundal (the force that produces an acceleration of one foot per second per second on a mass of one pound), but this is no longer used. Instead the term “pound force” is in widespread use and is a curious mixture of imperial and metric measurements. One lbf is about 4.44822 N, the force required to accelerate 1 lb by 9.80665 m/s^2 .
Load	An alternative term to mean force. The two are used interchangeably.
Load Cell	The transducer used in Mecmesin Universal testing machines. Load cells are based on strain gauge technology. This is a tried and tested method of measuring force using a change in electrical resistance brought about by

	dimensional changes in a metal block caused by increasing and decreasing force.
Load at Yield	The load reported at the point at which a specified deviation from proportionality of stress and strain occurs.
LOP (Limit of Proportionality)	This is the point on the stress/strain curve beyond which the stress does not increase proportionally with strain.
LVDT	Linear variable differential transformer. A type of electrical transformer used to measure linear displacement. Three coils are used, with the central coil being the primary. When an alternating current is present in the primary coil, a voltage is induced in the secondary coils, which is proportional to the inductance between them and the primary. As the coils are moved along a tube, their mutual inductances change and a voltage change is seen in the secondary coils. The difference between the two secondary coil voltages is proportional to the linear displacement along the tube.
Mass	Mass is related to force by the equation, $F = mg$, where mass is effectively constant, subject to the special theory of relativity where the mass of a body is measured in terms of its total energy content.
Magnetostriction	The effect seen when a change in the strain on a magnetic object produces a change in the measured magnetic flux in the object.
Measuring Range	The difference between maximum and minimum loads in a specific test or application. It must not exceed load cell capacity.
Median Force	This is the central value in a group of values, when placed in order. If there is an even number of values, then the median is halfway between the two central values.
Minimum Load	The lowest load in a specific test or application. This is not necessarily zero load, but includes the weight of any fixtures which are attached plus any intentional preload which is applied.
Modulus of Elasticity	Rate of change of strain as a function of stress. The slope line portion of a stress/strain diagram.
MPa	The SI derived unit of stress is the Pascal (1 N/m^2). The “megapascal” is 1 N/mm^2 .
Newton	The SI derived unit of force is the newton (the force that produces an acceleration of 1 m/s^2 on a mass of 1 kg).
Nonlinearity	Nonlinearity can be evaluated for either an individual data point or for a measurement run. For an individual point it is the difference between measured output at a specific load and the corresponding point on the straight-line fit to the output data. Normally expressed in units of %FSD. The nonlinearity for a measurement is usually defined as the maximum deviation between the measured output

and straight-line fit to the output data. This can be expressed in terms of the measurement units or %FSD.



Nonrepeatability
Overload

Sometimes used to mean the same as repeatability. Applying a force that exceeds the capacity of the load cell. Repeated overload, or significant overload can damage the load cell and require a recalibration or replacement. It is important to ensure that there is headroom built in to your system, that is, use a load cell that has a capacity of approximately 120% of the expected maximum load.

Photo Etch

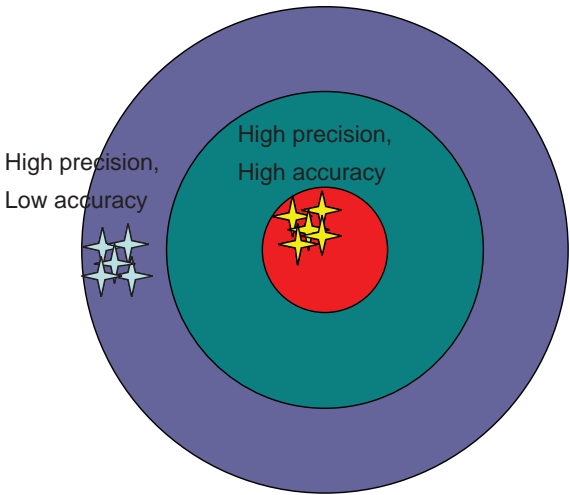
A manufacturing process utilizing light upon a photo chemically sensitive layer to create a template pattern on very thin material.

Piezoelectric effect

Whereby a material exhibits a voltage created when its dimensions are changed. Conversely, a piezoelectric material will exhibit a dimensional change if an electric field is applied.

Precision

The expression of how close multiple measurements of the same quantity are to each other. Not to be confused with accuracy.

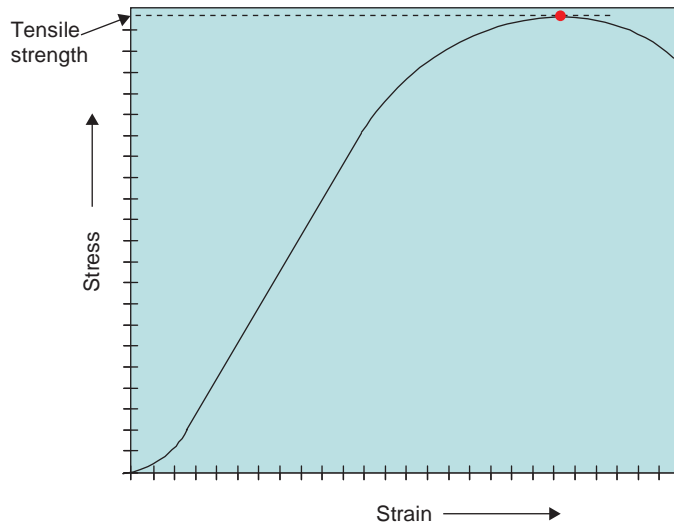


psi	“Pounds per square inch,” the imperial unit of stress.
Quasi-digital	An output signal that is time-based, for example, a frequency, that can be then coupled directly to a digital circuit with time measurement capability.
Refractive Index	The ratio of speed of light in a material to the speed of light in a vacuum
Repeatability	<p>The closeness of agreement between independent results obtained with the same method on identical test material, under the same conditions (same operator, same apparatus, same laboratory and after short intervals of time). Traditionally, repeatability is expressed as the standard deviation of results from repeated measurements of loadings under identical loading and environmental conditions using the same operator. When using this method, quoting 2 times the standard deviation gives approximately 95% confidence. Sometimes the maximum difference between output readings is used. Normally expressed in units of % RO or %FSD.</p> <p>In some contexts repeatability may be defined as the value below which the absolute difference between two single test results obtained under the above conditions, may be expected to lie with a specified probability.</p>
Reproducibility	Reproducibility, although similar, should not be confused with repeatability. It is the closeness of agreement between independent results obtained with the same method on identical test material, under the same conditions but with different operators.
Guide to the Expression of Uncertainty in Measurement	The resolution corresponds to the smallest two digits/values which can be read from a measurement device. Resolution should in no way be confused with Accuracy.
Resonance	Resonance occurs when the frequency of an input signal coincides with the natural resonant frequency of a material or object.
Sensitivity	The response of a device to a unit change input.
SI unit	<p>In 1960, the <i>Système International d’Unités</i> was introduced to replace systems of units based upon the meter/kilogram/second (mks), the centimeter/gram/second (cgs) and the foot/pound/second (fps). It is based on seven basic units:</p> <p>Length, meter (m) Mass, kilogram (kg) Time, second (s) Electrical current, ampere (A)</p>

	Temperature, Kelvin (K)
	Amount of a substance, mole (mol)
	Luminosity, candela (cd)
	All other measurement units can be derived by a combination of these basic units.
Signal	The absolute level of the measurable quantity into which a force input has been converted. Normally expressed as a displacement (as in a dial-type force gauge), or a voltage or current in electrical gauges and load cells.
Standard Deviation	This is the root mean square deviation from the mean of a random distribution.
Standard Uncertainty	A component of uncertainty in a measurement, expressed as a standard deviation S . All components of an uncertainty will be combined to form the uncertainty of measurement as quoted on a calibration certificate.
Static Force	In door closing applications: This is the average force from the end of the dynamic time until data acquisition stops, that is, at 6 s.
Strain	Usually used to denote the engineering strain. Quantitatively, it is the change in length of a specimen (as a result of an applied load) divided by the original length of the specimen; it is usually expressed as a percentage.
Strain Gauge	A device with electrical resistance that is a function of the applied strain. Normally in a Wheatstone bridge configuration
Stress	Usually used to denote the engineering stress. It is the load applied to a specimen, divided by the cross-sectional area of the specimen. Pressure. Can be measured in N/mm^2 , N/m^2 , psi, MPa, or kPa.
Tare	To automatically subtract either the weight of a container, fixture, or specimen, or the residual force being exerted by the specimen. Subsequent readings correspond either to the weight of the contents of a container, or to the force being exerted (through the fixturing) by a specimen under tensile or compressive load.
Temperature Range, Compensated	The range of temperature over which the load cell is compensated to maintain output and zero balance within specified limits.
Temperature Range, Operating	The extremes of ambient temperature within which the load cell will operate without permanent adverse change to any of its performance characteristics.
Tension	A force tending to stretch or elongate a specimen or material.

Tensile Strength

The ultimate strength of a material subjected to tensile loading. It is the maximum stress developed in a material during a tensile test.

**Tensile Test**

A tensile test is a way of determining how something will react when it is being pulled apart, or more correctly, when a force is applied to it in tension. Also known as a Pull test.

Torque

A twisting effect, or moment, exerted by a force acting at a distance on a body, equal to the force multiplied by the perpendicular distance between the line of action of the force and the center of rotation at which it is exerted.

Torque is often expressed in units: N m, Kgf m, kgf cm, lbf ft, and lbf in

Torsion Test

Method for determining behavior of materials subjected to twisting loads. Shear properties are often determined in a torsion test (ASTM E-143).

Torsional Modulus

AKA torsional modulus of elasticity, it is usually equal to the shear modulus. It is the modulus of elasticity of a material subjected to a twisting force.

Traceability

Traceability is the property of a measured result or value of a standard, whereby it can be related to stated references, usually national or international primary or secondary standards, through an unbroken chain of comparisons and all having stated uncertainties.

Transducer

Device that converts a physical input into an output of a different form.

True Strain

This differs from engineering strain in that it is determined from the rate of change in gauge length with respect to the instantaneous gauge length. It is expressed as a natural logarithm of engineering strain.

True Stress

True stress is the load divided by the instantaneous area of the specimen.

Ultimate Strength

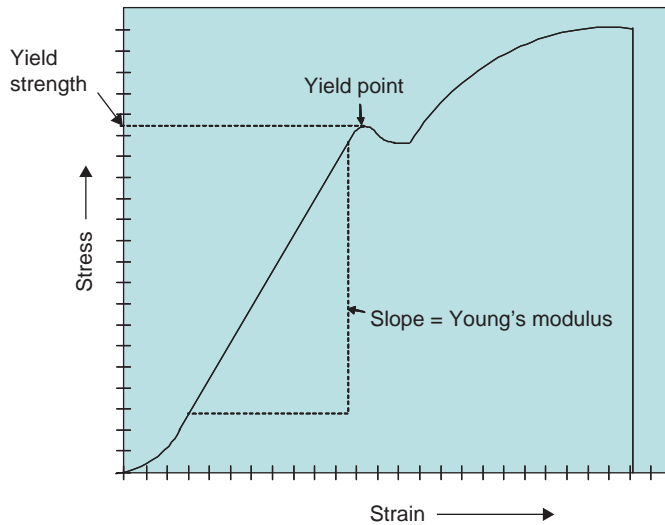
The highest engineering stress developed in material during a test and prior to rupture, or break. Normally, changes in area due to changing load and Necking are disregarded in determining ultimate strength.

Wave

Consisting of amplitude and frequency, a wave is a disturbance through a material or vacuum that is normally accompanied by a transfer of energy.

**Yield Point/
Yield Strength**

The point during a tensile test when a specimen ends elastic deformation and begins plastic deformation. There is an increase in strain with no corresponding increase in stress.



Yield strength is an indication of maximum stress that can be developed in a material without causing plastic deformation. It is the stress at which a material exhibits a specified permanent deformation and is a practical approximation of elastic limit.

**Young's
Modulus**

Hooke's law states

$$F = -kx$$

where F is the restorative force, x is the displacement, and k is a force (or spring) constant. Put another way

$$E = \text{stress/strain}$$

where stress is the force applied to the specimen and strain is the relative elongation of the specimen from its original length. E is known as Young's Modulus, represented as the gradient of the stress/strain curve within the initial linear region.

REFERENCES

- Aristotle. (4th century BC) *Physics: Bks. 1-4—Loeb Classical Library* No. 228, Vol 2. Cambridge: Harvard University Press;1934.
- Ausanio G et al. Magnetoelastic sensor application in civil buildings monitoring. *Sensors and Actuators A: Physical* 2005;123–124:290–295.
- Bienkowski A, Szewczyk R, Salach J. Industrial application of magnetoelastic force and torque sensors. *Acta Physica Polonica A* 2010;118:(5):1008–1009.
- Born M, Wolf E. *Principles of Optics*. 7th ed. Cambridge University Press;1999.
- Boynton R. *Precise Measurement of Mass*. Sawe Paper. No. 3147. Arlington (TX): S.A.W.E., Inc.;2001.
- Campbell CK. Applications of surface acoustic and shallow bulk acoustic wave devices. *Proceedings of IEEE* 1989;77(10):1453–1484.
- Charles W. The Bakerian Lecture: “An Account of Several New Instruments and Processes for Determining the Constants of a Voltaic Circuit”. *Philosophical Transactions of the Royal Society of London* 1843;133:303–327.
- Cheshmehdoost A, Jones BE. A new cylindrical structure load cell with integral resonators. Proceedings SENSORS VI: Technology, Systems and Applications. Bristol: IOP Publishing; 1993; p. 429–434.
- Cullen D, Montress T. Progress in the development of SAW resonator pressure transducers. Proceedings Ultrasonics Symposium;1980;2:519–522.
- Cullen D, Reeder T. Measurement of SAW velocity versus strain for YX and ST quartz. Proceedings Ultrasonics Symposium: 1975;519–522.
- Dan Mihai S. *Handbook of Force Transducers Principles and Components*. Berlin, Heidelberg: Springer-Verlag;2011.
- Dietrich J. Simmons and the strain gage. *Engineering and Science* 1986;50:19–23.
- Eernisse EP, Ward RW, Wiggins RB et al. Survey quartz bulk resonator sensor technologies. Proceedings IEEE International Frequency Control Symposium. 1996; p. 24–32.
- Einstein A. Die Grundlage der allgemeinen Relativitätstheorie. *Annalen der Physik* 1916;354: 769–822. Doi: 10.1002/andp.19163540702.
- Ekreem NB, Olabi AG, Prescott T, Rafferty A, Hashmi MSJ. An overview of magnetostriction, its use and methods to measure these properties. *Journal of Materials Processing Technology* 2007;191(1–3):96–101.
- European co-operation for Accreditation. *Expression of the Uncertainty of Measurement in Calibration*. Publication reference EA-4/02;1999
- Forcesensys Ltd. *Printed Tuning Fork Sensors*. Application Whitepaper, Summer; 2007.
- Galileo G. Dialogue Concerning the Two Chief World Systems (1632)—Translated by Drake. 1953; abridged.
- Gholamzadeh B, Nabovati H. Fiber optic sensors. *World Academy of Science, Engineering and Technology* 2008;42:297–307.
- International Organization for Standardization. *Guide to the Expression of Uncertainty in Measurement*. International Organization for Standardization;1995.
- Hauptmann P. *Sensors, Principles and Applications*. Prentice Hall;1991.
- Hill KO et al. Photosensitivity in optical fiber waveguides: application to reflection filter fabrication. *Applied Physics Letters* 1978;32:647–649.
- Hooke R 1678, *Lectures de potentia restitutiva, or, Of spring [microform]: explaining the power of springing bodies: to which are added some collections*. Printed for J. Martyn, London

- Ikelle LT, Amundsen L. Introduction to petroleum seismology: 12 (investigations in geophysics). *Society of Exploration Geophysics* 2005.
- Institute of Measurement and Control (London). *Force Measurement Guide*. Institute of Measurement and Control (London);1996.
- Intiang J et al. Characteristics of 9mm metallic triple-beam tuning fork resonant sensor. Proceedings Sensors and their Applications XIV (SENSORS07) IOP Publishing. *Journal of Physics: Conference Series* 2007;76.
- Jackson PGS. The early days of the Saunders-Roe foil strain gauge. *Strain* 1990;26:61–66.
- Joule JP. *On the Effects of Magnetism upon the Dimensions of Iron and Steel Bars*. London, Edinburgh and Dublin Philosophical Magazine and Journal of Science 1847;30 (Third Series: 76–87):225–241.
- Langdon RM. Resonator sensors—a review. *Journal of Physics E: Scientific Instruments* 1985;18:103.
- Lonsdale A, Lonsdale B. Method and apparatus for measuring strain. International patent public. No. WO91/13832, 19 Sept. 1991, Int. Applic. No. PCT/GB91/00328, Int. filing date: 4 March 1991, Priority: 9004822.4, 3 March 1990, GB.
- M3003. *The Expression of Uncertainty and Confidence in Measurement*. 2nd ed. UKAS;2007.
- Rayleigh L. On waves propagating along the plane surface of an elastic solid. *Proceedings London Mathematical Society*; 1885;7:4–11.
- Kreuzer M. White Paper. Strain Measurement with Fiber Bragg Grating Sensors. HBM, Darmstadt, Germany; 2007. <http://www.HBM.com>.
- Newton I. *Principia Mathematica*; 1685.
- Simmons EE Jr. Material testing apparatus. United States patent no. 2292549. 1942.
- Tatti M et al. Production and use of an interferometric optical strain gauge with comparison to conventional techniques. *Optics and Lasers in Engineering* 1997;27:269–284.
- Thomson W. On the electro-dynamic qualities of metals. *Philosophical Transactions of the Royal Society of London*; 1856;146:649–751.
- Toshiyuki H et al. Development and evaluation of tuning fork type force transducers. XVIII Imeko World Congress, Metrology for a Sustainable Development, Sept, 17–22, 2006; Rio de Janeiro, Brazil;2006.
- Kalinin V et al. High-speed high dynamic range resonant SAW torque sensor for kinetic energy recovery system. *Proceedings ETFT* 2010; Session 5.
- White RM, Voltmer FW. Direct piezoelectric coupling to surface elastic waves. *Applied Physics Letters* 1965;7:314–316.
- Wilson BL, Tate DR, Borkowski G. *Proving rings for calibrating testing machines*. U.S. National Bureau of Standards. Circular C454; 1946.
- Wold C et al. Temperature measurement using surface skimming bulk waves. *Proceedings Ultrasonics Symposium*. 1991;1:441–444.
- Yan T et al. Metallic triple beam resonator with thick-film printed drive and pickup. *17th European Conference on Solid State Transducers*;2003; p. 10–13.
- Young T. *Lectures on natural philosophy*. London; Vol.1;1807. p. 78–79.
- Young and Freedman. *University Physics with Modern Physics* 10th ed. p. 806; 2000.
- Young and Freedman. *Instrumentation for Engineering Measurements* 2nd ed.; p. 129–135
- Yun-Jiang R. In fibre bragg grating sensors. *Measurement Science Technology* 1997;8:355–375.

11

RESISTIVE STRAIN MEASUREMENT DEVICES

MARK TUTTLE

- 11.1 Preliminary discussion
 - 11.1.1 Scope
 - 11.1.2 Definition of strain
 - 11.1.3 Practical implications
 - 11.2 Resistance metal strain gages
 - 11.2.1 General description
 - 11.2.2 Strain sensitivity
 - 11.2.3 Strain gage alloys and calibration parameters
 - 11.2.4 Strain gage rosettes
 - 11.2.5 The Wheatstone bridge
 - 11.3 Semiconductor strain gages
 - 11.4 Liquid metal strain gages
- References

11.1 PRELIMINARY DISCUSSION

11.1.1 Scope

The measurement of strain is a fundamental necessity in many engineering disciplines and related industries. Consequently many strain measurement techniques have been developed. Methods of measurement can be roughly grouped into two major categories: those based on the behavior of light and those based on electronic devices. Examples of the former include reflection photoelasticity, geometric moiré, moiré interferometry, holographic interferometry, and digital image correlation (DIC). Examples of the latter include techniques that infer strain based on some change in an electrical characteristic of a strain measurement device, such as changes in resistance, inductance, or capacitance.

A significant discussion of all of these techniques would require a major treatise and is beyond the scope of this chapter. Rather, discussion is focused herein on one subcategory: strain measurements obtained by monitoring resistance changes. This focus is appropriate because at present strain measurement via resistance strain gages is by far the most widely used technique, although other methods (particularly DIC) are gaining in popularity. The reader interested in strain measurement methods beyond those described here is referenced to one of several excellent books devoted to these topics, including references (Sharpe, 2008; Shukla and Dally, 2010; Dally and Riley, 2005; Cloud, 1995; Sutton et al., 2009).

11.1.2 Definition of Strain

In this chapter, strain is discussed within the framework of continuum mechanics. In continuum mechanics, matter is studied at a physical scale large enough such that existence of individual atoms or molecules is not perceptible. That is, continuum mechanics analyses treat matter as a continuous substance, rather than as an assembly of discrete particles. This forms the basis of two major fields of study, *solid* mechanics and *fluid* mechanics. A solid substance is one that can resist deformation due to shear stress, whereas a fluid substance cannot. This chapter is devoted to strains encountered in solid mechanics.

Strain is essentially a description of the deformations that have occurred at a point in a solid body. Many different factors may have caused the deformation, including application of surface forces, gravity effects, electromagnetic forces, temperature changes, and so on or some combination thereof. The mechanism(s) that caused the deformations are immaterial to this discussion. It is sufficient to say that the deformation has occurred, and hence a nonzero *state of strain* may exist at any point in the solid body. The definition used to describe strain depends on the magnitudes of the deformation. *Finite strain theory* is used when the deformation is arbitrarily large, whereas *infinitesimal strain theory* is used when the magnitude of the deformation is small. Throughout this chapter it will be assumed that deformations are small, such that infinitesimal strain theory can be applied. This is usually the case for structures encountered in civil, mechanical, and aerospace structural engineering.

Displacements of a solid body in the x -, y -, and z -directions are typically denoted $u(x, y, z)$, $v(x, y, z)$, and $w(x, y, z)$, respectively. Infinitesimal strains are related to displacements as follows:

$$\begin{aligned} \epsilon_{xx} &= \frac{\partial u}{\partial x} & \gamma_{xy} &= \gamma_{yx} = \left(\frac{\partial v}{\partial x} + \frac{\partial u}{\partial y} \right) \\ \epsilon_{yy} &= \frac{\partial v}{\partial y} & \gamma_{yz} &= \gamma_{zy} = \left(\frac{\partial w}{\partial y} + \frac{\partial v}{\partial z} \right) \\ \epsilon_{zz} &= \frac{\partial w}{\partial z} & \gamma_{zx} &= \gamma_{xz} = \left(\frac{\partial u}{\partial z} + \frac{\partial w}{\partial x} \right) \end{aligned} \quad (11.1)$$

Strain components ϵ_{xx} , ϵ_{yy} , and ϵ_{zz} are called *normal strains*, whereas $\gamma_{xy} = \gamma_{yx}$, $\gamma_{yz} = \gamma_{zy}$, and $\gamma_{zx} = \gamma_{xz}$ are called *engineering shear strains*. To fully describe the state of strain at a point, the numerical values of all six independent strain components must be specified.

Although strains occur in three-dimensions, most strain measurements are limited to strains that occur on the surface of a solid. Let the surface of interest lie in the x - y plane. Therefore, the three strain components to be measured are ϵ_{xx} , ϵ_{yy} , and γ_{xy} .

A physical interpretation of normal and engineering shear strains that occur in the x - y plane can be developed based on the two-dimension sketch shown in Figure 11.1. Imagine that a rectangle with dimensions dx and dy has been drawn on a flat surface. Three corners of the rectangle are labeled as A , B , and C . Initially, the angle $\angle ABC$ is precisely $\pi/2$ radians (i.e., initially $\angle ABC = 90^\circ$). Now assume some mechanism(s) causes a displacement of the surface, and consequently the rectangle displaces as shown. The sides of the rectangle have increased in length and rotated, such that $\angle A'B'C' < \pi/2$.

Normal strains in the x - and y -directions, ϵ_{xx} and ϵ_{yy} , respectively, are defined as change in length divided by the original length of the rectangle sides in the x - and y -directions. From Figure 11.1, and assuming small displacements and displacement gradients, we have

$$\epsilon_{xx} = \frac{\left(\partial x + \frac{\partial u}{\partial x}\right) - (\partial x)}{\partial x} = \frac{\partial u}{\partial x}$$

and

$$\epsilon_{yy} = \frac{\left(\partial y + \frac{\partial v}{\partial y}\right) - (\partial y)}{\partial y} = \frac{\partial v}{\partial y}$$

These results are among those listed in Equation (11.1). A normal strain corresponding to an increase in length is called a *tensile strain* and is algebraically positive. Conversely, a normal strain corresponding to a decrease in length is called a *compressive strain* and is

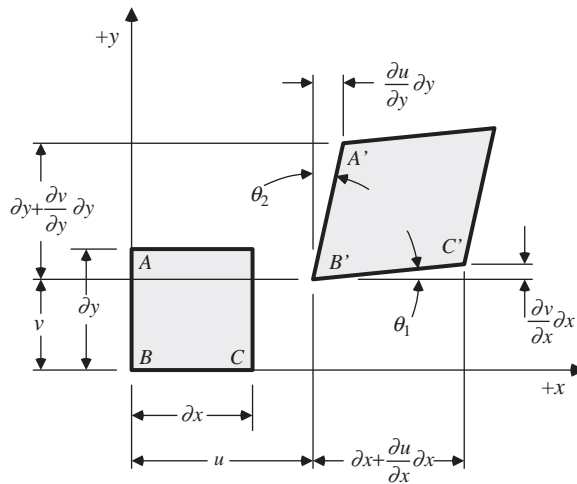


FIGURE 11.1 Sketch used to visualize infinitesimal strain.

algebraically negative. Note that normal strain is a unitless quantity, since it is a ratio of two lengths.

Referring again to Figure 11.1, engineering shear strain γ_{xy} is defined as the *change* in $\angle ABC$, expressed in radians. That is, engineering shear strain equals the *sum* of angles θ_1 and θ_2

$$\gamma_{xy} = \gamma_{yx} = \theta_1 + \theta_2$$

The tangent of angle θ_1 is given by

$$\tan \theta_1 = \frac{\frac{\partial v}{\partial x}}{\frac{\partial u}{\partial x} + \frac{\partial v}{\partial x}} = \frac{\frac{\partial v}{\partial x}}{1 + \frac{\partial u}{\partial x}}$$

Assuming small displacement gradients, the denominator in this expression is very close to unity, and hence

$$\tan \theta_1 \approx \frac{\partial v}{\partial x}$$

An identical process leads to

$$\tan \theta_2 \approx \frac{\partial u}{\partial y}$$

Again assuming small displacements and displacement gradients, which in the present context implies that both θ_1 and θ_2 are small angles expressed in radians, we have

$$\tan \theta_1 \approx \theta_1 \quad \text{and} \quad \tan \theta_2 \approx \theta_2$$

Combining the above, we find

$$\gamma_{xy} = \gamma_{yx} = \theta_1 + \theta_2 = \frac{\partial v}{\partial x} + \frac{\partial u}{\partial y}$$

This result is among those listed in Equation (11.1). A shear strain is positive if $\angle ABC$ has *decreased* (thus, Figure 11.1 illustrates a *positive* engineering shear strain). Since shear strain is an angle measured in radians, and since radians are defined as the ratio of two lengths and are, therefore, unitless, shear strain is a unitless quantity.

In passing, the reader should contrast the definition of engineering shear strain described earlier to the definition of *tensoral shear strain*, usually denoted $\varepsilon_{xy} = \varepsilon_{yx}$. Tensoral shear strains are defined as the *average* of angles θ_1 and θ_2 , rather than the sum:

$$\varepsilon_{xy} = \varepsilon_{yx} = \frac{\theta_1 + \theta_2}{2}$$

Hence, engineering shear strain differs from tensoral shear strain by a factor of 2:

$$\gamma_{xy} = \gamma_{yx} = 2\varepsilon_{xy} = 2\varepsilon_{yx}$$

Although the use of tensoral shear strain is mathematically elegant, engineering shear strains are far more commonly used in practice. The use of engineering shear strain will be assumed throughout the remainder of this chapter. In the following discussion, the phrase “engineering shear strain” will be abbreviated as “shear strain.”

Consider the tensile member shown in Figure 11.2a. A square grid pattern referenced to the x - y coordinate system is printed on the surface of the member. An axial tensile load is applied, causing both an increase in length and decrease in width (Figure 11.2b). It is apparent that the initially square grids will deform into rectangles. Qualitatively, the following strains are induced

$$\begin{aligned}\varepsilon_{xx} &< 0 \text{ (compressive)} \\ \varepsilon_{yy} &> 0 \text{ (tensile)} \\ \gamma_{yy} &= 0 \text{ (no change in angle)}\end{aligned}$$

Now consider the identical tensile member shown in Figure 11.3a. In this case, the square grid pattern is referenced to a x' - y' coordinate system oriented θ -degree counter-clockwise from the longitudinal axis. An identical axial load is applied, causing an identical increase in length and decrease in width. However, the initially square grids now deform into a rhombus. The insert implies that the following strains have been induced

$$\begin{aligned}\varepsilon_{x'x'} &< 0 \text{ (compressive)} \\ \varepsilon_{y'y'} &> 0 \text{ (tensile)} \\ \gamma_{y'y'} &> 0 \text{ (decrease in corner angle)}\end{aligned}$$

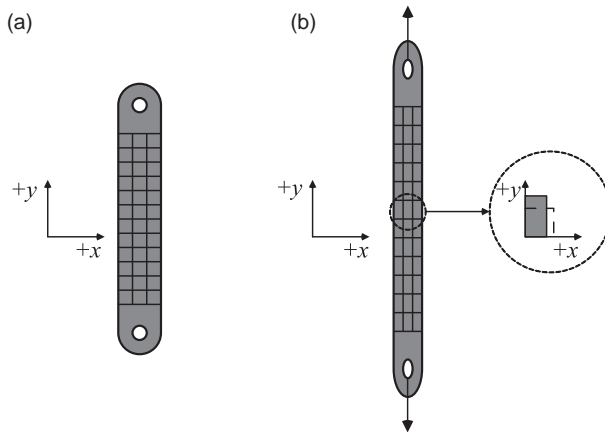


FIGURE 11.2 Beam with a square grid pattern parallel to beam axis. (a) Undeformed. (b) Deformed.

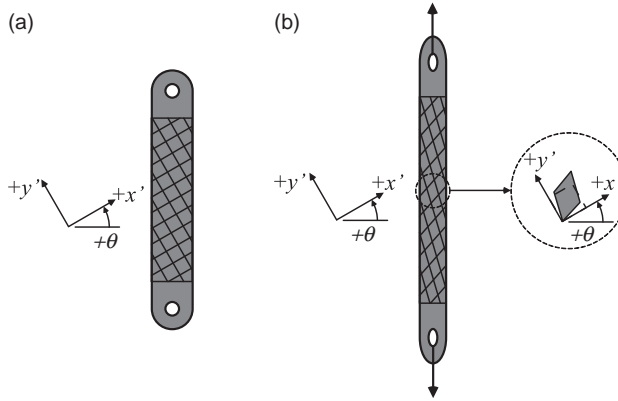


FIGURE 11.3 Beam with a square grid pattern inclined θ -degree from the beam axis. (a) Undeformed. (b) Deformed.

Figures 11.2 and 11.3 illustrate two important points. First, although strain is a description of deformation, the numerical values of the individual strain components that (collectively) describe the deformation depend on the coordinate system use. For example, the numerical values of the strain components illustrated in Figure 11.2 differ from the numerical values of the strain components illustrate in Figure 11.3, even though the tensile member deformation is identical in both cases. The description of the deformation, that is, the strains, can be rotated from one coordinate system to another using the *strain transformation equations*. Recall that the present discussion has been limited to strains within a plane. For this case, if strain components relative to the x - y coordinate system are known (ϵ_{xx} , ϵ_{yy} , γ_{xy}), then strain components referenced to a new x' - y' coordinate system-oriented θ -degree from the x - y coordinate system are given by

$$\begin{aligned}\epsilon_{x'x'} &= \epsilon_{xx}\cos^2\theta + \epsilon_{yy}\sin^2\theta + \gamma_{xy}\sin\theta\cos\theta \\ \epsilon_{y'y'} &= \epsilon_{xx}\sin^2\theta + \epsilon_{yy}\cos^2\theta - \gamma_{xy}\sin\theta\cos\theta \\ \gamma_{x'y'} &= 2\sin\theta\cos\theta(\epsilon_{yy} - \epsilon_{xx}) + \gamma_{xy}(\cos^2\theta - \sin^2\theta)\end{aligned}\quad (11.2)$$

The algebraic sign of angle θ is defined in accordance with the right-hand rule. For example, in Figure 11.3 angle θ is a positive angle since axis x' is rotated *from* the x -axis *toward* the y -axis.

The second important point illustrated by Figures 11.2 and 11.3 is that there is always a particular coordinate system, called the *principal strain coordinate system*, in which shear strains are zero. Obviously, the coordinate system used in Figure 11.2 is a principal strain coordinate system, whereas the coordinate system used in Figure 11.3 is not. The normal strains that exist in the principal coordinate system are called the *principal strains* and are denoted ϵ_1 , ϵ_2 , and ϵ_3 . Since in this discussion it has been assumed that out-of-plane shear strains $\gamma_{yz} = \gamma_{xz} = 0$, one of the principal strains is the out-of-plane normal strain ϵ_{zz} . The other two principal strains lie within the x - y plane. Ordinarily, the three principal strains are ordered such that $\epsilon_1 > \epsilon_2 > \epsilon_3$. However, since in the present case ϵ_{zz} is always a principal strain, the

convention adopted here is that $\varepsilon_3 = \varepsilon_{zz}$, and the two in-plane strains will be labeled ε_1 and ε_2 . If strain components ($\varepsilon_{xx}, \varepsilon_{yy}, \gamma_{xy}$) are known, then the in-plane principal strains are given by

$$\varepsilon_{1,2} = \frac{\varepsilon_{xx} + \varepsilon_{yy}}{2} \pm \sqrt{\left(\frac{\varepsilon_{xx} - \varepsilon_{yy}}{2}\right)^2 + \left(\frac{\gamma_{xy}}{2}\right)^2} \quad (11.3)$$

The angle θ_p from the x -axis to the 1-axis is given by

$$\theta_p = \tan^{-1} \left[\frac{2(\varepsilon_1 - \varepsilon_x)}{\gamma_{xy}} \right] = \tan^{-1} \left[\frac{\gamma_{xy}}{2(\varepsilon_1 - \varepsilon_y)} \right] \quad (11.4)$$

11.1.3 Practical Implications

As is evident from the preceding discussion, strain measurement involves two types of measurements. Normal strains require measurement of a change in length, and shear strains require measurement of a change in angle. Conceptually, these could be made with a simple ruled scale and protractor. For two reasons, these simple measurement devices are usually inadequate. First, the changes in lengths and angles involved are exceedingly small and difficult to measure precisely with simple measurement devices. A numerical example will serve to illustrate this difficulty. Many steel alloys will yield at a strain level of about 0.0012 m/m, or about $1200 \times 10^{-6} \text{ m/m} = 1200 \mu\varepsilon$ (read: “1200 microstrain”). This implies a steel element with an initial length of 10 mm will increase to a length of 10.012 mm at yielding. Recognizing that structural engineers are required to measure strains well below yielding, it is obvious that ruled scales do not provide the resolutions required. Similarly, many steel alloys subjected to a pure shear strain will yield at a shear strain level of about 0.0016 rad, or about $1600 \times 10^{-6} \text{ rad} = 1600 \mu\varepsilon$ (about 0.09°). Once again, simple protractors cannot be used to measure such small angular changes.

Second, since strains generally vary throughout a structure, the goal is usually to measure strains “at a point.” Thus, not only must very small changes in length and angle be measured, but also these changes must be measured over as small an initial area as possible.

Finally, for a simple isotropic structure such as that shown in Figure 11.2 the orientation of the principal axis is known by inspection. In contrast, for the complex structural shapes and loadings typically encountered in practice, the orientation of the principal coordinate system is rarely known a priori.

11.2 RESISTANCE METAL STRAIN GAGES

11.2.1 General Description

Nowadays, the most widely used strain measurement device is certainly the resistance metal strain gage and is originally known as “bonded wire strain gages,” they are based on a phenomenon first noticed by William Thomson (later named Lord Kelvin) in 1856: the resistance of an electrically conductive metal wire is increased if the wire is stretched

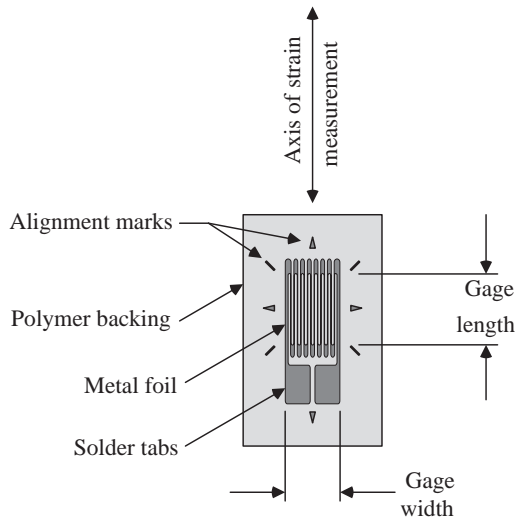


FIGURE 11.4 Basic elements of a metal foil strain gage.

and decreased if the wire is compressed. This phenomenon has since come to be known as the piezoresistive effect.¹

Based on this observation and working independently, Edward Simmons (in about 1937) and Arthur Ruge (in 1938) bonded small metal wires to structures they were studying and monitored the resistance change as the structures were loaded. The change in resistance provided an indirect measure of the change in wire length and hence strain. Simmons and Ruge are considered to be the coinventors of resistance strain gages. The first resistance strain gages became commercially available in 1941 and were marketed under the trade name “SR-4 gage,” in honor of Simmons, Ruge, and two colleagues who helped perfect and market the gage.

The next major development occurred in 1952, when engineers at Saunders-Roe Ltd. used the photoetch process to produce strain-sensing elements from thin metal foil rather than wire. The etching process makes it possible to produce large numbers of very small and complex sensing elements from a single parent sheet of metal foil. Although wire gages are still used in some circumstances, today most resistance metal strain gages are produced using the photoetch process.

The basic construction of a metal foil strain gage is summarized in Figure 11.4. The sensing element is etched from a parent metal foil in the desired pattern and bonded to a thin and flexible polymeric backing film. Backing films of polyimide or glass-reinforced epoxy are most commonly used, and serve to stabilize the delicate foil pattern during handling and also to electrically isolate the metal foil from the underlying substrate. The metal foil/backing film is subsequently adhesively bonded to the surface of interest. A good bond between the surface and the gage assembly is critically important, since the strain being measured must be transferred through the adhesive bondline and backing film to the metal foil. Detailed strain gage bonding instructions are provided by strain gage manufactures, and the user is well advised to follow these instructions precisely.

¹ The piezoresistive effect should not be confused with the piezoelectric effect. In the former, a mechanical strain results in a resistance change. In the latter, a mechanical strain causes an electrical potential to occur.

The gage pattern is designed such that strain is sensed primarily in the axial direction of the grid. Alignment marks are included as an alignment aid during the bonding process. Traditionally, leadwires are soldered to the solder tabs of the gage after it has been bonded to the surface, although it is possible to attach the leadwires prior to bonding, and in fact most strain gage manufacturers now offer gages with preattached leadwires.

11.2.2 Strain Sensitivity

Strain sensitivity is a measure of how much resistance changes per unit strain and is defined as

$$S = \frac{\left(\frac{\partial R}{R}\right)}{\varepsilon} \quad (11.5)$$

Strain sensitivity for pure metals ranges from about $-12 < S < 6$. Alloys commonly used in strain gages have strain sensitivities ranging from about 2 to 4. Note that a positive strain sensitivity implies that a tensile strain ($\varepsilon > 0$) causes an increase in resistance ($\partial R > 0$).

Strain sensitivity is often equated to the “gage factor” of the strain gage. In reality, there is a subtle difference between strain sensitivity, as defined by Equation (11.5), and the gage factor measured for a strain gage. The difference arises because of differences in the strain field present during measurement of strain sensitivity versus the strain field present during measurement of the gage factor. That is, strain sensitivity of an alloy is measured by applying uniaxial stress to an unconstrained metal specimen (typically a wire with circular cross-section) with initial resistance R . The resulting normal strain in the direction of loading ε and change in resistance ∂R are measured and used to calculate the strain sensitivity. Since the specimen is unconstrained, due to the Poisson effect a transverse strain equal to $-\nu\varepsilon$ also occurs, where ν is the Poisson ratio of the wire. The definition of strain sensitivity implies that this transverse strain is allowed to occur.

If the transverse strain actually differs from $-\nu\varepsilon$, which is a common occurrence for a strain gage bonded to a structure, then the resistance change of the bonded strain gage may not correspond to the strain sensitivity. This situation is described in greater detail in a following subsection entitled “Gage Factor and Transverse Sensitivity Coefficient.”

11.2.3 Strain Gage Alloys and Calibration Parameters

Strain gages have been produced using many different metal alloys since their introduction in 1941. The most common gage alloys are summarized in Table 11.1. These alloys have become preferred for various reasons, including the ability to roll the alloy into thin metal foils, ease of soldering leadwires to the alloy, linear change in resistance with strain, high strain to failure, good fatigue life, and thermal stability. Calibration parameters provided by the manufacturer include the gage resistance, the gage factor and transverse sensitivity coefficient, and the self-temperature compensation (STC) number.

Gage Resistance: In principle, a strain gage with any initial resistance can be used. In practice, 120, 350, or 1000 Ω gages are most common, since most commercial

TABLE 11.1 Common Resistance Strain Gage Alloys

Alloy	Nominal Composition	Gage Factor (S_g)	Comments
Constantan (or advance)	45% Ni, 55% Cu	2.1	Most widely used general purpose gage alloy Max strain $\sim 3\text{--}5\%$ (30,000–50,000 $\mu\epsilon$)
Karma	74% Ni, 20% Cr, 3% Al, 3% Fe	2.0	Excellent long-term stability More difficult to solder than Constantan Max strain $\sim 1.5\%$ (15,000 $\mu\epsilon$)
Isoelastic	36% Ni, 8% Cr, 0.5% Mo, 55.5% Fe	3.6	High thermal output Max strain $\sim 1.5\%$, but nonlinear at strains above 0.5% (5,000 $\mu\epsilon$) Excellent fatigue life
Nichrome V	80% Ni, 20% Cr	2.1	Used for high-temperature strain gages
Armour D	70% Fe, 20% Cr, 10% Al	2.0	Difficult to solder using standard methods
Alloy 479	92% Pt, 8% W	4.0	

strain gage amplifiers and data acquisition systems are designed to accept one of these resistances. The primary reason for selecting one gage resistance over another is the self-heating effect. That is, during the measurement process a voltage is applied to the gage, and the associated electrical current passing through the gage will cause an increase in gage temperature. Unstable temperature fluctuations may occur if a relatively high voltage is applied and the gage is bonded to a substrate that is a poor thermal conductor. The manufacturer provides recommendations on the maximum voltage (i.e., power) that can be safely applied based on the thermal properties of the substrate.

Gage Factor and Transverse Sensitivity Coefficient: A strain gage subjected to a biaxial strain field is shown in Figure 11.5. The resistance of the gage will be changed due to *both* the axial and the transverse strain components, ϵ_a and ϵ_t , respectively. This is an undesirable characteristic, since it would be ideal if the gage would respond *only* to the axial strain component. The gage factor and transverse sensitivity coefficient are calibration parameters that allow the user to remove the undesirable response to transverse strains from the strain signal.

Suppose a gage is subjected to an axial strain *only* (i.e., assume $\epsilon_t = 0$). The axial strain sensitivity of the gage, S_a , is defined under this condition

$$S_a = \frac{\left(\frac{\Delta R_a}{R_g}\right)}{\epsilon_a} \quad (11.6)$$

where R_g = original gage resistance; ΔR_a = change in gage resistance due to axial strain.

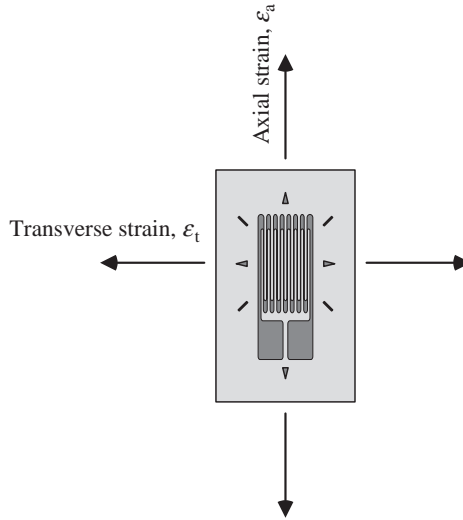


FIGURE 11.5 A strain gage subjected to a biaxial strain field.

Note that the definition of the axial strain sensitivity is similar to the expression used to calculate the strain sensitivity of an alloy (Eq. 11.5). However, a requirement associated with Equation (11.6) is that the transverse strain is zero, whereas this restriction is not imposed during measurement of the strain sensitivity of an alloy.

Now consider a case in which the gage is subjected to a transverse strain *only* (i.e., $\varepsilon_a = 0$). For this condition the transverse strain sensitivity of the gage can be defined as

$$S_t = \frac{\left(\frac{\Delta R_t}{R_g} \right)}{\varepsilon_t} \quad (11.7)$$

where ΔR_t = change in gage resistance due to transverse strain ε_t .

Gage manufacturers measure both the axial and the transverse strain sensitivities by subjecting a gage to pure axial or pure transverse strain, using a fixture specially designed for this purpose and described in an ASTM standard (ASTM Standard E 251, 2009). The *transverse sensitivity coefficient*, K_t , is defined as the ratio of the transverse and axial strain sensitivities:

$$K_t = \left(\frac{S_t}{S_a} \right) \quad (11.8)$$

The measured K_t is included in the calibration materials provided with each gage and is customarily defined as a percentage. For example, if a transverse sensitivity of 0.010 is measured for a particular gage, the manufacturer will report a value $K_t = 1.0\%$.

Now consider the case in which the gage is subjected to *both* the axial and transverse strains simultaneously. The total change in gage resistance ΔR in this case equals the sum of Equations (11.6) and (11.7):

$$\frac{\Delta R_g}{R_g} = \left(\frac{\Delta R_a}{R_g} \right) + \left(\frac{\Delta R_t}{R_g} \right) = S_a \varepsilon_a + S_t \varepsilon_t \quad (11.9)$$

Substituting the definition of K_t , this equation can be written as

$$\frac{\Delta R_g}{R_g} = S_a (\varepsilon_a + K_t \varepsilon_t) \quad (11.10)$$

During calibration, the strain gage is subjected to a *biaxial* strain field with a known ratio of $\varepsilon_t/\varepsilon_a$ (ASTM Standard E 251, 2009). This is achieved by mounting a gage in the axial direction on a tensile specimen with known Poisson ratio ν_o and applying an axial load (usually, $\nu_o = 0.285$). The resulting uniaxial state of stress causes a biaxial strain field, where the transverse strain is caused by the Poisson effect:

$$\varepsilon_t = -\nu_o \varepsilon_a \quad (11.11)$$

Substituting Equation (11.11) into Equation (11.10), the change in resistance for these conditions is

$$\frac{\Delta R_g}{R_g} = S_a (1 - \nu_o K_t) \varepsilon_a \quad (11.12)$$

The *gage factor* S_g is defined as

$$S_g = S_a (1 - \nu_o K_t) \quad (11.13)$$

Substituting this definition into Equation (11.12), we have

$$\frac{\Delta R_g}{R_g} = S_g \varepsilon_a \quad (11.14)$$

Through careful design of the gage grid strain gage manufacturers have successfully reduced transverse strain sensitivities to near-zero levels for many gage alloys and styles. If transverse strain sensitivity is reduced to a negligibly small value, then $K_t \approx 0$, $S_g \approx S_a$, and Equation (11.14) reduces to Equation (11.5). If transverse strain sensitivity is not negligibly small, then a strain gage measurement must be corrected for transverse sensitivity effects. Methods to correct for transverse sensitivity effects will be described in Section 11.2.4.

The strain sensitivity of an alloy, and consequently the gage factor of a strain gage, exhibits a modest temperature dependency. The magnitude and linearity of gage factor variation with temperature depends on the alloy used. As a rough rule of thumb, the gage factor for most commercial strain gages changes by 1–2% for a 100°C temperature change.

Self-Temperature Compensation Number: Consider a structure that is subjected to both external loading and to a temperature change ΔT . Some fraction of the strains induced in the structure is due to the external loading, while a fraction is due to thermal expansion or contraction caused by the temperature change. For modest temperature changes strains associated with thermal expansion/contraction (ε^T) can be calculated using the linear thermal expansion coefficient, α :

$$\varepsilon^T = \alpha(\Delta T) \quad (11.15)$$

The linear thermal expansion coefficient has been measured for most structural materials and is available in reference handbooks. For example, the thermal expansion coefficient for steel alloys is about $11 \mu\text{ε}/^\circ\text{C}$ ($6 \mu\text{ε}/^\circ\text{F}$).²

A strain measurement is said to be “temperature compensated” if that portion of the strain associated with free thermal expansion or contractions is subtracted from the total strain measurement. If this can be achieved, then only strains associated with stress will be measured.

A straightforward method of strain gage temperature compensation is to mount a gage on a small coupon of the material of interest, and to subject the unconstrained coupon to uniform temperature changes under carefully controlled laboratory conditions. The obtained measurements are plotted against temperature, generating a so-called “thermal apparent strain” curve. An identical strain gage is now bonded to a structure made from the same material, and the structure is subjected to the loads and temperature changes that occur in service. Apparent strains at a given temperature are subtracted from the strains measured in service, thereby compensating the strains measurements for temperature.

Initially, one might suppose that apparent strains would simply reflect the difference in thermal expansion of substrate and gage, $(\alpha_s - \alpha_g)\Delta T$. However, a change in the resistance of an *unbonded* strain gage also occurs with a change in temperature. Consequently even if the thermal expansion coefficients of the substrate and gage are identical, such that no mismatch exists, the resistance of a strain gage will change with temperature and an apparent strain will be measured. The change in resistance caused by a uniform temperature change ΔT can be written as

$$\left. \frac{\partial R}{R} \right|_{T_o} = \left[\beta_g + S_g \left(\frac{1 + K_t}{1 - \nu_o K_t} \right) (\alpha_s - \alpha_g) \right] \Delta T \quad (11.16)$$

where β_g is called the temperature coefficient of resistance of the gage alloy and is a measure of the change in gage resistance with temperature. In general, β_g is a highly nonlinear function of temperature, and can be algebraically negative or positive depending on temperature. For most alloys β_g can be manipulated over a wide range, through the use of cold-working and heat treatment processes. This gives rise to the concept of a

²This discussion assumes that the material is isotropic. Anisotropic materials, such as composites, usually exhibit three different thermal expansion coefficients, which are measured along the principal material coordinate system. This added complexity will not be addressed here. Suffice it to say that temperature-compensated strain gages can be used to measure strains in anisotropic materials, although extra care is needed to account for different thermal expansion coefficients in different material directions.

“self-temperature compensated” strain gage. The gage manufacturer manipulates β_g such the change in gage resistance due to a temperature change is equal and opposite to resistance changes due to the mismatch in thermal expansion coefficients. Because β_g is highly nonlinear compensation, it is not exact and can only be approximated over a limited temperature range. The *self-temperature compensation* number of a strain gage corresponds roughly to the thermal expansion coefficient of the test material the strain gage is intended to be bonded to. The system of temperature measurement used by the gage manufacturer is reflected in the STC number. That is, some manufacturers define the STC number in °F, whereas others define the STC number in °C. For example, a gage produced by manufacturer “A” and intended for use on steel could have an STC number of 06 (corresponding to $6\text{ }\mu\epsilon/\text{°F}$), whereas a similar gage offered by manufacturer “B” may have an STC number of 11 (corresponding to $11\text{ }\mu\epsilon/\text{°C}$).

A typical apparent strain is shown in Figure 11.6. In this case, a self-temperature compensated gage has been mounted to a sample of 17-4PH (a precipitation hardened stainless steel alloy), and the self-temperature compensated gage exhibits an apparent strain of less than $\pm 100\text{ }\mu\epsilon$ over the range $-75\text{°C} \leq T \leq 200\text{°C}$. The gage manufacturer typically fits the entire apparent structure curve to a fourth-order polynomial, and provides this curve-fit to the user. Curve fits based on temperatures measured in °C and °F are provided below the apparent strain curve shown in Figure 11.6.

11.2.4 Strain Gage Rosettes

The strain gage shown in Figures 11.4 and 11.5 is called a “uniaxial” strain gage, since it is intended to measure strain in only one direction. As discussed in the background

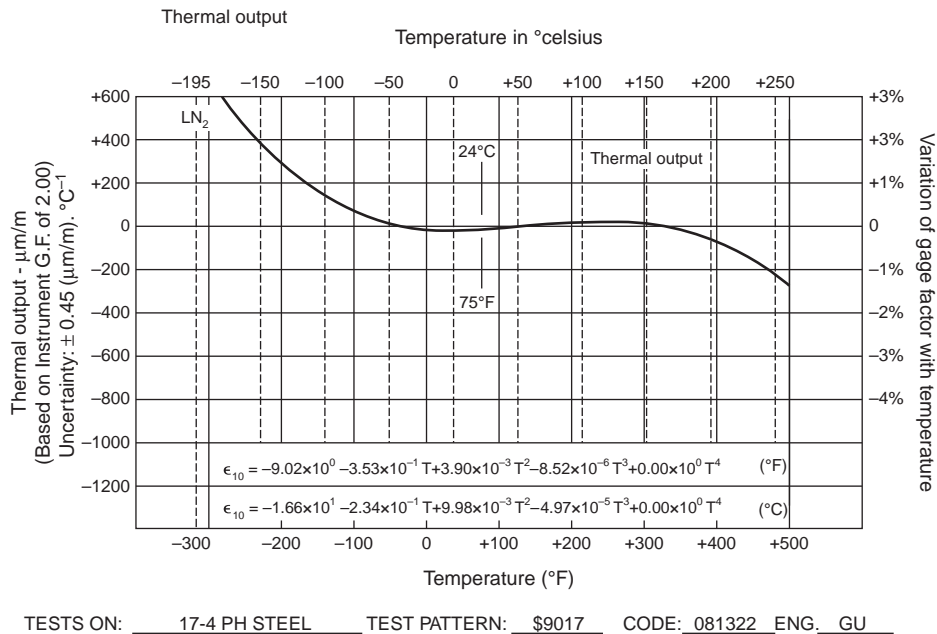


FIGURE 11.6 An apparent strain curve obtained for a strain gage with STC number 06, mounted in a coupon of 17-4 precipitation hardened stainless steel.

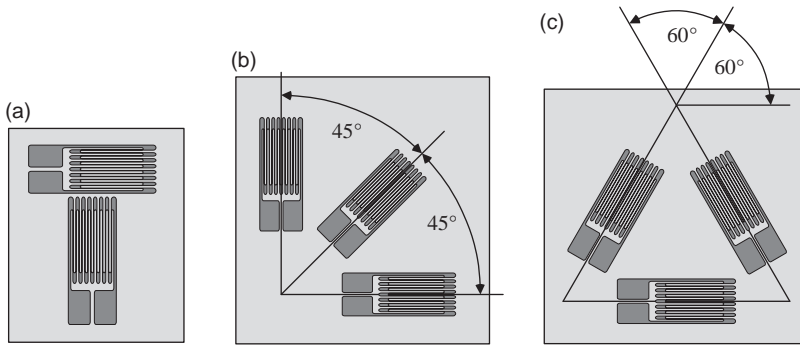


FIGURE 11.7 Common strain gage rosettes. (a) Biaxial rosette. (b) Rectangular rosette. (c) Delta rosette.

section 11.1.2, in practice three components of strain must be measured to determine the state of strain within a plane: ϵ_{xx} , ϵ_{yy} , and γ_{xy} . Once these strain components are known, the principal strains ϵ_1 and ϵ_2 and the orientation angle θ of the principal strain coordinate system can be calculated using Equations (11.3) and (11.4), respectively.

Since there are in general three unknowns (ϵ_{xx} , ϵ_{yy} , and γ_{xy} , or equivalently, ϵ_1 , ϵ_2 , and θ), three independent strain measurements are normally required to determine the state of strain within a plane. An exception occurs when the orientation angle θ is known a priori due to specimen geometry and loading (as in Figure 11.2, for example), in which case there are only two unknowns (ϵ_1 and ϵ_2), and only two independent strain measurements are required.

Conceptually, three uniaxial strain gages (or two uniaxial gages, if there are only two unknowns) could be used to measure strains in three (or two) different directions. However, as a practical matter it is easier to precisely align the gages in the intended orientation if three (or two) strain gages are bonded to the same backing by the gage manufacture. Strain gages produced in this manner are called *strain gage rosettes*.

Strain gage rosettes are commonly available in three forms, as shown in Figure 11.7. A biaxial rosette (also known as a T-rosette) consists of two perpendicular and electrically independent gage elements. Two types of three-element rosettes are available: the rectangular rosette, where the three gage elements are orientated in 45° increments, and the delta rosette, where the three gage elements are oriented in 60° increments.³

11.2.4.1 Rosette Equations Three strain gages oriented at distinct angles θ_a , θ_b , and θ_c relative to the x -axis are shown in Figure 11.8. Applying the first of Equation (11.2) to each of these gages in turn, we have

$$\begin{aligned}\epsilon_a &= \epsilon_{xx}\cos^2\theta_a + \epsilon_{yy}\sin^2\theta_a + \gamma_{xy}\sin\theta_a\cos\theta_a \\ \epsilon_b &= \epsilon_{xx}\cos^2\theta_b + \epsilon_{yy}\sin^2\theta_b + \gamma_{xy}\sin\theta_b\cos\theta_b \\ \epsilon_c &= \epsilon_{xx}\cos^2\theta_c + \epsilon_{yy}\sin^2\theta_c + \gamma_{xy}\sin\theta_c\cos\theta_c\end{aligned}\quad (11.17)$$

³ As is evident in Figure 11.7c, delta rosettes are so named because the pattern of the three strain gages often resembles the upper case Greek letter “Δ.”

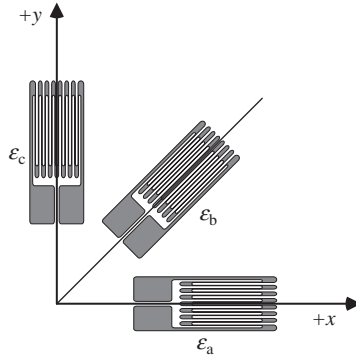


FIGURE 11.8 A rectangular rosette relative to an x - y coordinate system.

Equation (11.17) is called the *general rosette equations*. Since ε_a , ε_b , ε_c , θ_a , θ_b , and θ_c are all measured, we have three equations involving three unknowns, ε_{xx} , ε_{yy} , and γ_{xy} .

A rectangular rosette is shown relative to an x - y coordinate system in Figure 11.8. In this case,

$$\theta_a = 0^\circ \quad \theta_b = 45^\circ \quad \theta_c = 90^\circ$$

Substituting these angles into Equation (11.17) and solving for the unknown strains, we have

$$\begin{aligned} \varepsilon_{xx} &= \varepsilon_a \\ \varepsilon_{yy} &= \varepsilon_c \\ \gamma_{xy} &= 2\varepsilon_b - (\varepsilon_a + \varepsilon_c) \end{aligned} \quad (11.18)$$

Equation (11.18) is the rosette equations for use with rectangular rosettes.

A delta rosette is shown relative to an x - y coordinate system in Figure 11.9. In this case,

$$\theta_a = 0^\circ \quad \theta_b = 60^\circ \quad \theta_c = 120^\circ$$

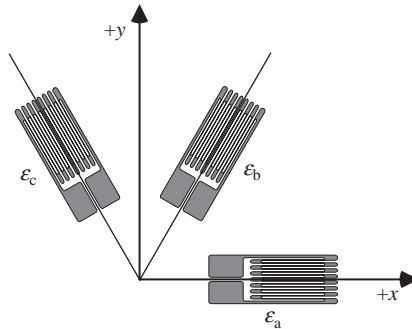


FIGURE 11.9 A delta rosette relative to an x - y coordinate system.

Substituting these angles into Equation (11.17) and solving for the unknown strains, we have

$$\begin{aligned}\varepsilon_{xx} &= \varepsilon_a \\ \varepsilon_{yy} &= \left(\frac{1}{3}\right)[2(\varepsilon_b + \varepsilon_c) - \varepsilon_a] \\ \gamma_{xy} &= \left(\frac{2}{\sqrt{3}}\right)[\varepsilon_b - \varepsilon_c]\end{aligned}\tag{11.19}$$

Equation (11.19) is the rosette equations for use with delta rosettes.

11.2.4.2 Corrections for Transverse Sensitivity Effects As previously discussed, strain gages generally respond to both axial and transverse strains. Therefore, a measured axial strain is generally in error due to transverse sensitivity effects. The transverse sensitivity coefficient K_t provided by the manufacturer indicates the severity of this problem for a given strain gage. Gage manufacturers have been successful in reducing K_t to low levels; for modern metal foil gages K_t is usually less than $\pm 5\%$ (i.e., $|K_t| < 0.05$). At these low levels transverse sensitivity can often be ignored, depending on the level of accuracy needed in a given application. Nevertheless, for the highest accuracy strain gages measurements should be corrected for transverse effects. At least two orthogonal strain measurements are required to correct for transverse sensitivity, which means that a 2- or 3-element strain gage rosette must be used.

Consider first the correction of strains measured using a biaxial rosette. Referring to Figure 11.7a, denote the uncorrected strains measured by the two gage elements as ε_{m0° and ε_{m90° . Strain measurements corrected for transverse sensitivity effects, ε_{0° and ε_{90° , are then given by

$$\begin{aligned}\varepsilon_{0^\circ} &= \frac{(1 - \nu_o K_t)(\varepsilon_{m0^\circ} - K_t \varepsilon_{m90^\circ})}{(1 - K_t^2)} \\ \varepsilon_{90^\circ} &= \frac{(1 - \nu_o K_t)(\varepsilon_{m90^\circ} - K_t \varepsilon_{m0^\circ})}{(1 - K_t^2)}\end{aligned}\tag{11.20}$$

where ν_o = Poisson's ratio of the calibration material used by the gage manufacturer during measurement of K_t (usually $\nu_o = 0.285$).

In the case of a 3-element rectangular rosette (Figure 11.7b), denote the uncorrected strain measurements as ε_{m0° , ε_{m45° , and ε_{m90° . The corrected strain measurements are given by

$$\begin{aligned}\varepsilon_{0^\circ} &= \frac{(1 - \nu_o K_t)(\varepsilon_{m0^\circ} - K_t \varepsilon_{m90^\circ})}{(1 - K_t^2)} \\ \varepsilon_{45^\circ} &= \frac{1 - \nu_o K_t}{1 - K_t^2} [\varepsilon_{m45^\circ} - K_t (\varepsilon_{m0^\circ} + \varepsilon_{m90^\circ} - \varepsilon_{m45^\circ})] \\ \varepsilon_{90^\circ} &= \frac{(1 - \nu_o K_t)(\varepsilon_{m90^\circ} - K_t \varepsilon_{m0^\circ})}{(1 - K_t^2)}\end{aligned}\tag{11.21}$$

Finally, for the case of a 3-element delta rosette (Figure 11.7c), denote the uncorrected strains as ε_{m0° , ε_{m60° , and ε_{m120° . The corrected strain measurements are given by

$$\begin{aligned}\varepsilon_{0^\circ} &= \frac{1 - \nu_o K_t}{1 - K_t^2} \left[\left(1 + \frac{K_t}{3} \right) \varepsilon_{m0^\circ} - \frac{2}{3} K_t (\varepsilon_{m60^\circ} + \varepsilon_{m120^\circ}) \right] \\ \varepsilon_{60^\circ} &= \frac{1 - \nu_o K_t}{1 - K_t^2} \left[\left(1 + \frac{K_t}{3} \right) \varepsilon_{m60^\circ} - \frac{2}{3} K_t (\varepsilon_{m0^\circ} + \varepsilon_{m120^\circ}) \right] \\ \varepsilon_{120^\circ} &= \frac{1 - \nu_o K_t}{1 - K_t^2} \left[\left(1 + \frac{K_t}{3} \right) \varepsilon_{m120^\circ} - \frac{2}{3} K_t (\varepsilon_{m0^\circ} + \varepsilon_{m60^\circ}) \right]\end{aligned}\quad (11.22)$$

Equations (11.20) and (11.21) are based on the assumption that all gage elements exhibit the same transverse sensitivity coefficient, which is usually the case. Analogous expressions that account for different transverse sensitivities for different gage elements are available (Errors due to Transverse Sensitivity in Strain Gages).

11.2.5 The Wheatstone Bridge

In practice, strain gages are routinely used to measure strains to a resolution of $1 \mu\varepsilon$. The change in gage resistance caused by $1 \mu\varepsilon$ is very small. Using Equation (11.14) and assuming a gage factor $S_g = 2.0$

$$\frac{\Delta R}{R} = S_g \varepsilon_a = (2)(1 \mu\varepsilon) = 2 \times 10^{-6}$$

This shows that the change in resistance caused by $1 \mu\varepsilon$ is only about 2 parts per million. For example, a strain gage with an initial resistance of precisely 120Ω will exhibit a resistance of about 119.9998Ω if subjected to a tensile strain of $1 \mu\varepsilon$. It is very difficult to measure these small resistance changes directly. Consequently strain gages are usually wired into special electrical circuits that convert the small resistance change into a relatively larger voltage change that is easier to accurately measure. The Wheatstone bridge is the most widely applied circuit of this type and will be reviewed here.

A basic Wheatstone bridge circuit is shown in Figure 11.10. Four resistances (R_1 to R_4) are wired into a four-arm pattern. An excitation voltage (V) is applied across junctions a and c , and an output voltage (E) is monitored between junctions b and d .

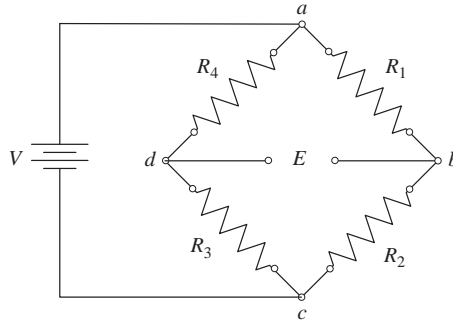


FIGURE 11.10 The Wheatstone bridge circuit.

Output voltage E is related to the excitation voltage V as follows:

$$E = \left[\frac{R_1 R_3 - R_2 R_4}{(R_1 + R_2)(R_3 + R_4)} \right] V \quad (11.23)$$

The bridge is said to be “balanced,” if the output voltage is $E = 0$. From Equation (11.23) it is seen that this occurs if $R_1 R_3 = R_2 R_4$. That is, the bridge is balanced if

$$\frac{R_2}{R_1} = \frac{R_3}{R_4} \quad (11.24)$$

In strain gage applications, the bridge is initially balanced. Often (but not always) the four resistances are initially identical ($R_1 = R_2 = R_3 = R_4$). As discussed later, one or more of the resistances may be a strain gage. If any of the four resistances change the bridge may become unbalanced ($E \neq 0$). Suppose that all four resistances are strain gages, that the initial resistance of all gages is identical ($R_1 = R_2 = R_3 = R_4 = R$, say), and that all four gages experiences a strain and consequently exhibit a change in resistance. It can be shown (Sharpe, 2008; Shukla and Dally, 2010; Dally and Riley, 2005) that the resulting output voltage is given by

$$E = \left(\frac{V}{4R} \right) [\Delta R_1 - \Delta R_2 + \Delta R_3 - \Delta R_4] (1 - \eta) \quad (11.25)$$

The quantity η represents a nonlinear term. That is, rigorously speaking the output voltage of the Wheatstone bridge is a nonlinear function of the change(s) in resistance. In most cases, the nonlinearity is very small and can be ignored, although if the bridge is initially unbalanced then errors due to nonlinearities can be significant. As a rule of thumb, in practical strain measurements the percent error due to bridge nonlinearities are roughly equal to the measured strain expressed as a percentage. Wheatstone bridge nonlinearities can usually be safely ignored and will not be further considered here.

Ignoring the nonlinear term η , Equation (11.25) becomes

$$E = \left(\frac{V}{4R} \right) [\Delta R_1 - \Delta R_2 + \Delta R_3 - \Delta R_4] \quad (11.26)$$

Equation (11.26) shows that the effects of resistance changes in adjacent arms (ΔR_1 and ΔR_2 , for example) are subtractive, whereas changes in opposite arms (ΔR_1 and ΔR_3 , for example) are additive. In fact, if $\Delta R_1 = \Delta R_2 = \Delta R_3 = \Delta R_4$, the output voltage is $E = 0$, and the bridge remains balanced.

In general-purpose strain analysis each strain gage is monitored using a separate Wheatstone bridge circuit as shown in Figure 11.11, rather than placing four gages in a single bridge. This is called a quarter-bridge circuit, since resistances R_2 – R_4 are now fixed, and only the gage resistance R_g changes as strain is applied. Equation (11.26) becomes

$$E = \left(\frac{\Delta R_g}{R_g} \right) \frac{V}{4} \quad (11.27)$$

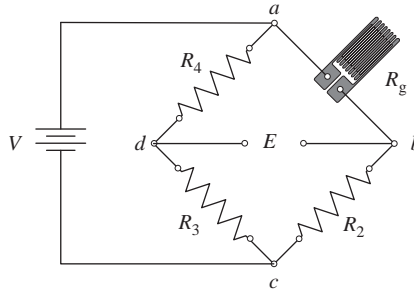


FIGURE 11.11 A quarter Wheatstone bridge circuit.

Combining Equations (11.14) and (11.26), we find

$$\varepsilon = \frac{4E}{S_g V} \quad (11.28)$$

Equation (11.28) summarizes how a resistance strain gage is used to measure strain. A strain gage with known gage factor S_g is bonded to a surface of interest. The gage is then wired into a quarter-bridge Wheatstone circuit, a known voltage V is applied, and the circuit is initially balanced. If the strain gage subsequently experiences a strain the resistance will change, the bridge becomes unbalanced, and the resulting output voltage E is measured and interpreted in terms of strain using Equation (11.28).

11.2.5.1 The Three-Leadwire System During practical implementation of the Wheatstone circuit fixed resistors R_2 – R_4 are incorporated into a strain gage amplifier and leadwires of the appropriate length are used to connect the strain gage to the rest of the circuit and complete the bridge. The situation that exists if only two-leadwires are used is shown schematically in Figure 11.12a. Note that the leadwires possess a measurable resistance R_L , so the total resistance in arm a – b is $(R_g + 2R_L)$. The magnitude of R_L depends, in part, on the length of the leadwire and is often a few ohms. Assuming $R_2 = R_3 = R_4 = R_g$, the resistance of the leadwires will cause an initial imbalance of the Wheatstone bridge. For this reason commercial strain gage amplifiers often incorporate variable resistors that

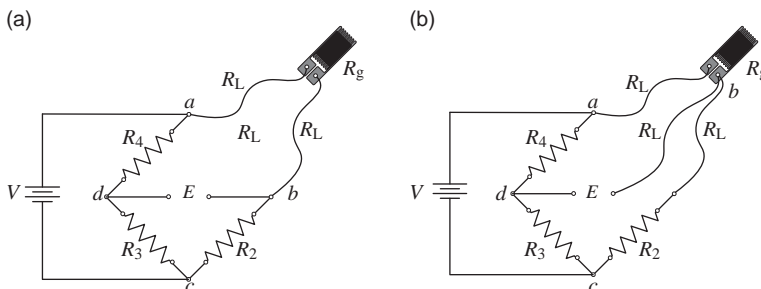


FIGURE 11.12 Illustration of the two-leadwire circuit versus three-leadwire Wheatstone bridge circuit. (a) Two-leadwire circuit, resulting in an increase of $2R_L$ in resistance of arm a – b . (b) Three-leadwire circuit, where equal an equal resistance of R_L is added to arms a – b and b – c .

allow the resistance of adjacent arms to be varied, so that the bridge can be initially balanced in accordance with Equation (11.24). However, if a temperature change occurs during the course of a strain measurement the *leadwire* resistance will change, since the resistance of the leadwires is governed in part by the temperature coefficient of resistivity of the leadwire alloy. A change in resistance due to leadwire effects is interpreted as strain, so changes in leadwire resistance results in an erroneous apparent strain measurement.

Due to the characteristics of the Wheatstone bridge this situation can be remedied through the use of three leadwires. A three-leadwire circuit is shown in Figure 11.12b. Note that use of three leadwires effectively places an additional resistance R_L in adjacent arms $a-b$ and $b-c$, and moves junction b to one of the strain gage solder tabs. The voltmeter used to measure the output voltage E has very high impedance (on the order of $M\Omega$), so adding a small resistance of R_L between junctions $b-d$ has no effect on measurement of the output voltage. The resistance of arm $a-b$ is $(R_g + R_L)$, whereas the resistance of arm $b-c$ is $(R_2 + R_L)$. All three leadwires are kept physically as close as possible so that the temperature along the length of the leadwires is essentially identical (this is easily achieved in practice, since in most cases all three leadwires are encased in the same sheathing). Therefore, identical leadwire resistance changes occur in adjacent arms and are canceled, as per Equation (11.26).

11.3 SEMICONDUCTOR STRAIN GAGES

Like metals, semiconductors exhibit the piezoresistive effect; their resistance changes when subjected to strain. In fact, the piezoresistive effect is much more pronounced in semiconductors than in metals or metal alloys, and consequently semiconductor strain gages have much higher gage factors than metal foil gages. Despite this advantage semiconductor gages are not widely used for general-purpose strain measurement. Two major drawbacks have limited their use. First, although the gage factor of semiconductors is very high, it is also a nonlinear function of strain. The gage factor of semiconductor strain gages varies substantially over the strain range of interest to structural engineers (often by a factor of 2–3), and this nonlinearity must be accounted for when interpreting the measured change in resistance. Second, the resistivity of semiconductors is far more sensitive to temperature changes than that of metals, so temperature compensation of semiconductor strain gages is generally less exact than for metal gages.

Semiconductor gages were first developed in the 1950s (Smith, 1954; Mason and Thurston, 1957) and were commercially available by about 1960. Most semiconductor gages are based on doped silicon. The two most common doping agents are phosphorous and boron. Silicon doped with phosphorous results in an n-type semiconductor, in which electrical conduction occurs due to the flow of electrons (a negative charge carrier). In contrast, silicon doped with boron results in a p-type semiconductor, where electrical conduction occurs due to vacancies in the atomic lattice (also called “holes”; a positive charge carrier).

The available literature and technologies describing semiconductor strain gages is not as detailed as for metal foil gages. The basic elements of uniaxial semiconductor gages are shown schematically in Figure 11.13. Since the piezoelectric effect exhibited by semiconductors is high there is no need for the undulating grid pattern that is characteristic of a metal foil gage (as shown in Figure 11.4, for example). Semiconductor gages are

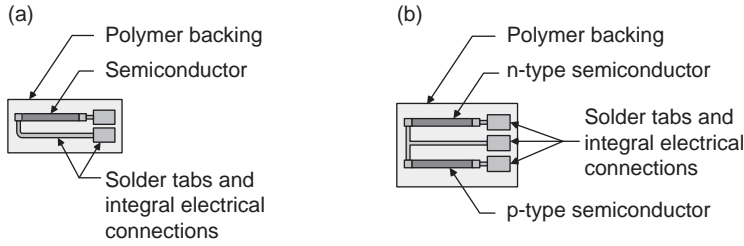


FIGURE 11.13 Typical semiconductor strain gages with polymer backing. (a) Uniaxial gage with a single sensing element. (b) Uniaxial gage with two sensing elements, providing thermal compensation.

commercially available with or without a polymer backing. Backings are usually made of phenolic resin, and serve to stabilize the gage element and as an aid during handling and bonding. Gages are often provided with preattached leadwires or leadwire ribbons. Both uniaxial and biaxial (T-rosettes) gages are available. Three-element rosettes are apparently not commercially available, at least as an off-the-shelf product. The uniaxial gage with two semiconductor elements shown in Figure 11.13b is a particularly interesting configuration because, as will be discussed later, the combined use of one n-type semiconductor and one p-type semiconductor in a Wheatstone bridge results in self-temperature compensation.

The transverse sensitivity of semiconductor strain gages has apparently not been measured, or at least a transverse sensitivity coefficient is not provided with commercially available semiconductor strain gages. The gage factor for semiconductor gages is simply equated to the strain sensitivity:

$$S_g = S = \frac{\left(\frac{\partial R}{R}\right)}{\epsilon}$$

The *magnitude* of gage factors of n- and p-types silicon semiconductors are similar, and range from about 40 to 200, compared with only 2–3 for metallic foil gages. Interestingly, the gage factor of p- and n-types semiconductor gages differ in algebraic sign. That is, the resistance of p-type silicon increases with a tensile strain (as do metal conductors); therefore, p-type semiconductor strain gages have a positive gage factor. In contrast, the resistances of n-type silicon decrease with tensile strain; therefore, n-type semiconductor gages have a negative gage factor.

The change of resistance of a semiconductor gage bonded to an unconstrained test sample and subjected to a change in temperature is given by Equation (11.16), except that it is now assumed that $K_t = 0$:

$$\left.\frac{\partial R}{R}\right|_{T_0} = [\beta_g + S_g(\alpha_s - \alpha_g)]\Delta T \quad (11.29)$$

It would be ideal if a semiconductor gage could be produced for which $\beta_g = -S_g(\alpha_s - \alpha_g)$, which would eliminate the resistance change due solely to free expansion or contraction due to ΔT and allow development of a self-temperature

compensated semiconductor gage using a single semiconductor element. Unfortunately, the temperature coefficient of resistance β_g for semiconductors is established strictly by the level of doping employed and cannot be modified by cold working or thermal treatments. However, thermal compensation can be achieved using the configuration shown in Figure 11.13b. This configuration placed sensing elements in two arms of a half-bridge Wheatstone circuit, where the p- and n-types gage elements are placed in adjacent arms. Compensation is achieved because the gage factor of p-type semiconductors is positive, whereas the gage factor of n-type semiconductors is negative, and because resistance changes in adjacent arms of a Wheatstone bridge are subtractive. As long as β_g and α_g for both elements are the same, resistance changes associated with the thermal coefficient of resistance and the mismatch in thermal expansion coefficients occur in adjacent arms and cancel. Since gage factors of p- and n-types silicon have opposite algebraic signs, resistance changes due to mechanically induced strains do not cancel, and in fact the overall strain sensitivity is increased. A practical requirement is that both gage elements must experience the same strain and temperature change. This requirement is generally satisfied since the gage elements are small and bonded to the same backing.

11.4 LIQUID METAL STRAIN GAGES

Liquid metal strain gages can be used to measure very large strains (in excess of 150% or more). They are often used to measure strain in biological tissues such as ligaments or tendons, and are also used in engineered structures that experience large deformations during service such as automobile, truck, or aircraft tires.

The first liquid metal strain gage was developed and applied to the study of biological structures by Whitney in the late 1940s (Whitney, 1949, 1953). Various applications have been described in the literature since that time, and design and fabrication guidelines appeared in 1983 (Stone et al., 1983). The demand for these devices has never been large enough to warrant commercialization, so liquid metal gages are usually custom-made and calibrated by the user. A typical configuration is shown in Figure 11.14. The gage consists of a column(s) of liquid metal, usually mercury, encased in an elastomeric casing capable of large extensions, often silicon rubber. Leadwires in contact with the liquid metal and embedded within the elastomeric body and are connected to external instrumentation.

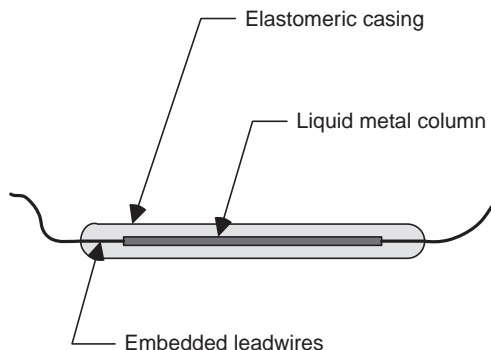


FIGURE 11.14 Typical configuration of a liquid metal strain gage.

The method of attachment to the structure varies according to application; in biological studies, the gage is often sutured to the tissue of interest.

The resistance of the gage is established by the liquid metal used and by length and diameter of the liquid metal column. Initial resistance is often very low, on the order of a few ohms. If a tensile strain is applied the length is increased and diameter is decreased (due to the Poisson effect), resulting in an increase in gage resistance. A gage factor is defined as normal and is determined by measuring the resistance change as a function of strain:

$$S = \frac{\left(\frac{\Delta R}{R}\right)}{\varepsilon}$$

A typical gage factor for liquid metal gages cannot be defined, since these devices are custom made. In general, gage factors are about 2, at least at small strain levels. Although these gages can sustain very high strain levels repeatedly and without failure, a cautionary note is that the gage factor often becomes nonlinear at high strain levels. The strain level at which nonlinear effects become pronounced cannot be specified, again because these gages are custom made, but in general substantial nonlinearities occur at strain levels greater than about 40%.

REFERENCES

- ASTM Standard E251. *Standard Test Methods for Performance Characteristics of Metallic Bonded Resistance Strain Gages*. West Conshohocken, PA: ASTM International;2009; DOI: 10:1520/E0251-92R09. www.astm.org
- Cloud G, *Optical Methods of Engineering Analysis*. Cambridge University;1995. p. 503. ISBN 0-521-45087-4.
- Dally JR, Riley WF. *Experimental Stress Analysis*. 4th ed. College House Enterprises LLC;2005. p. 686, ISBN 0-9762413-0-7.
- Errors due to Transverse Sensitivity in Strain Gages. Appendix, Tech Note TN-509, Micro-Measurements, Raleigh, NC 1982. Available at www.vishaypg.com/docs/11059/tn509tn5.pdf
- Mason WP, Thurston RN. Piezoresistive materials in measuring displacement, force, and torque. *Journal of Acoustic Society of America* 1957;29(10):1096–1101.
- Sharpe WN Jr., editor. *Handbook of Experimental Solids Mechanics* Springer;2008. p. 1098. ISBN 978-0-387-26883-5.
- Shukla A, Dally JW. *Experimental Solid Mechanics*. College House Enterprises LLC;2010. p. 668. ISBN 0-9792581-8-9.
- Smith CS. Piezoresistive effect in germanium and silicon. *Phys. Rev.* 1954;94(1):42–49.
- Stone JE, Madsen NH, Milton JL, Swinson WF, Turner JL. Developments in the design and use of liquid-metal strain gages. *Experimental Mechanics* 1983;23(2):129–139.
- Sutton MA, Ortu J-J, Schreier H. *Image Correlation for Shape, Motion and Deformation Measurements: Basic Concepts, Theory and Applications*. New York: Springer-Verlag;2009. p. 344. ISBN 0387787461.
- Whitney RJ. The measurement of changes in human limb volume by means of a mercury-in-rubber strain gauge. *Proceedings of Physical Society*. (March 26, 1949).
- Whitney RJ. The measurement of volume change in human limbs. *Journal of Physiology*. 1953;121:1–27.

12

VIBRATION MEASUREMENT

SHERYL M. GRACEWSKI AND NIGEL D. RAMOUTAR

- 12.1 Introduction
- 12.2 One-degree-of-freedom system response
 - 12.2.1 Free vibration response—time domain
 - 12.2.2 Time harmonic excitation—frequency domain
- 12.3 Multi-degree-of-freedom systems and the frequency response function
 - 12.3.1 Free vibration response—time domain
 - 12.3.2 Frequency domain response
 - 12.3.3 The frequency response function and discrete fourier transforms
- 12.4 Vibration measurement equipment and techniques
 - 12.4.1 Excitation sources
 - 12.4.2 Response and force measurement
 - 12.4.3 Vibration analyzers
- 12.5 Experimental modal analysis
 - 12.5.1 Modal analysis using frequency response functions (FRFs)
 - 12.5.2 Operational data analysis
- 12.6 Applications of vibration measurement
 - 12.6.1 Structural system characterization
 - 12.6.2 Machine condition monitoring
 - 12.6.3 Other machine fault characteristics
- Nomenclature
- References

12.1 INTRODUCTION

Vibration can be defined as an oscillation about an equilibrium point. This chapter will focus on techniques of vibration measurement used to characterize vibrations of structural systems. However, many of the concepts discussed here can be applied to other systems, such as oscillations in electrical circuits or fluidic systems. Vibrations are desirable in a wide range of structural systems, for example, musical instruments,

sieves, compactors, ultrasonic welding machines, and massaging cushions. In other systems, vibrations can be detrimental, possibly generating excessive noise and/or large oscillating displacements or forces that could lead to rapid wear or fatigue failure. In either case, vibration measurements can be used to obtain information about the dynamic response of a system so that the system's performance may be improved or structural failure avoided.

Vibration measurements can be used to obtain natural frequencies and mode shapes of a structure. The natural frequencies and mode shapes define the dynamic characteristics of the structure and can be used to predict the structure's response to other loading conditions. This information is useful when designing or modifying a system to decrease deleterious vibrations and/or to improve the system's performance. Often design specifications set a lower limit on a structure's natural frequencies to avoid a resonance response. Knowledge of a structure's dynamic characteristics is also useful when designing vibration isolation or vibration absorber systems for the structure.

Vibration measurements can also be used to determine a machine's dynamic response to forces that are generated during operation. Measurement of a structure's vibration response to operating forces can be used to identify optimum operating conditions or to monitor the condition of a machine. Any imbalance of a rotating shaft or component will produce a periodic force during operation. These periodic forces will cause vibrations that may affect operating performance. For example, in grinding machines, these vibrations can cause undesirable cutter marks. Identifying operating conditions that minimize vibration amplitudes can be used to improve the performance of a machine. In addition, vibrations can increase with part wear, and therefore the condition of a machine may be diagnosed by monitoring these self-generated vibrations. Machine condition monitoring can be used to identify, for example, bent, eccentric, or unbalanced shafts, faulty bearings or gears, and loose mechanical parts.

Measuring a system's response to impulsive, periodic, or random external excitations is useful in a wide range of applications. Shock or impulse excitation can be used to determine a structure's resistance to failure when dropped or when subjected to an earthquake or other impulsive loading. Time harmonic or more general periodic excitation can be used to determine a system's response to excitations caused by nearby rotating components/machinery or to predict a structure's fatigue life. Random excitation can also be used when it more accurately replicates real world environments, such as a vehicle's travel over a bumpy road.

An overview of the current state of the art in vibration measurement will be presented in this chapter. First, basic notation and concepts needed to interpret vibration measurements will be introduced in Section 12.2, by discussing the response of a one-degree-of-freedom system. These concepts will be generalized to multi-degree-of-freedom systems in Section 12.3 to emphasize the need to analyze vibration measurements in the frequency domain. Discrete Fourier transforms, the frequency response function, and parameters that control the accuracy of a dynamic measurement will also be discussed in Section 12.3. In Section 12.4, a range of vibration measurement equipment and techniques will be described that can be used to excite a structure and to measure and analyze its response. A summary of modal analysis techniques that can be used to obtain the natural frequencies, mode shapes, and modal damping parameters of a structure will be given in Section 12.5. Finally, in Section 12.6, some applications of vibration measurement will be discussed.

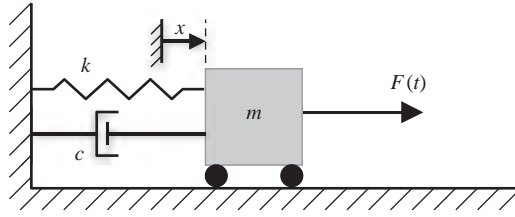


FIGURE 12.1 Single-degree-of-freedom system. The mass displacement x is measured from the static equilibrium position.

12.2 ONE-DEGREE-OF-FREEDOM SYSTEM RESPONSE

Fundamental vibration concepts and terminology will first be introduced by describing the vibration response of a system with a single degree of freedom. This one-degree-of-freedom system can adequately represent the measured response of a structure, when the lowest natural frequency of the structure dominates its response.

A one-degree-of-freedom vibration system can be modeled by a spring, with spring constant k , and a damper, with viscous damping constant c , both fixed at one end and attached to a mass m at the other end, as shown schematically in Figure 12.1. The displacement of the mass from its static equilibrium position is denoted by x . If an excitation force $F(t)$ is applied to the mass, the governing equation for the displacement $x(t)$ can be written as

$$m\ddot{x} + c\dot{x} + kx = F(t) \quad (12.1)$$

where a superposed dot denotes a time derivative, that is, $\dot{x} = dx/dt$. The initial displacement $x(0) = x_0$ and the initial velocity $\dot{x}|_{t=0} = v_0$ are needed to completely define the response of this system to a given excitation force.

12.2.1 Free Vibration Response—Time Domain

A single-degree-of-freedom system can be characterized by its free vibration response, that is, by its response when there is no externally applied force. The free vibration response of an undamped system ($c = 0$) is a harmonic oscillation of constant amplitude that can be expressed as

$$x(t) = x_0 \cos(\omega_n t) + \frac{v_0}{\omega_n} \sin(\omega_n t) \quad (12.2)$$

In Equation (12.2)

$$\omega_n = \sqrt{\frac{k}{m}} \quad (12.3)$$

is the *natural frequency* of the system in units of rad/s. The system will vibrate at its natural frequency when given an initial displacement or when excited by an impulse force. In vibration measurements, frequency f is typically specified in units of Hertz (Hz), where 1 Hz is equal to 1 cycle per second. Therefore, the natural frequency in Hz is defined as $f_n = \omega_n/2\pi$.

Any real vibrating system will lose energy to the environment. In Equation (12.1), the energy loss is modeled by a viscous damping term with damping force, $F_d = -c\dot{x}$, proportional to the mass's velocity. This is the easiest form of damping to include analytically. Damping in many systems is small so that this approximation is adequate for representing the dissipation in the system's response. The *nondimensional damping ratio*

$$\zeta = \frac{c}{2\sqrt{km}} \quad (12.4)$$

is often used to characterize the damping of a system. For underdamped systems ($\zeta < 1$), the free vibration will be oscillatory, but the amplitude of the oscillation will decay exponentially with time. The free vibration response of an underdamped system, satisfying Equation (12.1) (with force $F(t) = 0$) can be written as

$$x(t) = e^{-\zeta\omega_n t} \left(x_0 \cos(\omega_d t) + \frac{v_0 + \zeta\omega_n x_0}{\omega_d} \sin(\omega_d t) \right) \quad (12.5)$$

where

$$\omega_d = \omega_n \sqrt{1 - \zeta^2} \quad (12.6)$$

is the frequency of damped vibration. In typical systems, damping is small ($\zeta \ll 1$) so the frequency of damped vibration is approximately equal to the natural frequency, that is, $\omega_d \sim \omega_n$ or $f_d \sim f_n$.

The periodic nature and exponential decay of the free vibration response of a single-degree-of-freedom system is illustrated in Figure 12.2. The *period* T_d of the oscillation is the time between successive peaks. The frequency of damped vibrations can be obtained from the period as

$$f_d = \frac{1}{T_d} \text{ or } \omega_d = \frac{2\pi}{T_d} \quad (12.7)$$

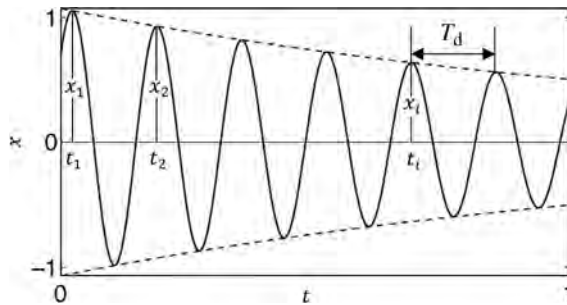


FIGURE 12.2 Transient response of a single-degree-of-freedom system with viscous damping for $\zeta = 0.02$. The dashed lines show the exponential decay envelope. The i th peak displacement $x(t_i)$ occurs at time t_i . The period of oscillation is denoted by T_d .

The exponential decay can be characterized by the *logarithmic decrement* δ defined as

$$\delta = \ln \left(\frac{x(t_i)}{x(t_{i+1})} \right) \quad (12.8)$$

where $x(t_i)$ is the amplitude of the i th oscillation peak. More accuracy can sometimes be obtained using peaks separated by n cycles with the equation

$$\delta = \frac{1}{n} \ln \left(\frac{x(t_i)}{x(t_{i+n})} \right) \quad (12.9)$$

where n is an integer. The nondimensional damping coefficient can be obtained from the logarithmic decrement as

$$\zeta = \frac{\delta}{\sqrt{(2\pi)^2 + \delta^2}} \sim \frac{\delta}{2\pi} \quad (12.10)$$

where the last term is a good approximation when $\zeta \ll 1$.

In summary, the natural frequency f_n and nondimensional damping ratio ζ are the parameters of a single-degree-of-freedom system that characterize its dynamic response. Both of these parameters can be obtained from a time domain measurement of the system's free vibration response.

When a structure's response is dominated by a single frequency, similar time domain measurements can be used to characterize the structure's dynamic response. However, if additional frequencies contribute significantly to the response of the structure, then it is often not practical to separate these frequency contributions from a time domain measurement. Therefore, most vibration measurements are analyzed in the frequency domain. The easiest way to understand frequency domain measurements is to consider the steady-state response of a system to time harmonic excitation as discussed in Section 12.2.2.

12.2.2 Time Harmonic Excitation—Frequency Domain

The natural frequency f_n and nondimensional damping ratio ζ of a single-degree-of-freedom system can also be obtained from its response represented in the frequency domain. Consider excitation of a single-degree-of-freedom system by a harmonic force $F(t) = F_0 \cos(2\pi f t)$. After transients decay to negligible levels, the steady-state response of the system is given by

$$x(t) = X \cos(2\pi f t - \phi) \quad (12.11)$$

This is a vibration at the forcing frequency f with magnitude

$$X = \frac{F_0/k}{\left[\left(1 - \left(\frac{f}{f_n} \right)^2 \right)^2 + \left(2\zeta \frac{f}{f_n} \right)^2 \right]^{1/2}} \quad (12.12)$$

and phase lag between the excitation and the response given by

$$\phi = \tan^{-1} \left[\frac{2\zeta \frac{f}{f_n}}{1 - \left(\frac{f}{f_n}\right)^2} \right] \quad (12.13)$$

The dependence of the response amplitude and phase lag on the forcing frequency is shown in Figure 12.3.

There is a peak in the amplitude plot (Figure 12.3a) that occurs at

$$f_{\text{peak}} = f_n \sqrt{1 - 2\zeta^2} \quad (12.14)$$

When $\zeta \ll 1$, $f_{\text{peak}} \sim f_n$. Therefore, peak frequency is a good approximation for the natural frequency when damping is sufficiently small. The large amplitude response that occurs when a system is excited near its natural frequency is called *resonance*. One reason for determining a structure's natural frequencies is so that operational parameters can be chosen to avoid resonance excitations.

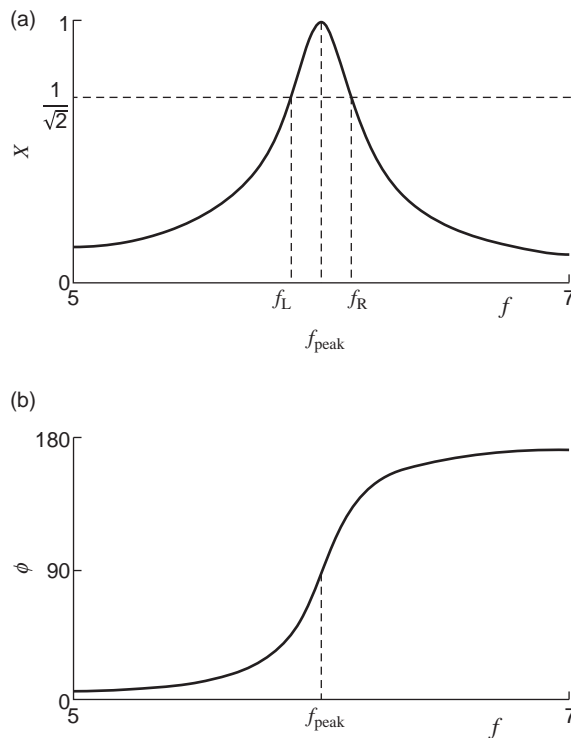


FIGURE 12.3 The (a) amplitude and (b) phase lag of the steady-state response plotted versus forcing frequency f for a single-degree-of-freedom system. The same system parameters were used to obtain the results for Figures 12.2 and 12.3.

The nondimensional damping ratio can be approximated from a plot of response amplitude $X(f)$ versus frequency f (as in Figure 12.3a), using the equation

$$\zeta \sim \frac{f_R - f_L}{2f_{\text{peak}}} \quad (12.15)$$

In Equation (12.15), f_L and f_R are the half power points, defined by

$$X(f_L) = X(f_R) = \frac{1}{\sqrt{2}}X(f_{\text{peak}}) \quad (12.16)$$

The *quality factor* or *Q factor* is a dimensionless quantity often used to characterize the amount of damping in the system. The quality factor

$$Q = \frac{1}{2\zeta} \sim \frac{f_{\text{peak}}}{f_R - f_L} \quad (12.17)$$

is inversely proportional to the damping coefficient. A system with low damping will have a large Q factor. A peak in the frequency response plot for a system with low damping will appear to be proportionally narrower than a peak for a system with high damping.

Figure 12.3b shows that the phase lag is equal to 90° when the response amplitude is near its maximum. The response is nearly in phase ($\phi \sim 0^\circ$) with the force for excitations below the natural frequency and out of phase ($\phi \sim 180^\circ$) for excitations above the natural frequency.

In Section 12.2, a single-degree-of-freedom system was used to illustrate how the natural frequency and damping ratio can be obtained from a plot of response versus time or from a plot of the response amplitude versus excitation frequency. In Section 12.3, this analysis is generalized to systems with more than one degree of freedom to illustrate why typical vibration measurements are analyzed in the frequency domain.

12.3 MULTI-DEGREE-OF-FREEDOM SYSTEMS AND THE FREQUENCY RESPONSE FUNCTION

A two-degree-of-freedom system will be used in this section to describe important phenomena that occur when more than one frequency contributes to a vibration measurement. While most structural components, such as beams, plates, and shells, have distributed mass and elasticity, their dynamic responses can be approximated by discrete multi-degree-of-freedom models. For example, a finite element model of a continuous structure is a discrete multi-degree-of-freedom system that can be used to obtain the important vibration characteristics of the structure (see, e.g., Petyt, 2010 or Friswell and Mottershead, 2010). Because the vibrational behavior of multi-degree-of-freedom systems and continuous structures are similar, concepts discussed here for multi-degree-of-freedom systems can be generalized for continuous systems. Only the most important concepts for vibration testing are introduced in this chapter, so the reader is referred to texts on mechanical vibration, such as de Silva (2000), Kelly (2000), and Rao (2011), for more in depth treatments of the subject.

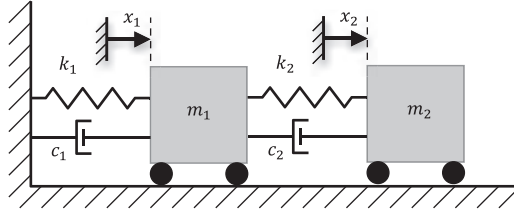


FIGURE 12.4 Schematic of a two-degree-of-freedom system. The two degrees of freedom, denoted by x_1 and x_2 , are the displacements of masses m_1 and m_2 , respectively, measured from their equilibrium positions.

A system with two degrees of freedom, denoted by x_1 and x_2 , shown schematically in Figure 12.4, will be used as a simple illustration of a system with more than one degree of freedom. The coupled equations for a discrete multi-degree-of-freedom system with viscous damping can be written in a matrix form similar to Equation (12.1) as

$$M\ddot{\mathbf{x}} + C\dot{\mathbf{x}} + K\mathbf{x} = \mathbf{F}(t) \quad (12.18)$$

where, for the two-degree-of-freedom example, the unknown displacement vector is

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \quad (12.19)$$

The initial displacements $\mathbf{x}(0) = \mathbf{x}_0$ and the initial velocities $\dot{\mathbf{x}}|_{t=0} = \mathbf{v}_0$ for each degree of freedom are needed to completely define the response of this system to a given excitation force $\mathbf{F}(t)$.

12.3.1 Free Vibration Response—Time Domain

A two-degree-of-freedom system such as the one shown in Figure 12.4 has two natural frequencies, ω_1 and ω_2 , and two corresponding mode shapes, $\boldsymbol{\varphi}^{(1)}$ and $\boldsymbol{\varphi}^{(2)}$. Similar to the one-degree-of-freedom system discussed in Section 12.2, these frequencies and mode shapes are important in characterizing the system's dynamic behavior. For example, the free vibration response of a system with negligible damping can be written as the sum of oscillations at each of these frequencies

$$\mathbf{x}(t) = A_1\boldsymbol{\varphi}^{(1)}\sin(\omega_1 t + \theta_1) + A_2\boldsymbol{\varphi}^{(2)}\sin(\omega_2 t + \theta_2) \quad (12.20)$$

where θ_1 and θ_2 are the phase shifts of the two contributions.

The coefficients A_1 and A_2 and phase shifts θ_1 and θ_2 are determined by the initial displacements and velocities. Note that the first term $A_1\boldsymbol{\varphi}^{(1)}\sin(\omega_1 t + \theta_1)$ represents a vibration at first frequency ω_1 with the relative displacement amplitudes of the two masses determined by the first mode shape $\boldsymbol{\varphi}^{(1)}$. Similarly, the second term, $A_2\boldsymbol{\varphi}^{(2)}\sin(\omega_2 t + \theta_2)$, represents a vibration at second frequency ω_2 with the relative displacement amplitudes determined by the second mode shape $\boldsymbol{\varphi}^{(2)}$. The *mode shape* $\boldsymbol{\varphi}^{(i)}$ is a vector that contains the normalized displacement amplitude of each mass when an undamped system is vibrating freely at the natural frequency ω_i . For the system shown in

Figure 12.4, the masses will be oscillating in phase for the first mode shape and out of phase for the second mode shape.

A damped system will have a similar behavior, but the amplitudes of oscillation will decay with time. The exact nature of the damping in a system is difficult to determine and quantify. Various loss mechanism exist that gradually convert vibrational energy into heat and/or sound, including coulomb friction between contacting parts, fluid–structure interactions, and internal heating due to time varying strains. Fortunately, damping is small in most of the systems so that the exact causes of damping are typically not important, and an approximation can be used to specify losses when characterizing the dynamic response of a structure.

During vibration testing, a nondimensional *modal damping constant* ζ_i corresponding to each important frequency ω_i is typically measured and used to approximate the damping in the system. The damped free vibration response can be written in terms of the modes shapes using modal damping as

$$\mathbf{x}(t) = A_1 e^{-\zeta_1 \omega_1 t} \boldsymbol{\phi}^{(1)} \sin(\omega_{d1} t + \theta_1) + A_2 e^{-\zeta_2 \omega_2 t} \boldsymbol{\phi}^{(2)} \sin(\omega_{d2} t + \theta_2) \quad (12.21)$$

where $\omega_{d1} = \omega_1 \sqrt{1 - \zeta_1^2}$ and $\omega_{d2} = \omega_2 \sqrt{1 - \zeta_2^2}$ are the damped natural frequencies. The general forced response of the system can also be written in terms of the mode shapes, $\boldsymbol{\phi}^{(1)}$ and $\boldsymbol{\phi}^{(2)}$, with the time dependence of each term determined by the excitation. This illustrates that the frequencies, damping coefficients, and mode shapes are important for determining the dynamic response of a system.

These concepts can be extended to a system with N degrees of freedom, where N is a positive integer. The dynamic response of an N -degree-of-freedom system is governed by Equation (12.18) with the displacement vector written as

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_N \end{bmatrix} \quad (12.22)$$

An N -degree-of-freedom system has N natural frequencies, ω_i , $i = 1, 2, \dots, N$. The frequencies are numbered in ascending order, because the lowest frequencies typically have the largest contribution to the dynamic response of a system. The lowest natural frequency is also called the *fundamental frequency*. Design requirements often stipulate that the fundamental frequency be greater than a specified value to avoid resonance behavior during operation. There is a mode shape $\boldsymbol{\phi}^{(i)}$, $i = 1, 2, \dots, N$, corresponding to each natural frequency. Equation (12.21) can be generalized to express the free vibration response of an N -degree-of-freedom system in terms of the natural frequencies and mode shapes as

$$\mathbf{x}(t) = \sum_{i=1}^N A_i e^{-\zeta_i \omega_i t} \boldsymbol{\phi}^{(i)} \sin(\omega_{di} t + \theta_i) \quad (12.23)$$

The free vibration responses of the masses in the two-degree-of-freedom system shown in Figure 12.4 to a specific set of initial conditions (e.g., resulting from an instrumented hammer impact during vibration measurement) are plotted in Figure 12.5. Even for this

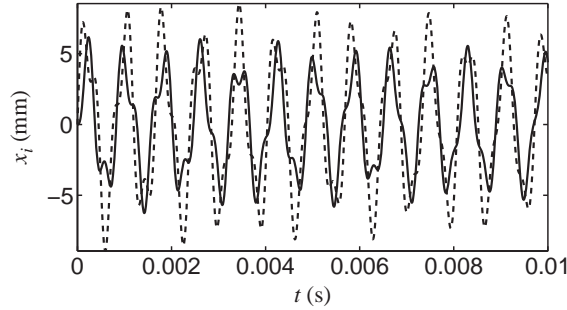


FIGURE 12.5 Transient responses, displacements x_1 (solid) and x_2 (dashed) versus time, for the two-degree-of-freedom system in Figure 12.4 to an initial impact. There are two frequencies contributing to the responses. The higher frequency decays more rapidly so, as time increases, the response will be dominated by the lower frequency.

simple two-degree-of-freedom system, it is not easy to identify both natural frequencies from the time domain responses. As more degrees of freedom are added to the system, the time domain responses become more difficult to interpret, because additional frequencies contribute to the responses. Therefore, typical vibration measurements are plotted in the frequency domain, as discussed in the next section.

12.3.2 Frequency Domain Response

The natural frequencies, ω_i or f_i , modal damping ratios, ζ_i , and mode shapes, $\phi^{(i)}$, of a structure can be obtained from responses represented in the frequency domain. Characteristics of the time domain response can be inferred from these structural parameters using Equation (12.23). Frequency domain plots, therefore, contain the same information as time domain plots, but often in a more useful form.

In vibration testing, frequency domain plots are typically obtained by taking the discrete Fourier transform (DFT) of time domain responses. Alternatively, frequency domain plots can be obtained from measurements of the steady-state response to harmonic excitations across a range of frequencies. The latter method is typically not used because it is more time consuming, but the concepts are useful in interpreting frequency response plots. In this section, discrete Fourier transform methods are used to obtain the frequency domain plots.

Denote the discrete Fourier transform of a response, $x_i(t)$, by $X_i(f)$. The discrete Fourier transform, $X_i(f)$, consists of a set of complex numbers, each of which can either be represented in terms of its magnitude and phase or in terms of its real and imaginary parts. Figures 12.6–12.9 illustrate different representations of the discrete Fourier transform of the free vibration responses for the two masses shown in Figure 12.4. The magnitude and phase of the discrete Fourier transform data are plotted in Figure 12.6a and b, respectively. The real and imaginary parts of the discrete Fourier transform data are plotted in Figure 12.7a and b, respectively. A Nyquist plot of the real and imaginary parts in polar coordinates is shown in Figure 12.8. Finally, the magnitude of the discrete Fourier transform data is replotted using log and dB scales in Figures 12.9a and b, respectively. All of these plot formats can be useful in analyzing vibration measurements.

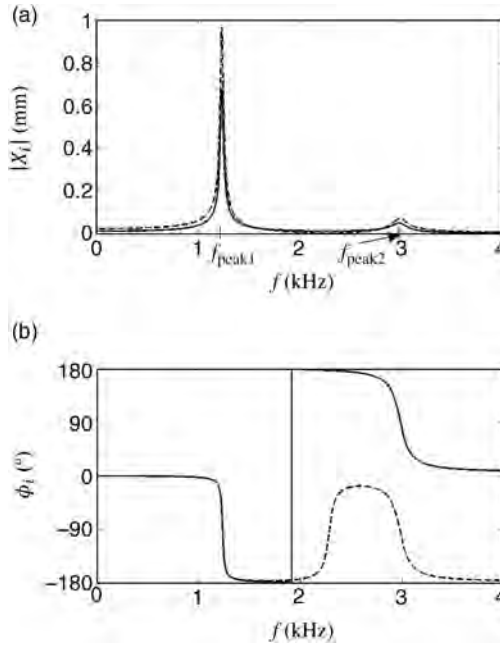


FIGURE 12.6 (a) Amplitude and (b) phase versus frequency of the discrete Fourier transform of the responses x_1 (solid) and x_2 (dashed) of the two-degree-of-freedom system in Figure 12.4 to an initial impact. The two natural frequencies contributing to the responses are easy to identify. The amplitude and phase of the steady-state response plotted versus forcing frequency f for excitation of the two-degree-of-freedom system at the same location would produce a similar plot. The approximately vertical line (at $f \sim 1800$ Hz) is an artifact of phase wrapping, that is, confining the phase between $+180^\circ$ and -180° . The line connects points at $+180^\circ$ and -180° , two representations of the same phase.

Plots similar to those in Figure 12.6 are obtained by graphing the magnitude and phase of the steady-state harmonic responses to a harmonic excitation, as described for a single-degree-of-freedom system in Section 12.2. Therefore, $|X_i|(f)$ and $\phi_i(f)$ can be also be thought of as the magnitude and phase of a steady-state response of mass i to a harmonic force with frequency, f .

There are two peaks in the magnitude versus frequency plot for this two-degree-of-freedom system (Figure 12.6a) at frequencies (denoted by f_{peak1} and f_{peak2}) near the two natural frequencies. If the system is excited at one of these two frequencies, resonance will occur, possibly resulting in a large amplitude response. Therefore, a typical goal of vibration measurement is to identify natural frequencies of the system so that operating conditions can be chosen or the structure can be modified to avoid resonance excitation.

The two natural frequencies of the two-degree-of-freedom system can simply be approximated from the location of the two peaks in the magnitude versus frequency plot (Figure 12.6a), $f_i \sim f_{\text{peak}i}$, analogous to a single-degree-of-freedom system. Also, the relative width of a peak is indicative of the damping for oscillations near that frequency.

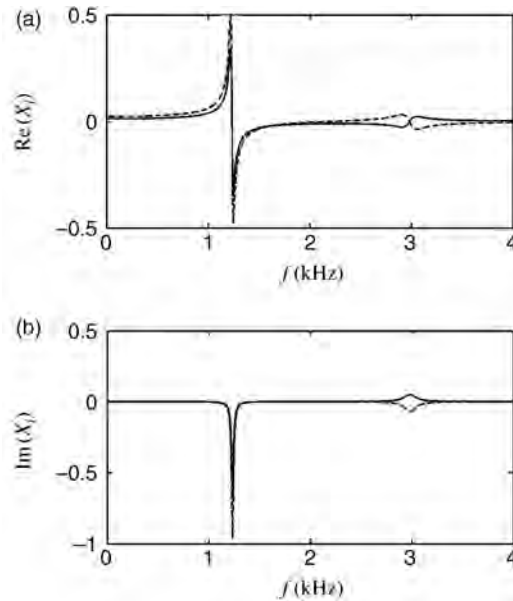


FIGURE 12.7 (a) Real part and (b) imaginary part versus frequency of the discrete Fourier transform of the responses x_1 (solid) and x_2 (dashed) of the two-degree-of-freedom system in Figure 12.4 to an initial impact. The real parts are equal to zero at the natural frequencies, f_1 and f_2 . The peak values of the imaginary parts approximate the amplitude and phase of the mode shapes.

If the two peaks are sufficiently separated, as in Figure 12.6, then the natural frequencies and their nondimensional damping coefficients can be approximated using generalizations of Equations (12.14) and (12.15), respectively. In addition, the peak amplitudes at a resonance can be used to approximate the mode shape for that frequency. Small amplitude peaks can be identified and analyzed more easily if the magnitude is plotted on a log or

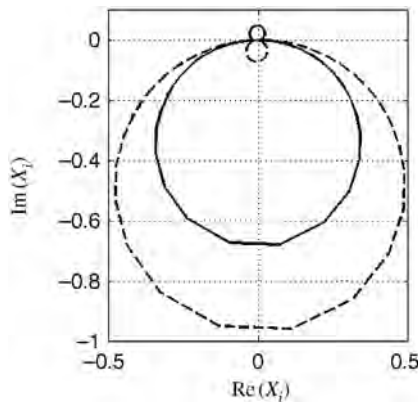


FIGURE 12.8 The Nyquist plot of the discrete Fourier transform of the responses x_1 (solid) and x_2 (dashed) of the two-degree-of-freedom system in Figure 12.4 to an initial impact. The imaginary part is plotted versus the real part, tracing out an approximate circle as the frequency increases through each resonance. The larger and smaller circles correspond to f_1 and f_2 , respectively.

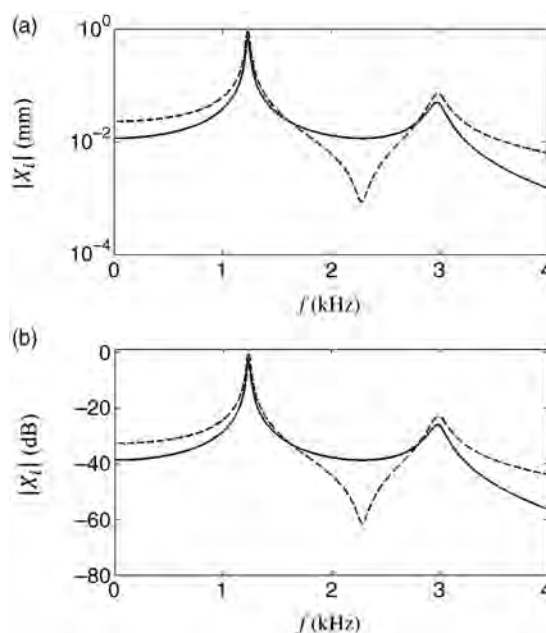


FIGURE 12.9 The amplitude of the discrete Fourier transform of the responses x_1 (solid) and x_2 (dashed) of the two-degree-of-freedom system in Figure 12.4 to an initial impact, on (a) log and (b) dB scales. The dB values are $20 \log_{10}(|X|/X_R)$, with $X_R = 1$ mm. An anti-resonance occurs between the two resonance peaks for the dashed curve, but not for the solid curve. The existence of an anti-resonance depends on whether successive peaks are nearly in phase or out of phase.

dB scale, as shown in Figure 12.9. (See Section 12.5.1.3 for more detail. Techniques for obtaining natural frequencies, modal damping coefficients, and modes shapes from responses with and without significant peak overlap are discussed in Section “Single-Mode Analysis Methods” and “Multi-Mode Analysis Methods,” respectively.)

The phase plot in Figure 12.6b shows that the two masses move in phase with each other for frequencies below the lower resonance frequency. The masses move out of phase with each other for frequencies above the upper resonance frequency. The relative phase changes at a frequency (~ 2400 Hz for this system) between the two resonance frequencies.

Both phase curves in Figure 12.6b cross 90° at the first natural frequency, f_1 , similar to the single-degree-of-freedom system response. Near the second natural frequency, f_2 , the phase curves either cross 90° or -90° . The real part of the response (plotted in Figure 12.7a) is equal to zero when the phase is equal to $\pm 90^\circ$. Therefore, zero crossings of plots of the real part of the response versus frequency can be used to identify natural frequencies of the system. When the amplitude of the real part of the response is zero, the amplitude of the imaginary part is equal to the total amplitude. Therefore, peaks in plots of the imaginary part of the response can be used to estimate the amplitude and phase of the mode shapes, as illustrated in Section “Single-Mode Analysis Methods.”

Frequency is a parameter in the Nyquist plot, and a circle is approximately traced as the frequency increases through each natural frequency, as illustrated in Figure 12.8.

Semicircular curve fits to Nyquist plot data can be used to estimate the natural frequencies and damping coefficients, as discussed in Section 12.5.1.3.

In summary, plots of vibration measurements in the frequency domain can be used to identify resonance frequencies, damping coefficients, and mode shapes that characterize the dynamic response of a structure. Vibration testing can also be used to monitor the condition of a machine by identifying and monitoring frequencies that indicate specific machine failure modes. (See Section 12.6.2 for more detail on condition monitoring.) Therefore, vibration measurement data are typically displayed and analyzed in the frequency domain.

A variety of numerical errors can be introduced by the discrete Fourier transform used to obtain a frequency domain plot from a time domain response. For example, to obtain the accuracy and resolution for the plots in Figures 12.6–12.9 and avoid *leakage* (defined in Section 12.3.3.2), the discrete Fourier transform was taken of a longer time span than plotted in Figure 12.5. A summary of discrete Fourier transforms and techniques used to minimize numerical errors in vibration measurements will be discussed in the next section.

12.3.3 The Frequency Response Function and Discrete Fourier Transforms

The *frequency response function (FRF)* is a quantity often used in vibration measurements and modal analysis. The frequency response function, $H(f)$, is obtained by dividing the Fourier transform of the response, $H(f)$, by the Fourier transform of the excitation, $F(f)$

$$H(f) \stackrel{\text{def}}{=} \frac{X(f)}{F(f)} \quad (12.24)$$

The response data $X_i(f)$ used to generate Figures 12.6–12.9 are essentially scaled frequency response functions, since the Fourier transform of an ideal impulsive excitation is independent of frequency, that is, $F(f)$ was constant for the data in these plots.

If the Fourier transform of the displacement, $X(f)$, is used, then the frequency response function, $X(f)/F(f)$, is called the *receptance* or *dynamic compliance*. The *dynamic stiffness* is equal to the reciprocal of the receptance. This definition for dynamic stiffness is an extension of the definition of static stiffness, $k = \text{force/displacement}$, to time harmonic excitation.

If the Fourier transform of the velocity, $V(f)$, is used, then the frequency response function, $V(f)/F(f)$, is called the *mobility*. The reciprocal of the mobility is the *mechanical impedance*. If the Fourier transform of the acceleration, $A(f)$, is used, then the frequency response function, $A(f)/F(f)$, is called the *accelerance*. The reciprocal of the accelerance is the *apparent mass*.

In theory, frequency response functions depend only on the characteristics of the system and are independent of the time dependence of forcing function used to excite the system. For example, for a single-degree-of-freedom system, the expressions for the amplitude X/F_0 and phase ϕ of the frequency response function, given by Equations (12.12) and (12.13), respectively, are functions of the natural frequency, damping and static stiffness of the system.

A measured frequency response function (defined as the discrete Fourier transform of the response measurement divided by the discrete Fourier transform of the excitation measurement) only approximates the ideal frequency response function because of errors

introduced by finite measurement times, discrete sampling, and measurement noise. An overview of discrete Fourier transforms will next be given and techniques to minimize errors introduced by sampling limitations and discretization error will be discussed. The reader is referred to Brigham (1988), Chapters 14 and 19 of Piersol and Paez (2010), and signal processing references for more detailed treatments of the subject.

Digital signal analyzers obtain and analyze discrete samplings of an analog signal, so it is important to be aware of how this discretization can affect measurement results. Given a constant *sampling rate*, Δt , (time between samples) and the *number of samples*, N , an analog signal $x(t)$ can be represented by the discrete set of samples $\{x(0), x(\Delta t), x(2\Delta t), \dots, x((N-1)\Delta t)\}$. The duration of the sampling interval is $t_{\max} = N\Delta t$. Often, N is restricted to be a power of 2, for example, 1024, because more efficient DFT algorithms, called fast Fourier transforms (FFTs), are then available.

The maximum frequency that can be represented with a sampling rate Δt is determined by the *Nyquist criterion*, $f_{\max} = 1/2\Delta t$. The *frequency resolution*, Δf , that is, the distance between points in the frequency domain, is $\Delta f = (f_{\max}/[N/2]) = (1/t_{\max})$. The discrete Fourier transform consists of only $N/2$ points, even though the time domain signal had N data points; however, each data point in the frequency domain is complex. Complex numbers can be represented in terms of an amplitude and phase or a real and imaginary part, so discrete Fourier transform still has N pieces of information.

Digital signal analyzers have a variety of options that can be selected to improve the accuracy of discrete Fourier transforms and frequency response functions. The most important options will be discussed here.

12.3.3.1 Filtering—to Minimize Aliasing The Nyquist criterion states that at least two samples per cycle are required to identify a frequency. A frequency greater than $f_{\max} = 1/(2\Delta t)$ that contributes to an analog signal will be undersampled when a sampling time step equal to Δt is used. As a result, this contribution will appear to be at a lower frequency, as illustrated in Figure 12.10. If you use a modern digital signal analyzer, then a combination of analog filtering, oversampling, and digital filtering will be used automatically to minimize aliasing. However, you may need to implement filtering techniques

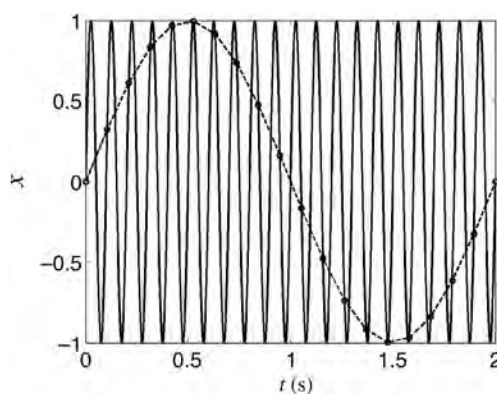


FIGURE 12.10 Illustration of aliasing. The 10 Hz signal (solid line) is sampled at 9.5 Hz (dashed line). The solid dots represent the sample points. The frequency measured from the sampled signal appears to be at approximately 2 Hz. A discrete Fourier transform of the sampled signal would have a peak at approximately 2 Hz, rather than at the actual frequency of 10 Hz!

yourself when using a digital oscilloscope or computer application to obtain the discrete Fourier transforms of a signal. To avoid aliasing, analog signals must be low pass filtered to remove contributions from frequencies above $f_{\max} = 1/2\Delta t$, before the discrete Fourier transform is applied. No filter is ideal. However, a higher order filter will be better at minimizing the effects of aliasing.

12.3.3.2 Windowing—to Minimize Leakage or the Effect of Noise The discrete Fourier transform assumes periodic signals. If the sampling interval is denoted by $t_{\max} = N\Delta t$, then $x(t + t_{\max}) = x(t)$ is assumed, as illustrated in Figures 12.11–12.13. That is, a discrete Fourier transform of a signal in the time range $0 \leq t \leq t_{\max}$ assumes that the signal repeats in the time range $t_{\max} \leq t \leq 2t_{\max}$ and again in the time range $2t_{\max} \leq t \leq 3t_{\max}$, and so on. The periodic extension can introduce discontinuities in the signal amplitude or slope.

The periodic extension of the sampled data in Figure 12.11a is shown in Figure 12.11b. The periodic extension did not introduce any discontinuity at the beginning or end of the data, because an integer number of periods occurred in the sampling interval. Therefore,

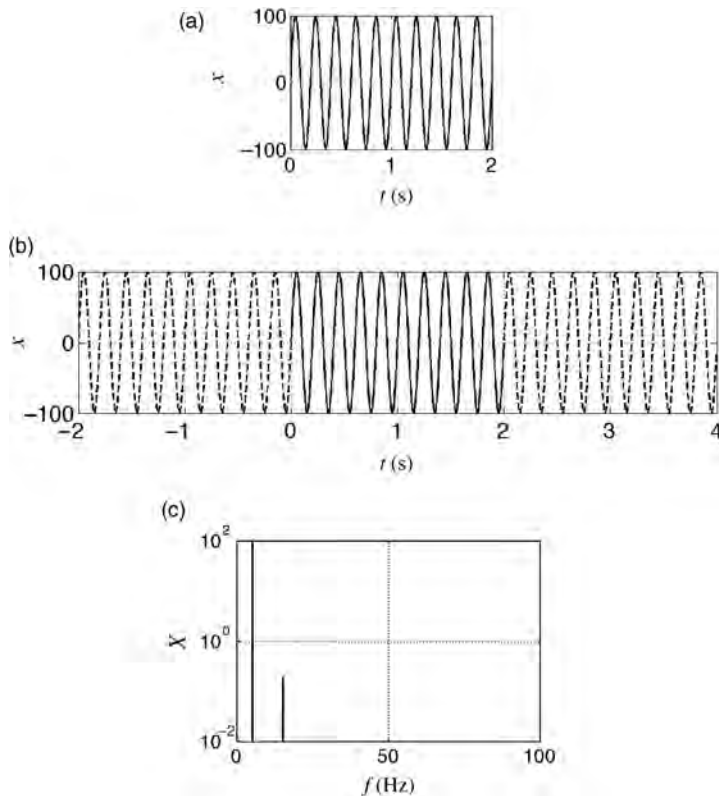


FIGURE 12.11 Illustration of a discrete Fourier transform of an essentially harmonic signal that does not exhibit leakage, including (a) the sampled data, (b) the periodic extension of the sampled data, and (c) the discrete Fourier transform of the sampled data. The discrete Fourier transform has two sharp peaks at the fundamental frequency of 5 Hz and the 2nd harmonic at 15 Hz.

the discrete Fourier transform of Figure 12.11c shows no evidence of leakage, that is, the two peaks are clearly visible and the amplitude everywhere else is many orders of magnitude smaller than the peaks. The largest peak is at 5 Hz, the fundamental frequency. The smaller peak at 15 Hz is due to a contribution at the 2nd harmonic that slightly distorts the signal. The sampled data in Figure 12.12a have a fundamental frequency of 4.75 Hz, so that the sampling interval does not contain an integer number of periods. In this case, the periodic extension introduces a slope discontinuity, as shown in Figure 12.12b. A large range of frequency components are needed to represent this discontinuity, raising the baseline of the frequency domain plot, as shown in Figure 12.12c. The distortion of the discrete Fourier transform due to discontinuities introduced by the periodic extension is called *leakage*. In Figure 12.12c, the 2nd peak is obscured due to leakage.

In vibration measurements, leakage may be a problem when the excitation is continuous, such as when using a shaker driven by random noise or using the rotation of a component of the structural system as the excitation. In these cases, leakage can be minimized by multiply the time domain signal by a weighting function before taking the discrete

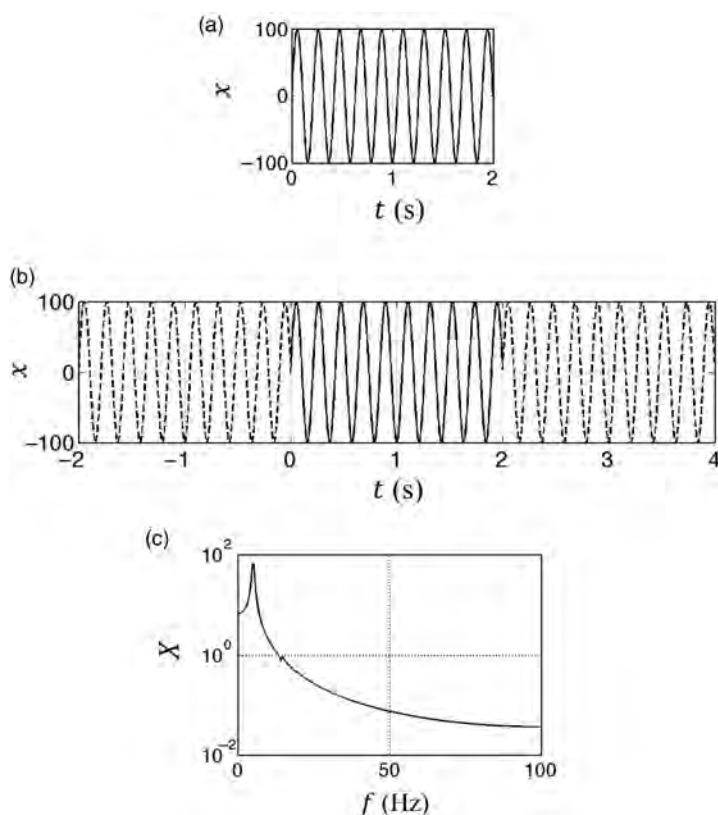


FIGURE 12.12 Illustration of a discrete Fourier transform of an essentially harmonic signal that does exhibit leakage, including (a) the sampled data, (b) the periodic extension of the sampled data, and (c) the discrete Fourier transform of the sampled data. The discrete Fourier transform has a peak at the fundamental frequency of 4.5 Hz but the peak at the 2nd harmonic, 13.5 Hz, is below the raised baseline.

Fourier transform. This process is called *windowing* and the weighting function is called the *window function*.

There are a large variety of windowing functions, but two of the most popular are the *Hanning window* and the *flat-top window*. Both of these functions weight the signal at the beginning and end of the time interval less than the central portion of the signal so that the leakage due to the discontinuity introduced by the period extension is reduced. The flat-top window allows better accuracy in determining the peak amplitude, whereas the Hanning window has better frequency resolution. One caution when using windowing: a window introduces an artificial source of damping that will increase the width and decrease the amplitude of the peaks in the response spectrum, so estimates of damping obtained from the frequency domain plot will be less accurate.

Figure 12.13a shows the sampled data of Figure 12.12a after the Hanning window shown in Figure 12.14 was applied. Figure 12.13b shows that the periodic extension does not introduce a discontinuity for the windowed data. The discrete Fourier transform of the windowed function plotted in Figure 12.13c shows that leakage has been reduced by the windowing process so that the 2nd peak is now visible. However, both peak amplitudes have been reduced and the peaks are wider than those in Figure 12.11c, obtained for a comparable set of sampled data.

Windowing should be used with care, because windowing can cause serious distortion of discrete Fourier transforms. Windowing typically should not be used for sine swept signals produced by a shaker. Sine swept excitations are designed to be periodic, with a period equal to the sampling interval. The amplitude of the sine sweep is typically not zero at the beginning and end of the sampling interval. Therefore, use of a nonuniform window will alter the amplitudes of the frequencies at the beginning and end of the sampling interval resulting in distortion and error in the discrete Fourier transform of the sine swept signal.

Impulse responses obtained by exciting a system with an instrumented hammer often begin at rest and decay to near zero amplitude within the sampling interval. In this case, a Hanning or flat-top window may also cause distortion and error in the frequency response plot if the window alters the initial response. Instead, most digital signal analyzers have force and exponential response windows that can be used to reduce leakage or to minimize the effect of noise that may be present after the excitation or response signals have decayed.

If windowing is used for impact testing, a force window should be applied to the excitation signal and an exponential decay window should be applied to the response signal. The force window should essentially be equal to 1 during the time of impact to preserve the entire force signal, but should be equal to 0 outside of this time range so that any extraneous noise would be eliminated. The exponential response window similarly can be used to minimize the effects of extraneous vibrations that may occur after the impact response has decayed below the noise level. In addition, an exponential window can be used to eliminate leakage if the vibration response has not decayed to zero within the sampling duration. This exponential window also will introduce artificial damping, so damping calculated from the frequency spectrum of exponentially windowed data will be overestimated.

12.3.3.3 Ranging—to Minimize Discretization Effects Not only is the signal sampled only at discrete times, but the voltage amplitude of each sample is also converted (using an analog-to-digital converter) to a discrete number so that it can be stored digitally for further analysis. This discretization of the signal amplitude can also introduce error into

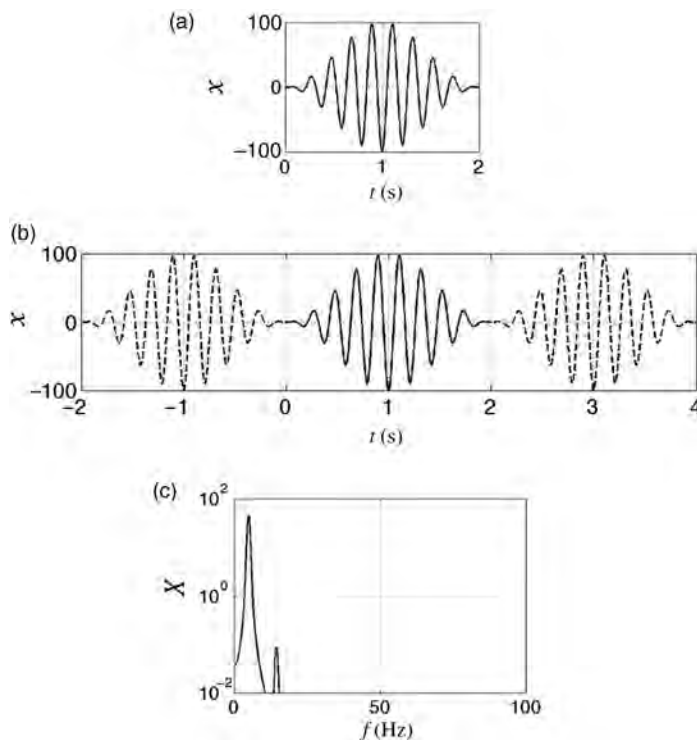


FIGURE 12.13 Illustration of the use of a Hanning window to minimize leakage, including (a) the windowed sampled data, (b) the periodic extension of the windowed sampled data, and (c) the discrete Fourier transform of the windowed sampled data. The sampled data are the same as in Figure 12.12a, weighted by the Hanning window shown in Figure 12.14. The discrete Fourier transform has two peaks at the fundamental frequency of 4.5 Hz and the 2nd harmonic at 13.5 Hz. The windowing function caused a decrease in the peak amplitude as well as a broadening of the peaks, compared to Figure 12.11c.

the discrete Fourier transform. The *input range*, that is, the voltage range that is converted, is defined before a measurement is taken. It is ideal to set the input range on a digital signal analyzer to approximately half the amplitude of the maximum signal (referred to as *half-ranging*) to obtain the best resolution from the analog-to-digital

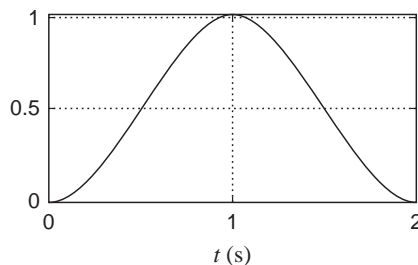


FIGURE 12.14 The Hanning window used to obtain the plots in Figure 12.13.

conversion. It is also important to avoid signals with amplitudes that exceed the input range (often referred to as *overloading*). Overloading would cause peaks of the time domain signal to be out of range, and therefore the peaks would be clipped, causing distortion of the discrete Fourier transforms.

Because the range for the display of the signal on many digital signal analyzers can be set independently from the input range, the input range on a digital signal analyzer is not as immediately apparent as it is on most oscilloscopes. Therefore, analyzers typically have indicators that identify when the signal amplitude exceeds half the input range (half-ranging) and when the signal amplitude exceeds the input range (overloading).

Often digital signal analyzers have an *autorange* option that automatically increases or decreases the input range until overloading is just prevented. This option works well if the amplitude of the signal is fairly constant, such as when a swept sine excitation is continuously repeated. However, if an instrumented hammer is used for excitation, then the autorange option is not useful if it decreases the range to a low level in between hammer hits. When using an instrumented hammer, an *autorange up only* option can be used to set the correct input range on each channel (receiving either hammer or response signals). This is done by producing a number of consistent hammer hits, so that the input ranges for both the excitation and the response channels are increased until overloading is just avoided. If a new hammer tip is used or the impact location changes, then this procedure should be repeated to set new input ranges for all channels, starting again from a low level.

12.3.3.4 AC Coupling—to Minimize DC Bias Effects A signal with a small amplitude oscillation could have poor resolution after analog-to-digital conversion, if it is superimposed on a large constant (DC) bias. An input range large enough to include the DC bias may have inadequate resolution to capture small variations in the signal. *AC coupling* with a coupling capacitor can be used to remove the DC bias. The capacitor acts as a high pass filter, filtering out the low frequencies. Therefore, AC coupling not only will remove the DC bias, but may also degrade the low frequency components of the signals. The choice between using AC and DC coupling depends on the frequency range of the vibration of interest. For example, when measuring the transmission of floor vibrations to a structure for frequencies on the order of 1 Hz, DC coupling is appropriate. However, when measuring the dynamic response of a machine for a frequency range above 50 Hz, AC coupling may be more appropriate.

12.3.3.5 Averaging—to Minimize Noise Effects The frequency response function assumes that the measured response of a structure is a direct result of the measured excitation. However, nearby machinery and people may cause additional structural vibrations. The measured response will be the sum of the desired response due to the excitation and the response due to environmental noise. *Averaging* a set of frequency response functions can increase the signal to noise ratio, by reducing the effects from random contributions. The input range should be established for all measurements before averaging begins. Analyzers typically have an option to automatically reject a data set if overloading occurs during the acquisition of the data. This option can save valuable time because without this option, all prior measurements would need to be repeated if an overload occurred.

The *coherence function* is used during averaging as an indicator of frequency response functions repeatability. The coherence function is a measure of how much of the output signal is caused by the measured input signal, assuming the excitation and response points

and directions remain the same. The coherence function is plotted versus frequency and its values range between 0 and 1. The coherence function is equal to 1 if all of the output is entirely caused by the input and equal to 0 if none of the output is caused by the input. For example, if the background vibration (noise) in the test environment is high relative to the excitation source, then the coherence will be low. Low coherence can also occur when using an instrumented hammer, if the instrumented hammer does not strike the same point on the structure in the same direction each time. Practice with the instrumented hammer can improve coherence in this case. The coherence should be approximately equal to 1 in the frequency ranges of interest (e.g., near resonance peaks). Lower coherence that can occur in regions of lower response amplitude may be acceptable if no response peaks of interest occur in these regions.

For illustration, the coherence and amplitude of a frequency response function are shown in Figure 12.15 for one set of averaged data. The data were obtained from accelerometer measurements of the transverse response of a plate excited by an instrumented hammer impact. In Section 12.5, this plate structure is used as an example when discussing experimental modal analysis techniques. The coherence is approximately equal to 1 at each of the peaks in the frequency response function, indicating good repeatability of the peak locations and amplitudes. However, the coherence is lower at frequencies below 10 Hz due to inconsistencies in foam pad used to support the plate and low accelerometer sensitivity at low frequencies. The coherence is also lower at the frequency corresponding to the antinode. Because no peaks of interest occur in these regions, the lower coherence would not significantly impact analysis of the response.

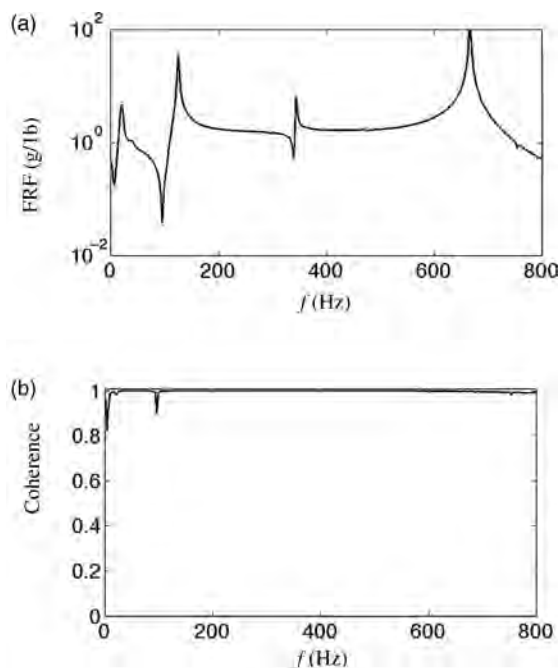


FIGURE 12.15 Measured (a) amplitude and (b) coherence versus frequency for the frequency response function (FRF) obtained for transverse vibrations of plate.

12.4 VIBRATION MEASUREMENT EQUIPMENT AND TECHNIQUES

Practical aspects of vibration measurement are considered in this section, with specific emphasis on the equipment used to obtain and analyze vibration data. Even with the best equipment and analysis techniques, however, it can be difficult to obtain meaningful data, especially for complex systems in noisy environments. Therefore, it is important to first clearly define the purpose of the measurements and to carefully consider the experimental setup. An understanding of fundamental vibration theory is needed to choose and apply the correct equipment and analysis methods for a particular purpose as well as to obtain valuable conclusions from the vibration testing results.

Vibrations measurements are obtained for a wide variety of purposes including, characterizing a structure's dynamic behaviors, monitoring the condition of a machine, and testing the durability or lifespan of a component. Because the desired goals for vibration measurements are so diverse, it is not possible to cover all possible scenarios. However, the types of equipment used for vibration testing are similar for the various applications. Therefore, the objective of this section is to give an overview of the most common equipment used to obtain and analyze vibration measurements.

Experimental vibration testing involves measuring structural motions in response to input loads. The response depends on both the structure and the excitation, as shown schematically in Figure 12.16. Generally, one of three things is under investigation in vibration testing: (1) the structural system itself is being characterized; (2) the input that excites the system is being determined; or (3) the system's response to the excitation is being observed.

Vibration measurements can be used to obtain dynamic characteristics of a product or structural component, necessary for further product development. The equipment needed to measure the dynamic characteristics of a structure include (1) an excitation source, (2) a device to measure the system's response, and (3) an analyzer to process the data to obtain the desired information.

Modal analysis methods for characterizing the dynamic behavior of a structure are discussed in Section 12.5. For modal analysis, a controlled excitation is applied to a structure and both the vibration response and the applied excitation are simultaneously measured. A signal analyzer is used to collect and analyze the measurement data to determine the dynamic characteristics of the structure. The quality of the data will depend on both the excitation type (impact, swept harmonic, random, burst, and so on) and the response measurement type (displacement, velocity, acceleration, strain, and so on). Excitation and response measurement types will be discussed in Sections 12.4.1 and 12.4.1.1, respectively. Signal analyzers will be discussed in Section 12.4.2.

Because a test structure's dynamic characteristics are affected by how the structure is supported, the support fixture is also an important consideration for vibration testing.

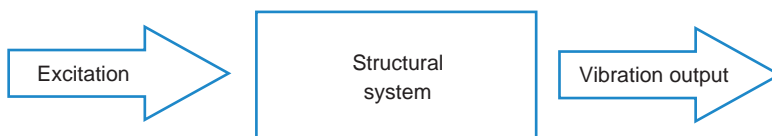


FIGURE 12.16 Schematic of the components of a test system for characterizing the dynamic response of a structure.

To obtain natural frequencies that closely match the natural frequencies of a structure under operating conditions, vibration measurements should be made in situ whenever possible. When in situ testing is not possible, supporting structures should either be designed to mimic the in situ constraints or to approximate the constraints used in analytical or numerical models of the structure.

Some systems, such as an airplane or satellite in flight, typically operate without any boundary constraints. To approximate free boundary conditions, the structure can be suspended from elastic cords or placed on a soft cushion. As a rule of thumb, to mimic free boundary conditions, the support system stiffness should be low enough so that the rigid body mode frequencies of the structure on its supports are less than approximately 10% of the lowest deformable mode frequency of the structure.

Other structures, such as a grinding tool in a chuck, are held almost rigidly. Perfectly rigid boundary conditions are difficult to approximate for many structures, because of the compliance of the fasteners used to attach the test component to the supporting fixture and the vibration of the supporting structure itself. The motion of the supporting structure at the attachment points can be measured; however, a low response does not necessarily imply that the support has the same effect as a rigid support.

Finite element models can be used to determine the effects of different boundary constraints on the dynamic characteristics of a component. To correlate measurement data with finite element predictions, support fixtures for the vibration measurements are often designed to mimic the either completely free or fixed boundary conditions typically used in finite element models. Once the finite element model predictions and measured responses for a component agree to sufficient accuracy, the model can be used to investigate how other boundary condition will affect the component's dynamic behavior. In particular, the model of the component can be attached to a model of the overall system to predict the component's in situ behavior. For example, a finite element model of an airplane wing can be verified experimentally with free boundary conditions and then used in a complete finite element model of an airplane to determine the wing's in flight response.

Vibration measurements can be used to characterize the excitation experienced by a structure or component when in operation. Knowledge of the excitation may be needed for a number of purposes. For example, more realistic predictions of product life and durability may be obtained from fatigue testing when the excitation applied during the test more accurately represents the in-service excitation. A variety of multiaxis shaker tables have been developed so that different vibration environments can be closely approximated.

Vibration testing can be used to measure the dynamic motion of a machine during operation or to monitor its performance over a period of time. Measurement of the machine vibration can be used to identify sources of vibration problems or operating conditions that minimize vibrations. As discussed in Section 12.6.2, some machine component failure can be identified in their early stages, from measurements of machine vibrations generated by defects in the component. Therefore, vibration monitoring can be used as a sensitive indicator of machine failure.

For any of these vibration measurement applications, there must be an excitation source, a system response measurement, and analysis of the measurement data to obtain meaningful conclusions. While the purposes of the vibration measurements may differ, the equipment used for many of the measurements are similar. In this section, an overview will be presented of some of the most commonly used vibration measurement equipment and some of their advantages and disadvantages.

12.4.1 Excitation Sources

Many types of excitation equipment are available, including instrumented hammers, shakers, and multiaxis tables. Only a brief overview of this equipment will be given here. Impact hammers and electrodynamic shakers will be described in more detail because these are perhaps the most common excitation sources and they can produce the most commonly used types of excitations.

12.4.1.1 Instrumented Hammers—Impulse Excitation An instrumented hammer is used to strike the test structure to produce an impulse excitation. Instrumented hammers have a force transducer near their impact tip, so the time history of the force imparted to the structure can be measured and recorded. This force measurement can be used to obtain frequency response functions for the test structure, as described in Section 12.3.3.

Instrumented hammers are perhaps the easiest source of excitation to use, because no additional fixturing is required. A test structure is excited simply by striking it with the instrumented hammer at the desired load location and in the desired direction. Responses for an array of excitation locations can quickly be measured by changing the impact location of the instrumented hammer.

The size of the hammer and stiffness of the hammer's impact tip control the duration of the impact and thus the frequency content of the excitation. The amplitude of the applied impulse is controlled by the mass and velocity of the hammer impact. Often hammer mass can be adjusted by changing weights attached to the hammer. Two instrumented hammers are shown in Figure 12.17, along with a variety of hammer tips. Time and frequency plots for soft, medium, and hard tips are shown in Figure 12.18a and b, respectively. The more compliant (softer) the tip, the longer the impact time. Conversely, the



FIGURE 12.17 Two instrumented hammers in the mid-size range with additional impact tips that can be used to vary the impulse duration and thus the frequency range of the excitation. Each instrumented hammer has a force sensor adjacent to the attached tip (outlined in white).

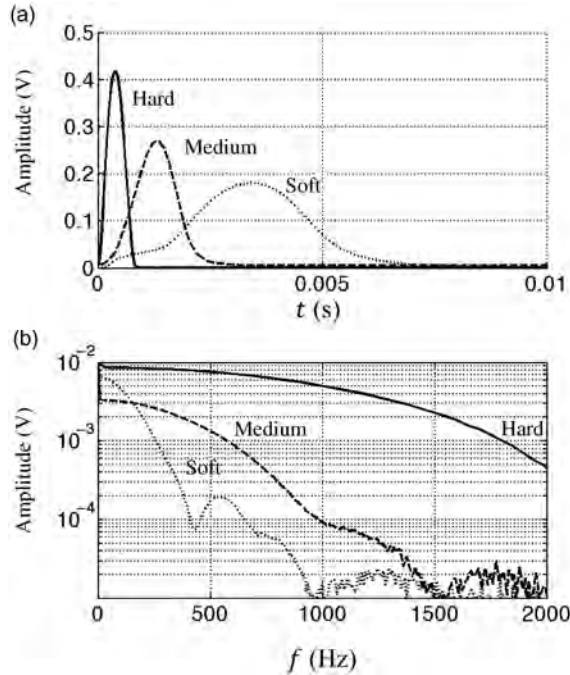


FIGURE 12.18 Signal amplitude versus (a) time and (b) frequency generated by an instrumented hammer striking a rigid surface, with a soft (dotted), medium (dashed), and hard (solid) tip. The usable frequency ranges for this hammer with the soft, medium, and hard tips are from 0 Hz to approximately 400 Hz, 1000 Hz, and 2000 Hz, respectively. The signal amplitudes depend on the impact strength. The noise level is approximately 2×10^{-5} , so signal below this level is not meaningful.

stiffer the tip, the shorter the impact time. Because higher frequencies are needed to represent shorter pulses, the stiffer tip imparts energy over a broader frequency range than the soft tip as shown in Figure 12.18b. The hammer size and tip stiffness are chosen based on the desired impact energy and structural frequency range of interest. Large hammers can impart higher energy to the test structure and therefore are used when testing larger/stiffer structures.

Instrumented hammers have two main limitations. First, the energy imparted to the structure by the hammer impact may be insufficient to excite the modes of interest, resulting in a low signal to noise ratio. Second, it is sometimes difficult to get consistent hammer hits, especially in recessed locations. Double hits, where the hammer actually bounces on the structure instead of hitting it cleanly, typically cause obvious artifacts in the frequency response function and therefore can easily be identified. A double hit can be positively identified if a double peak exists in the time response. Other variations in the location and/or direction of the hammer hit are typically not as obvious when observing the frequency response functions. Therefore, frequency response functions obtained using an instrumented hammer should be averaged, so that the coherence (see Section 12.3.3.5) can be used as a measure of the repeatability of the hammer impacts. A high coherence (~ 1) indicates consistent impact and response data.



FIGURE 12.19 Two shakers, a 4 lbf (a) and a 40 lbf (b) shaker, are shown. A stinger can be attached to the driven section of the shaker at the central threaded hole. The 4 lbf shaker is shown attached to a mount that allows rotation of the shaker to align the direction of motion (along the axis of the shaker) with the desired excitation direction on the test structure.

12.4.1.2 Shakers—Harmonic, Swept Harmonic, Random, or Burst Excitation A shaker produces a unidirectional force when appropriately attached to a test structure. Two shakers of differing size are shown in Figure 12.19. The size of the shaker determines its frequency and forces ranges, with smaller shakers typically having a higher frequency range, but a lower force range. Frequencies can range from a few Hertz to tens of kiloHertz and forces can range from a few Newtons to hundreds of kiloNewtons.

Shakers can produce a wide range of excitation time histories, including harmonic, swept harmonic, burst, or random. The desired excitation signal is specified using a vibration controller, often within the digital signal analyzer, and amplified to drive the shaker. If the test structure's dynamics response is only required at a small number of distinct frequencies, then single frequency harmonic excitation can be used. However, single frequency harmonic excitation is inefficient if the response is required over a broad frequency range. A swept harmonic excitation is often used to obtain the structure's response over a given frequency range. Just as the name implies, swept harmonic excitation applies a harmonic force with a frequency that sweeps through a frequency range, starting at the specified lower frequency and continuously increasing until the upper frequency is reached. A swept sine excitation for the frequency range from 0 to 800 Hz is shown in Figure 12.20. The amplitude versus frequency plot (not shown) indicates that the signal amplitude is approximately constant over this frequency range. Random and burst (burst swept harmonic or burst random) excitation can also be used to obtain the frequency response. Choice of excitation signal depends on what sources are available and how similar the source is to expected forces on the test structure during operation.

Although the ideal frequency response function (defined in Section 12.3.3) for a linear system depends only on the test structure's dynamic characteristics, frequency response functions obtained using different excitation methods will differ somewhat due to errors introduced by limitations in the measurements. To obtain the most accurate results,

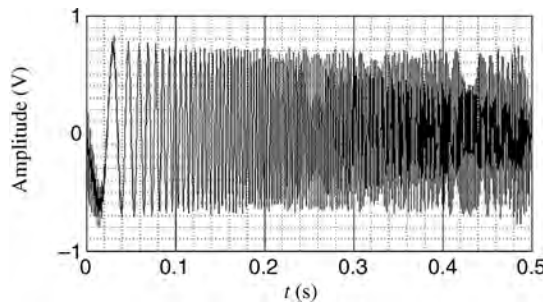


FIGURE 12.20 Source signal versus time for a swept sine from 0 to 800 Hz. A discrete Fourier transform of this signal would indicate that the amplitude is nearly constant over the specified frequency range from 0 to 800 Hz. The signal smoothly transitions from the high to low frequency so that the periodic extension of the signal does not introduce a discontinuity.

appropriate parameters need to be specified for data acquisition and for the discrete Fourier transform, as described in Section 12.3.3. In particular, an appropriate windowing function must be used, as described in Section 12.3.3.2. Swept sine and burst excitation require no (sometimes called uniform or rectangular) windowing, while random excitation typically requires a Hanning or similar windowing function.

Electromagnetic (also referred to as electrodynamic) shakers are the most commonly used shakers in vibration testing. Other types of shakers are available and may be better suited for specific applications. For example, electrohydraulic shakers have a lower frequency range and can be used to apply a static load superimposed on the dynamics excitation.

In an electrodynamic shaker, the output from the amplifier passes through a coil located within a magnetic field, producing a force proportional to the current. The force actually applied to the structure (needed to determine the frequency response function of the structure) depends on the response characteristics of the shaker and structure, so *a separate force transducer must be used* to measure this force. This force transducer should be located at the point of excitation on the structure to ensure that the force applied to the structure is accurately measured. In addition, for a small test structure, a small force transducer should be chosen so that the added mass does not significantly affect the structure's dynamic response.

Care must be taken to ensure that only the desired excitation force is applied to the test structure. Often this requires extra fixturing both to support the shaker and to connect the shaker to the structure.

The connection between the shaker and the structure should allow only the desired force, along the direction of the shaker motion, to be transmitted to the structure. A shaker is much stiffer in the direction perpendicular to its motion than in the direction of its motion. Therefore, connecting the shaker directly to the structure will effectively stiffen the structure, detrimentally affecting the frequency response measurements. Typically a shaker is connected to the force transducer attached to the test structure through a *stinger* or push-rod, as illustrated in Figure 12.21. A stinger is a short rod or wire (perhaps ~ 1 cm in free length and 1 mm in diameter), designed to be stiff in the axial direction but flexible in the transverse direction so that essentially only the shaker's axial forces are transmitted to the test structure. The stinger should be stiff enough so that its resonance frequencies

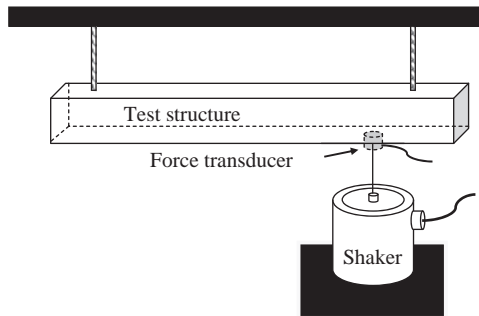


FIGURE 12.21 Experimental setup showing a test structure suspended from above with elastic cords to approximate unconstrained boundary conditions. A shaker, used to excite the structure, is grounded from below and connected to the test structure by a push rod. A force transducer, mounted between the push rod and the structure, is used to measure the excitation force transmitted to the test structure.

are high enough above the frequency range of interest so that they do not interfere with the vibration measurement. In addition, if the path to the desired excitation point is obstructed, then a stiff, but light-weight extension may be required to connect the shaker output to the structure.

The base of the shaker should be supported so that it is isolated from the test structure and does not create an additional excitation force on the structure. Because the frequency response function assumes that the only excitation is that measured by the force transducer, all other excitation of the structure must be minimized for accurate frequency response function measurements. The base of the shaker can be grounded, for example, attached to the floor, while the test structure is isolated from floor vibrations. For example, larger structures can be supported on a vibration isolation system and smaller structures can be placed on a foam pad or hung from the ceiling. Alternatively, the base of the shaker can be suspended so that no force is transmitted to any surrounding structure, except to the test structure through the stinger. However, to create the same excitation force amplitude, a larger shaker mass is required for a suspended shaker than for a grounded shaker.

12.4.1.3 Multiaxis Tables Multiaxis shaker table test systems are designed to excite up to six (three translational and three rotational) degrees of freedom simultaneously, by controlling the output of multiple shakers working in unison. Typically a stiff vibration table is used to apply ground excitation to the test structure, as shown in Figure 12.22. Multiaxis loading not only reduces testing time but also can produce more relevant data.

The shaker table motion can be controlled to replicate actual loading conditions for a structure thereby producing vibration test results that may be more applicable to the actual operation of the structure. For example, operational deflection shapes obtained using realistic multiaxis excitations may indicate points of collision that could be missed if only single axis excitations are applied. In addition, multiaxis excitation can more accurately reproduce operational stresses for accelerated life and durability testing.

However, developing the load data for multiaxis testing is often not a trivial task. Typical in situ loading conditions must be measured and the relevant features identified, often using power spectral density representations of the data. The data also must be filtered and



FIGURE 12.22 A multiaxis shaker table showing ground excitation of an automobile seat. From: http://en.wikipedia.org/wiki/File:Multi_Axis_Shaker_Table.jpg.

conditioned to meet the requirements defined by the physical limitations of the multiaxis table.

12.4.1.4 Operational Loads The dynamic response of a machine to purely operational loads can also be measured and analyzed. Vibration response to operational loads can be used to obtain information about the dynamic characteristics of a machine (see Section 12.5.2) or to monitor the performance of a machine (see Section 12.6.2). The loads experienced by a machine during operation can be generated either externally or internally.

Vibrations are generated internally by moving parts within the machine. Any rotating shaft will have some unbalance that will produce a harmonic force at the shaft rotation frequency. The response of the machine to a range of excitation frequencies can be obtained by varying the shaft rotation rate and taking repeated measurements. Other moving parts such as bearings, reciprocating pistons, and gears will produce more complex loadings.

External loads can be generated by a wide range of sources and transmitted to the machine through the ground supports or through the air. A machine on a factory floor may be excited, for example, by vibration of nearby machines or motions generated by people working in the surrounding area. If the structure of interest is a part of a machine, then the vibration of mounting surface caused by the machine's operation may be considered as the external load of interest. Ground motion caused by earthquakes or nearby explosions may be the relevant excitation, especially for civil structures. Wind loadings may also be important sources of vibration for civil structures and other outdoor structures such as antennas and windmills. Automobiles, trucks, and other land vehicles will

experience a random excitation generated as they are driven over rough terrain. Components in a car crash or dropped object will experience shock loadings. Measurement of a structure's response to loads it experiences in service may provide information to improve its performance or to suggest design modifications.

12.4.2 Response and Force Measurement

In this section, a brief overview of vibration measurement equipment will be given. A device to measure motion is typically required for any vibration test and a wide range of contact and noncontact devices are available. Measurement of additional quantities, such as force, pressure, or sound may also be required for vibration testing.

Transducers are devices that convert a mechanical measurement to an electrical signal that can be used as input to a signal analyzer. The most important characteristics to consider when choosing a transducer for vibration measurement are the (1) sensitivity, (2) resolution, and (3) operating frequency range.

Sensitivity is defined as the ratio of the electrical output to the mechanical input. For example, for a force transducer, the sensitivity may be specified in millivolts per Newton (mV/N). A good transducer has a high sensitivity in the intended measurement direction and a low *transverse sensitivity*, that is, a low sensitivity normal to the intended measurement direction.

The *resolution* of a transducer is the smallest increment of the mechanical input that produces a discernable change in the electrical output. Measurement resolution depends on the equipment used to capture the signal as well as on the transducer and can be limited by noise levels in the measurement system.

Ideally, a transducer's sensitivity would be constant over its entire operating range, although in practice some variation of sensitivity will occur. The *operating frequency range* is specified as the frequency range over which the sensitivity varies less than a specified percent from its rated value. A sensor's natural frequencies essentially determine its operating frequency range. Sensors are typically designed so their natural frequencies are far from the operating frequency range to avoid resonance effects that would cause variations in the sensor's sensitivity.

Environmental factors such as temperature and humidity may affect the operation of a transducer. Often transducers include internal electronic compensation to minimize environmental effects on their output signal and to reduce other undesirable effects such as drift. Specifications for environmental factors should be consulted, especially when choosing a sensor to be used in extreme environments or environments that vary over time.

Sensors should be calibrated periodically, especially when high precision data are required. A sensor can be calibrated by measuring its response to a known calibration excitation or by comparing its response to that of another sensor of known calibration (see, e.g., Chapter 11 of Piersol and Paez, 2010).

12.4.2.1 Relating Acceleration, Velocity, Displacement Acceleration, velocity, and displacement are the most common measures of motion, although indirect measures, such as strain, are also possible. The most appropriate quantity to measure depends on the amplitude and frequency range of the motion as well as on the application. Sections 12.6.2, 12.6.3, and 12.6.4 cover common methods for measuring acceleration, velocity, and displacement, respectively. The relationship between these three quantities will be presented in this section.

Acceleration, velocity, and displacement are related, so theoretically, if one of these quantities is measured, then the other two can be calculated. For example, a time derivative can be applied to the displacement data to obtain velocity, $v(t) = (dx/dt)t$, and a time derivative can be applied to the velocity data to obtain acceleration, $a(t) = (dv/dt)t$. Conversely, acceleration data can be integrated in time to obtain velocity and similarly, velocity data can be integrated to obtain displacement.

Vibration data are typically analyzed in the frequency domain where the relationships between acceleration, velocity, and displacement reduce to simple multiplications. Specifically, in the frequency domain, a time derivative is equivalent to a multiplication by $2\pi jf$, and a time integral is equivalent to a multiplication by $1/(2\pi jf)$, where f is the frequency in Hz and $j = \sqrt{-1}$.

These relationships can be more easily understood by considering a time harmonic response

$$x(t) = A \cos(2\pi f t + \phi) \quad (12.25)$$

at a single frequency f . In Equation 12.25, the phase, $\phi = \phi(f)$, can be a function of frequency. The velocity

$$v(t) = \frac{dx}{dt}(t) = 2\pi f A \sin(2\pi f t + \phi) \quad (12.26)$$

is obtained by taking the time derivative of the displacement. Similarly, the acceleration

$$a(t) = \frac{dv}{dt}(t) = -(2\pi f)^2 A \cos(2\pi f t + \phi) \quad (12.27)$$

is obtained by taking the time derivative of the velocity.

A comparison of Equations (12.25) and (12.27) shows that the amplitude of the acceleration of a time harmonic motion can be obtained by multiplying the displacement amplitude by $(2\pi f)^2$. Therefore, for the same displacement amplitude, higher frequency oscillations will have higher acceleration amplitudes.

Equations (12.25)–(12.27) can be rewritten in complex notation as

$$x(t) = \text{Re}[A e^{j(2\pi f t + \phi)}] \quad (12.28)$$

$$v(t) = \text{Re}[2\pi j f A e^{j(2\pi f t + \phi)}] \quad (12.29)$$

and

$$a(t) = -(2\pi f)^2 \text{Re}[A e^{j(2\pi f t + \phi)}] \quad (12.30)$$

This notation shows that multiplication by j in the frequency domain corresponds to a 90° phase shift (e.g., from $\cos(2\pi f t)$ to $\sin(2\pi f t)$) in the time domain. Therefore, acceleration and displacement are 180° out of phase.

By superposition, the time domain signal for a linear system can be represented as a summation of time harmonic components, each component represented by one point in the frequency spectrum. Therefore, for example, a plot of the amplitude of the velocity

spectrum can be obtained from a plot of the amplitude of the displacement spectrum simply by multiplying each point of the plot by $2\pi f$.

The relationships among displacement, velocity, and acceleration are useful in determining which sensor to use to measure a specific dynamic response. Equations (12.25)–(12.27) imply that, in general, accelerometers will be more sensitive to high frequency motions, while displacement sensors will be more sensitive to low frequency motions. However, because numerical error is introduced in converting from one measure to another, it is often best to measure the quantity most representative of the phenomena under study. Other factors to consider include the sensitivity, resolution, and linear frequency range of the sensor.

12.4.2.2 Acceleration Measurement An accelerometer is a device used to measure acceleration. Currently, accelerometers are the dominant choice for motion measurement, so they will be described in detail.

Accelerometers are seismic transducers that consist of a spring–mass system attached to an outer housing. The displacement of the internal mass relative to the housing is approximately proportional to the housing acceleration, within the operating frequency range of the accelerometer. The upper limit of the operating frequency range is limited to about one-third of the accelerometer’s natural frequency. Therefore, the spring–mass system within an accelerometer is design to have relatively high stiffness and small mass to achieve a high natural frequency. Accelerometers come in a range of sizes, with the smaller accelerometers typically designed for higher frequency measurement.

The relative displacement between the accelerometer mass and housing causes a compression or expansion of the spring element. Therefore, either a measure of the spring compression or relative mass displacement can be used to produce an output signal proportional to the acceleration of the housing. When the housing is firmly attached to a vibrating structure, the output signal will be proportional to the acceleration of the structure at the attachment point.

The three basic types of accelerometers, piezoelectric, piezoresistive, and capacitive, differ by how the relative displacement of the mass (or spring deformation) is measured. The operating frequency range, sensitivity, and measurement resolution typically differ between the types. However, these differences are becoming less significant as transducer design improves.

Piezoelectric accelerometers are the most common and are available in a range of sizes. A variety of piezoelectric accelerometers are shown in Figure 12.23. In piezoelectric accelerometers, a stiff piezoceramic or a piezoelectric crystal serves as the spring and also as the sensing element. Deformation of the piezoelectric crystal produces a charge or electronic potential proportional to the strain. The main advantages of piezoelectric accelerometers are their high operating frequency ranges, smaller size and weight, and their insensitivity to temperature. However, piezoelectric accelerometers are not suited for quasi-static (near zero frequency) measurements because the charge generated in the piezoelectric material will slowly bleed off through internal resistances. In addition, piezoelectric accelerometers are typically not recommended for high shock levels due to strain limits on the sensing element.

There are two types of piezoelectric accelerometers: charge mode piezoelectric (PE) and voltage mode internal electronic piezoelectric (IEPE). Internal electronic piezoelectric accelerometers have a built-in charge amplifier that eliminates the need for an external constant current source and for special low-noise cables. Therefore,

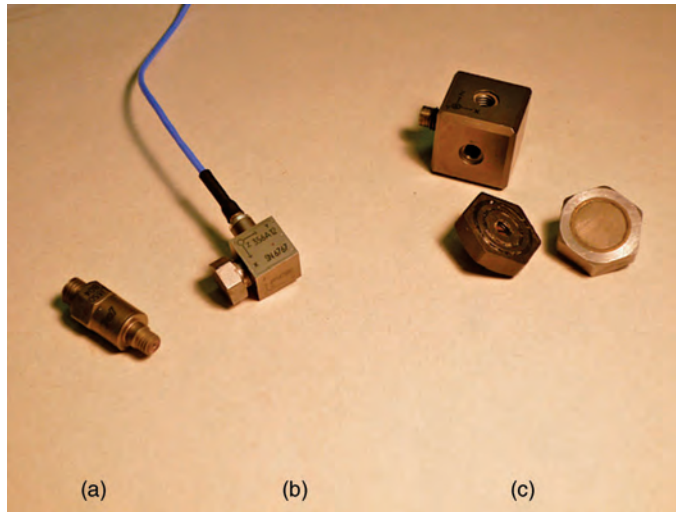


FIGURE 12.23 A variety of piezoelectric accelerometers are shown with different mounting attachments. The uniaxial accelerometer (a) has a threaded stud. The smaller triaxial transducer (b) has an attached magnetic mount. The magnetic and wax mounting surfaces shown can be attached to the larger triaxial accelerometer (c).

internal electronic piezoelectric accelerometers are more popular than charge mode accelerometers.

In a *piezoresistive accelerometer*, the sensor is a strain gage attached to the spring element. The strain gage generates a change in electrical resistance proportional to the compression of the spring element. Historically, piezoresistive accelerometers have had lower operating frequency ranges than piezoelectric accelerometers, but recent design improvements have reduced this distinction. A major advantage of piezoresistive accelerometers is their ability to measure low, near zero, frequency signals. Piezoresistive accelerometers have low sensitivity, so they are better suited to shock applications, such as drop or crash tests, than for vibration testing and monitoring.

Capacitive accelerometers use a capacitance probe to measure the gap between the mass and the housing. Capacitive accelerometers are good for measuring lower frequency vibrations. Microelectrical–mechanical systems (MEMS) type capacitive accelerometers have become popular in video game controllers and cell phones for sensing control inputs and also for sensing tilt.

For vibration measurement, an accelerometer can be mounted on a vibrating structure using a threaded stud, cement, wax, adhesive tape, or magnet. Stud, wax, and magnetic mount attachments are shown in Figure 12.23. The choice of attachment method depends on a number of factors. However, the connection should be as strong as possible, because weak attachments will decrease the operating frequency range of an accelerometer.

For long-term placement, an accelerometer can be mechanically fastened with a threaded stud or bolt or it can be chemically bonded with a cement. These attachment methods are typically the strongest, and therefore they have the least effect on the operating frequency range.

A variety of temporary mounting options exist for accelerometers that need to be frequently repositioned. A magnetic accelerometer base can be used on steel and iron

ferromagnetic surfaces. A stronger magnet will have a smaller effect on the operating frequency range, but the additional mass may affect the motion of the test structure. Wax or double sided tape can be used to temporarily attach an accelerometer to a surface. A thinner bonding layer will have higher stiffness and thus have less effect on the operating frequency range. Flat, smooth, and clean surfaces are needed to ensure good contact between the accelerometer and the test structure for these temporary attachment methods. Holding an accelerometer by hand against the test surface will produce the poorest results, but may be useful for preliminary measurements.

If an accelerometer is not electrically insulated from a test structure, extraneous signals may result. Electrical noise transmission is more likely when a mechanical fastener or magnetic mount is used on a machine with operating motors. A strong 60 Hz signal is a good indication that electrical noise is present and electrical isolation is needed.

Attaching an accelerometer to a structure effectively increases its mass. To avoid significant changes in the dynamic response of the structure, the accelerometer mass should be less one tenth of the dynamic mass of the structure at resonance. Care should also be taken to ensure that the accelerometer cable does not affect the vibration measurement.

As with any linear motion sensor, care should be taken to align the sensor to the desired measurement direction. An angle misalignment θ_{err} from intended direction introduces an error proportional to $(1 - \cos(\theta_{\text{err}}))$. Fortunately, this error is typically negligible for small angles of misalignment.

12.4.2.3 Velocity Measurement Similar to an accelerometer, traditional velocity transducers also consist of a spring-mass system attached to an outer housing. The mass is a permanent magnet that can move along the axis of a cylindrical coil of wire attached to the housing. A voltage signal proportional to housing velocity is produced by the relative motion between the magnet and the coil.

For velocity sensors, the spring is weak and mass is large, so the natural frequency of the spring-mass system is low (a few Hz). The mass is surrounded by a damping fluid, usually oil, to damp out vibration at the natural frequency. The operating frequency range is above the natural frequency, but typically not as high as the operating ranges for typical accelerometers.

These traditional velocity transducers are sturdy and most need no external power supply or charge amplifier, but tend to be larger than comparable accelerometers (see Figure 12.24). Many machinery monitoring standards were written in terms of velocity measurements because these measurements used the technology available at that time. However, due to advances in accelerometer technology, these velocity transducers are not as popular as they once were.

More recently, laser Doppler vibrometer systems have been developed that use the Doppler shift of laser light wavelength to measure the velocity of a vibrating structure (Maurer et al., 2010). Laser Doppler vibrometers and other optical system can be more expensive than other measurement devices, but they offer a number of advantages that have lead to their steady gain in popularity. Because optical systems are noncontact systems, they can be used in harsh environments and on rotating shafts where it would be impossible to mount a motion sensor. In addition, using a noncontact system eliminates mass loading effects. Laser systems usually have higher spatial resolution and the laser beam can be scanned across the test structure so that a large number of data points can be obtained rapidly. Also, laser systems typically have a broader operating frequency range, including near zero Hz.



FIGURE 12.24 A velocity transducer showing its typical large size.

12.4.2.4 Displacement Measurement Displacements may be the most appropriate quantity to measure in applications where clearance between bodies is critical. The most popular displacement transducers are noncontact proximity probes, although contact probes, such as linear variable differential transformers (LVDTs) and string potentiometers, are used in some vibration measurement applications.

Eddy current probes are probably the most common noncontact proximity sensor used in vibration monitoring. Eddy current probes can only measure the vibration displacements of electrically conductive (metallic) surfaces. Eddy current probes are sometimes referred to as inductive proximity sensors, because the electromagnetic field generated at the probe tip induces eddy currents in a nearby metal surface. The interaction between the electromagnetic field and the induced eddy currents depends on the distance between the probe and the metallic surface. The probe senses this interaction and produces an electrical voltage proportional to the gap thickness. The interaction depends on the conductive properties of the metallic surface, so eddy current probes must be calibrated for each material.

Eddy current probes are often used to measure the vibration of rotating shafts. Direct measurement of shaft vibrations is especially useful when the shaft is supported by oil-film bearings. Vibrations transmitted through oil-film bearings are difficult to measure since they are significantly dampened and of much lower amplitude than the vibration of the shaft itself. Because the eddy current probe output is proportional to the gap thickness, probe vibration as well as shaft vibration will affect the output signal. Therefore, proximity probes must be securely mounted to a nonvibrating surface for accurate measurement of shaft displacement. The probe casing is often threaded, as shown in Figure 12.25, to allow for a secure attachment. In addition, shaft run out, imperfections in the shaft material, and shaft surface quality will affect the eddy current probe signal. These latter affects can be measured while the shaft is rotating slowly (so no vibrations are excited) and subtracted from the measurements obtained during vibration testing. A rotary encoder or other angular position measurement is needed to synchronize the angular positions of the two measurements.

Capacitive or optical proximity sensors can be used if the material of the vibrating structure is nonconductive. Capacitive sensors generate an electric field that induces a



FIGURE 12.25 A proximity probe (b) and mounting fixture (a). The proximity probe has a threaded casing used to securely mount the probe into a fixture.

charge in the test surface. The capacitive sensor measures the capacitance of the air in the gap between the probe and test surface to generate a signal that is proportional to the gap thickness. A variety of optical sensors are available, most based on fiber optic technology.

Proximity probes can be used to measure static as well as dynamically changing gaps. Therefore, there is no lower limit to the operating frequency range. However, the upper frequency limit is typically not as high as for accelerometers or velocity sensors.

12.4.2.5 Force Measurement The excitation force imparted by a shaker, hammer, or other source must be measured for many vibration testing applications. Instrumented impact hammers typically have a load cell built into the tip, while external load cells, such as the one shown in Figure 12.26, are needed for other excitation sources. A load cell is a transducer that generates an electrical signal proportional to an applied normal or shear force. The force amplitude is obtained indirectly from a measurement of the strain or deformation of an internal component of the load cell. This is analogous to determining the amplitude of a force, F_s , applied to a spring (with spring constant k) by measuring the extension, x_s , of the spring, using the equation, $F_s = kx_s$. For dynamic measurements, the mass of the load cell may also affect the measurement as well as the response of the structure (McConnell and Varoto, 2008).

Piezoelectric force transducers are the most common load cells used for vibration measurements. Similar to piezoelectric accelerometers, a piezoceramic or piezoelectric crystal acts as the “spring” element and produces an electronic potential proportional to the applied load. Piezoelectric force transducers have advantages and disadvantages similar to piezoelectric accelerometers because the sensing element is the same.

12.4.3 Vibration Analyzers

A signal analyzer is the core element of a vibration measurement system. Digital signal analyzers have been developed to handle both data collection and analysis functions for



FIGURE 12.26 A normal force load cell with mounting stud (foreground) used to measure the force applied to a test structure by a shaker. The stinger with collets attached at both ends (background) is used to connect the shaker to the load cell.

many vibration applications. These functions include signal conditioning, filtering, analog-to-digital (AD) conversion, analysis, storage, and presentation of the vibration data. In addition, many analyzers can output a control signal for shaker or other source excitation. Analyzers are comprised of a hardware unit and software tools. The software runs either on the analyzer itself or on a linked computer. Discrete frequency domain analysis is the main focus of these tools, although capabilities for time domain and other analysis types may also be available. This section covers the basic functions of a digital signal analyzer and the general differences between current analyzers types.

One major function of an analyzer is to acquire measurement data. Most vibration measurement applications require at least two input channels. For example, simultaneous measurements of both excitation and response signals are required to generate frequency response functions (defined in Section 12.3.3) needed for modal analysis (discussed in Section 12.5). Often analyzers have more than two input channels so that multiple responses can be measured simultaneously. Simultaneous measurement not only can save time but also can improve the quality of the analysis results.

Typical force or response measurement devices require a conditioner to convert or amplify the weak analog signal generated by the sensing element of the device to an analog voltage signal in a range that can be used by the analyzer. Modern analyzers may have built-in signal amplification and conditioning functions so that separate conditioning units are unnecessary.

Because vibration analysis is typically done in the frequency domain, digital signal analyzers ultimately generate a frequency spectrum of a time varying voltage. Before the data are acquired, the user must specify the frequency range and number of points for the frequency spectrum. These two quantities determine not only the spectrum's frequency resolution but also the time interval for data acquisition. Higher frequencies require faster

sampling rates and therefore, shorter data acquisition times. Conversely, slower sampling rates and longer data acquisition times are used for lower frequencies.

The frequency spectrum is obtained by applying a discrete Fourier transform to the time domain data. Refer to Section 12.3.3 for a discussion of discrete Fourier transforms. Before the discrete Fourier transform can be applied, the analog voltage input signal must first be low-pass filtered and converted to a digital signal using an analog-to-digital (AD) converter. To prevent aliasing (defined in Section 12.3.3.1), most signal analyzers automatically filter out contributions from frequencies above the user-specified frequency range. Digital signal analyzers use a combination of analog and digital filtering to efficiently reduce contributions from these frequencies to the noise level.

Digital signal analyzers allow the user to select parameters that can affect the accuracy of the resulting spectrum, such as those discussed in Section 12.3.3. Values for these parameters must be properly chosen so that the most accurate time and frequency domain representations of the data are obtained. The input voltage range, coupling type, and triggering parameters must be specified to obtain an accurate discrete representation of the time domain signal needed for analysis and storage. The input voltage range specifies the amplitude range for analog-to-digital conversion. Input signals with voltages outside this range will be cut off and signals that only use a small portion of this range will have poor resolution. Ideally, the range for each input channel should be chosen so that the input signal voltages use most of the input range without exceeding the range limit. The input signal can be either DC or AC coupled to the input channel. AC coupling is essentially a high pass filter that can be used to eliminate a large DC bias to achieve better resolution for small amplitude higher frequency oscillations. Triggering can be used to capture transient signals or to synchronize signals and preserve phase information.

Analyzers typically have a variety of windowing functions that can be selected to prevent leakage (defined in Section 12.3.3.2) or to minimize the effect of noise on the frequency spectrum. Some windowing functions that are useful for vibration measurements are described in Section 12.3.3.2. The effect of windowing on the frequency spectrum can be determined experimentally by comparing the frequency spectrums obtained for different windowing functions.

A frequency spectrum of the response signal is the most basic output of an analyzer and can be useful in a variety of applications. However, frequency response functions are used for most vibration analysis applications and therefore can be calculated by most digital signal analyzers. The frequency response function is defined in Section 12.3.3 as the response spectrum divided by the excitation frequency spectrum. The frequency response function can typically be displayed in a number of formats to allow easier analysis of the results.

Analyzers can also have a variety of software options for further analysis of the vibration data. These analysis options can more rapidly produce meaningful test information for special purpose applications. For example, modal analysis software can be used to determine the natural frequencies, modal damping coefficients, and mode shapes of a structure (see Section 12.5). Software options for other applications, such as machine fault monitoring or fatigue testing, may also be available.

Data are typically stored on digital media. With a link to a computer, analysis results can be transferred and output in various formats including simplified screen shots, spreadsheet data, and video animations files. Modal analysis software often allows for animations of mode shapes or operational deflection shapes so the problem areas of a structure can be quickly identified. Standardized file formats have enabled the transfer of data

across hardware and software platforms, for example, for comparing measured frequency response functions and mode shapes with finite element predictions.

Analyzers come in various sizes and with a variety of capabilities. Factors to consider when selecting an analyzer include the number of channels, the available frequency range, the quality of the signal processing, the analysis software and display capabilities, the analyzer size and portability, and the cost.

Vibration analyzers have experienced a significant reduction in size along with improvements in capability and signal processing quality over the last few decades. Hand-held and very small units with high frequency bandwidths and multiple channels now exist. Medium size, but still portable, multiple channel systems often offer more flexibility and software analysis options for customizable multipurpose applications. As a result, single-chassis portable multichannel analyzers have become commonplace in a variety of industries. Systems with high numbers of channels can be realized by linking individual analyzers with synchronizing connections. Racks of multiple modules can simultaneously record dozens, hundreds, or even more channels. Multiple racks can be synchronized over significant distances between racks, allowing for short rack to sensor connections even when acquiring measurements on very large structures.

12.5 EXPERIMENTAL MODAL ANALYSIS

Experimental modal analysis is the determination of the dynamic characteristics of a structure from vibration measurement data. The basic parameters that define the dynamic response of a structure are the natural frequencies, damping coefficients, and mode shapes. These parameters were introduced in Section 12.3 for multi-degree-of-freedom systems and will be reviewed using a plate model in this section, before experimental modal analysis techniques are presented. Two basic assumptions are typically made: (1) the structure's response is linear and (2) the structure's dynamic characteristics do not change with time.

The natural frequencies, ω_i , are the easiest to measure and often the most important. For example, if a structure is excited at one of its natural frequency, then resonance can occur that might result in large amplitude vibrations. Corresponding to each natural frequency, ω_i , there exist a mode shape, φ_i , that specifies the shape of the vibration at that frequency. Mode shape information is needed to fully characterize the structure's resonance response and to determine the locations of the largest vibration amplitudes. The damping at each natural frequency can be characterized by a modal damping coefficient, ζ_i . Damping will limit the amplitude of vibration when a structure is excited at one of its natural frequencies and will cause the vibration amplitude to decay when the excitation is stopped.

These quantities, natural frequencies, mode shapes, and modal damping, are not only useful in specifying a structure's response to harmonic excitation at one of its natural frequencies, but they also can be used to predict the structure's response to any excitation. These predictions can be made using either a model generated directly from the modal analysis data or a finite element model of the structure. In the latter case, the experimental modal analysis data can be used to verify or update the finite element model. Either model can then be used to investigate design modifications that can minimize the effects of the structure's vibration during operation.

Modal parameters cannot be directly measured, but instead must be inferred from measured vibration responses. There are numerous methods for obtaining the

measurement data and also for extracting modal information from the data. The experimental measurement data can either be analyzed in the frequency domain or the time domain. Frequency domain analysis will be used in this section because it is used in the most common methods. Typically, FRFs (defined in Section 12.3.3) are the basis of the analysis. To obtain frequency response functions, both the excitation and the response must be measured.

Techniques for obtaining modal information when the excitation forces are unknown, for example, from operational excitation, will be briefly discussed in Section 12.5.2. These output-only techniques are useful when vibration measurements can only be obtained on a machine during operation or when a structure is too large to excite with available sources. Also, it may be necessary to measure the natural frequencies of a shaft while at operating speed, because the natural frequencies can change with rotational speed, especially at high rotation rates. See Avitabile (2000) and Batel (2002) for practical discussions of modal analysis and operation modal analysis, respectively, and Bucher and Ewins (2001) for a review of modal analysis for rotating structures.

12.5.1 Modal Analysis Using Frequency Response Functions (FRFs)

In this section, it is assumed that both the excitation and the response are measured when obtaining the experimental data for modal analysis. First, a brief description of modal analysis concepts will be given in Section 12.5.1.1 in the context of a simple thin rectangular plate. These concepts are useful when choosing the excitation and response locations and directions for vibration measurements, and will also add insight when interpreting modal analysis data. In Section 12.5.1.2, experimental measurement data typically used in modal analysis will be discussed. Finally, in Section 12.5.1.3, an overview of methods for estimating the modal parameters from the measurement data will be given.

12.5.1.1 Natural Frequencies, Mode Shapes, and Modal Damping The concepts of natural frequencies, mode shapes, and modal damping discussed for multi-degree-of-freedom systems in Section 12.3.1 can be generalized to structures that have a continuous distribution of mass and stiffness. The rectangular plate, shown schematically in Figure 12.27, will be used for illustration in this section as a simple example of a continuous structure. Result obtained from both experiment measurements and a finite element model of the plate will be used to illustrate basic vibration phenomena. For this illustration, free

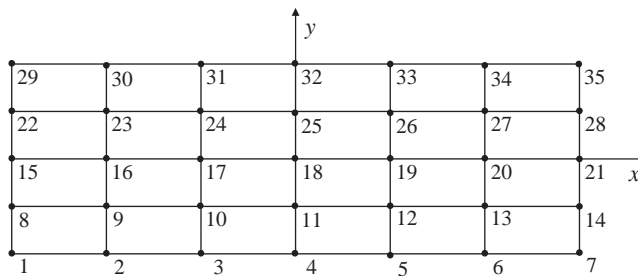


FIGURE 12.27 Schematic view of the top surface of a rectangular plate (12 in. by 4 in. by 0.088 in.) with an overlaid rectangular grid of 35 points. For the analysis, the plate is unconstrained. During experimental measurements, the plate was placed on soft foam to approximate the unconstrained boundary conditions.

boundary conditions are used in the finite element model. In other words, no constraints are applied to the boundaries in the finite element model, so the finite element results predict the behavior of a plate floating in space. Experimentally, free boundary conditions can be approximated by supporting the plate on a soft foam pad or suspending the plate with elastic cords.

Theoretically, a continuous structure has an infinite number of natural frequencies. However, the structure's response is dominated by contributions from a finite number of modes and a finite range of frequencies. The relevant range of frequencies is determined by the frequency content of the excitation. Therefore, frequencies are numbered in ascending order, often with the lowest frequencies dominating the structure's response.

Corresponding to each natural frequency, ω_i , is a mode shape, $\boldsymbol{\varphi}^{(i)}(x, y, z)$, that specifies the shape of the structure when vibrating at that frequency. Specifically, $\boldsymbol{\varphi}^{(i)}(x, y, z)$ is a vector that specifies the normalized x , y , and z displacements as a function of location (x, y, z) in the structure.

For illustration, the first five mode shapes of the rectangular plate with free boundary conditions are shown in Figure 12.28. The modes shapes are in order of increasing frequency. Modes 1, 3, and 5 correspond to bending modes of the plate, which are analogous to the bending modes of an unsupported beam. There are similar bending modes across the width of the plate, but these occur at higher frequencies because the width is smaller than the length of the plate, causing a higher stiffness in this direction. Modes 2 and 4 correspond to twisting modes for this plate.

For the modes shapes of the thin plate, the transverse displacements (in the z direction, normal to the plate) are significantly larger than the in-plane displacements (in the x and y directions). Also, the displacements do not vary significantly across the thickness of the plate (in the z direction). Therefore, the i th mode shape for the rectangular plate can be approximated by a scalar, $\varphi^{(i)}(x, y)$, that specifies only the transverse displacement as a function of the (x, y) location on the plate surface. This implies that a good approximation for the plate mode shapes can be obtained experimentally using only normal excitation forces and measuring responses normal to the plate. This demonstrates how an understanding of the expected dynamic behavior of a structure can be used to dramatically reduce the quantity of data required for experimental modal analysis.

Damping is often neglected when determining the frequencies and mode shapes of a structure from analytical or numerical (e.g., finite element) models. In the absence of damping, the predicted phase between the excitation and response at any point is either 0° (in phase) or 180° (out of phase). In a mode shape animation for an undamped structure, all points across the structure will reach their maximum deflection at the same time and will simultaneously pass through zero deflection. Therefore, an animation of an undamped mode shape will appear to be a standing wave.

Any real structure has some damping, although in most structures that damping is small. Therefore, measured mode shapes will be approximately the same as the mode shapes predicted using models without damping. However, the phases from measured data will not be exactly either 0° or 180° . Measured responses of lightly damped structures often will be nearly in phase or out of phase. When the modal responses at different points are nearly in phase or out of phase, the vibration mode is described as having a *low modal complexity*.

Theoretically, if the damping is “proportional damping,” that is, distributed throughout the structure in the same way as the stiffness or mass, then damping will not affect the modal complexity. However, typically joint friction contributes significantly to the

damping of a structure, resulting in “nonproportional damping.” Nonproportional damping can increase modal complexity, especially when successive frequencies are close and their response peaks significantly overlap. A mode is said to have *high modal complexity* if the phase difference between structural points within the mode shape vary significantly from 0° or 180° . In this case, a stationary plot of a complex mode shape would not properly represent the vibration at that frequency. An animation of a mode with high modal complexity will appear more like a traveling wave than a standing wave.

The modal response for repeated roots of symmetric structures can also exhibit modal complexity, even in the absence of nonproportional damping. For example, vibration of a bar with a circular cross-section at its fundamental frequency can be in either of the two

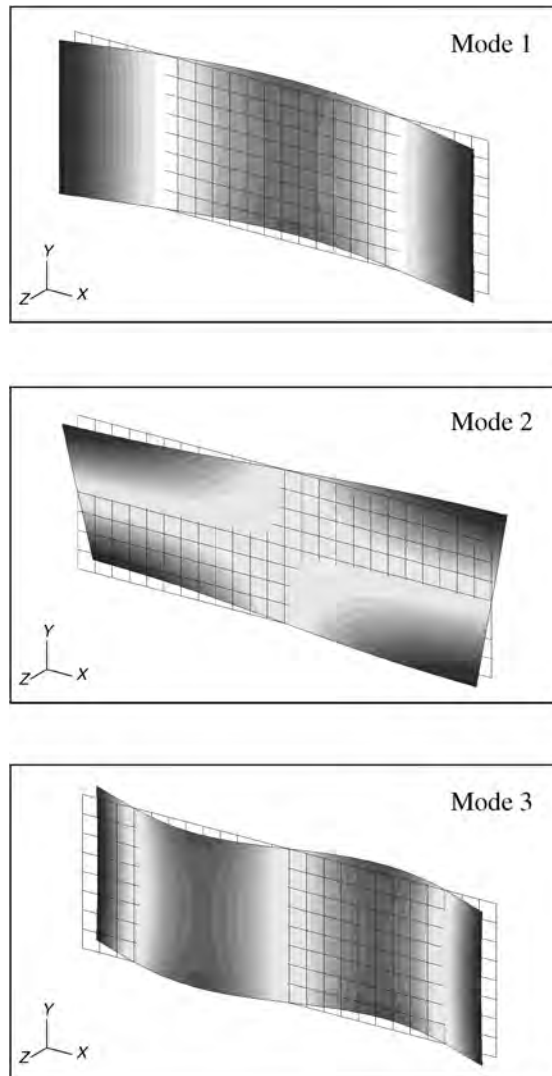


FIGURE 12.28 First five mode shapes of the unconstrained rectangular plate shown in Figure 12.27, predicted by finite element analysis.

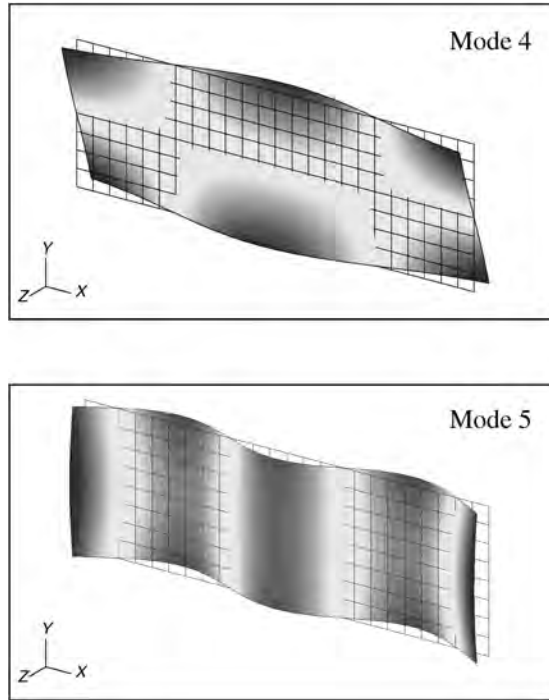


FIGURE 12.28 (Continued)

transverse directions. If these two responses are combined with a 90° phase between them, then the end of the bar will follow a circular or elliptical path in space. Care must be taken in experimental modal analysis as an oversimplified study would not reveal this behavior or enable differentiation between a complex mode, repeated roots, or closely spaced modes.

As described for multi-degree-of-freedom systems in Section 12.3.1, mode shapes are not only useful for understanding the response of a structure when excited at resonance but are also useful for predicting the response of the structure to a more general excitation. For example, when damping is negligible, the transverse displacement response of the plate, $w(x, y)$, can be written in terms of the modes shapes as

$$w(x, y, t) = \sum_{i=1}^{\infty} \varphi^{(i)}(x, y) T_i(t) \quad (12.31)$$

where the time dependence $T_i(t)$ is determined by the excitation.

Damping will limit the amplitude of the response to a continuous excitation and cause the vibration amplitude to decay with time after an excitation has ceased. Although sources of damping are difficult to identify and quantify, an estimate of the magnitude of the damping, ζ_i , can be obtained for each mode from experimental modal analysis data. Analogous to Equation (12.23), the free vibration response of a damped system, with modal damping coefficients ζ_i , can be approximated by

$$w(x, y, t) = \sum_{i=1}^{\infty} A_i \varphi^{(i)}(x, y) e^{-\zeta_i \omega_i t} \sin(\omega_{di} t + \theta_i) \quad (12.32)$$

The coefficients A_i and phase shifts θ_i are determined by the initial displacements and velocities. For the free vibration response, the time dependence for each mode is a decaying oscillation at its damped natural frequency, $\omega_{di} = \omega_i \sqrt{1 - \zeta_i^2}$.

The dynamic response of a structure (given by the infinite sum in Equation 12.32) can be accurately approximated by a sum of contributions from a finite number of modes. Typically only modes with the lowest frequencies or with frequencies in the frequency range of the excitation contribute significantly to the response. This reduced expression for the system response is what makes modal analysis so powerful for interpreting vibration behaviors.

12.5.1.2 Modal Analysis Measurement Data Estimates of a structure's important natural frequencies, damping parameters and modes shapes can be obtained by analyzing a set of vibration response measurements. While some of the resonance frequencies can be estimated from a single measurement, multiple measurements with different excitation and/or response locations are needed ensure that all important resonance frequencies are identified and to obtain mode shape information. For example, to obtain mode shapes for the plate of Figure 12.27, measurements can be taken over a regular grid, such as locations 1 through 35. The responses at these locations can then be analyzed to estimate the modal parameters, including the normalized displacement at each location for the mode(s) of interest. A model using this coarse grid can then be animated to visualize the mode shape(s).

The most basic measurement consists of a single excitation and a single response, often denoted as *single-input/single-output (SISO)*. Both the location and the direction of the excitation and response need to be specified. As stated above, for the plate example, excitation and measurements normal to the plate will provide the best data. Therefore, in this discussion, it is assumed that both the excitation and the response are in the z direction.

The *FRF*, $H_{ij}(f)$, for a single measurement is obtained by dividing the discrete Fourier transform of the response, $X_i(f)$, at location i by the discrete Fourier transform of the excitation, $F_j(f)$, at location j . Because each discrete Fourier transform is a frequency spectrum consisting of a list of complex values (one complex number corresponding to each frequency in a given frequency range), the frequency response function is also a complex frequency spectrum. Therefore, frequency response functions can be represented in terms of either their amplitude and phase or real and imaginary parts, as illustrated in Section 12.3.3.

As an example, the amplitude and phase of a measured frequency response function are shown in Figures 12.29 and 12.30, respectively. The measurements were obtained from the plate of Figure 12.27 for an impact excitation at location 19 and an acceleration measurement at location 18. Note the similarities between the measured frequency response function plot (Figures 12.29 and 12.30) and the plot in Figure 12.6 for the response of a two-degree-of-freedom system. There are two major peaks in the amplitude plots of Figure 12.29 at two of the plate's natural frequencies. The phase plots in Figure 12.30 show a phase change of approximately 180° as the frequency increases through these two natural frequencies. Much of the discussion in Section 12.3 can be extended to the analysis of the dynamic response of continuous systems. For example, as shown in Figure 12.31, the real and imaginary parts of the frequency response functions can be plotted, similar to Figure 12.7a and b or a Nyquist plot similar to Figure 12.8 can be obtained. Analyzer postprocessing software can create these standard plots.

Because the plots shown in Figure 12.29–12.31 are obtained from experimental data, they slightly deviate from predictions of an ideal model. For example, the plate was supported on a soft form pad to approximate a plate with no constraints, that is, with free

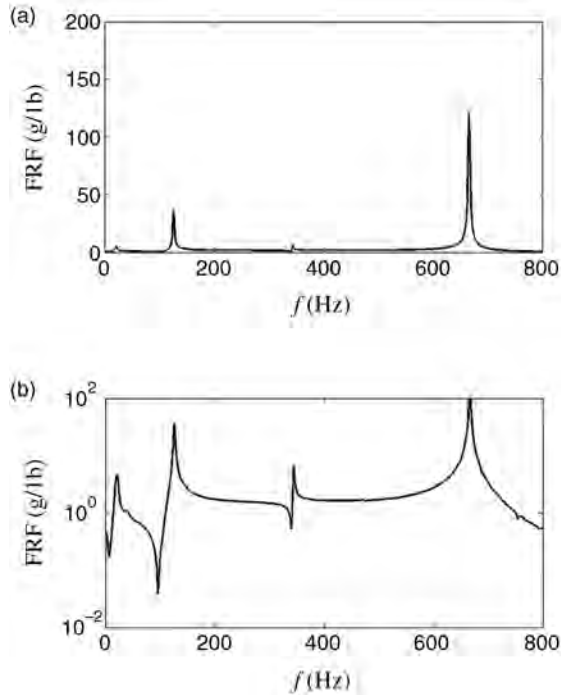


FIGURE 12.29 Amplitude of the frequency response function (FRF) for the plate shown schematically in Figure 12.27, measured with an accelerometer at location 18 for a hammer excitation at location 19. The same data are plotted on (a) a linear scale and (b) a logarithmic scale. The low amplitude resonance peaks and anti-resonances (frequencies at which the response amplitude is very low) are more easily identified on the logarithmic scale.

boundary conditions. The resonance frequency of a rigid body plate vibration on this soft support adds a low amplitude peak at 22 Hz in the frequency response function of Figure 12.29. Because this frequency is less than 1/5th of the fundamental (lowest) plate frequency, the support does not significantly affect the response of the plate at its natural frequencies. Effectively, no mass, stiffness, or damping is added to the plate by the foam.

The two major peaks in Figure 12.29 correspond to the 1st and 5th modes shapes of Figure 12.28. Because location 18 is at a nodal point for the 2nd, 3rd, and 4th mode shapes, an ideal model would predict that there would be no response at location 18 at these natural frequencies. In the experimental results of Figure 12.28, there are no discernable peaks in the frequency response function at the 2nd and 4th plate natural frequencies, but there is a low amplitude peak at the 3rd plate natural frequency, likely due to the finite size of the accelerometer. If either the response *or* excitation locations are at a nodal point for a mode shape, then this mode will not have a significant contribution to the response. Therefore, excitation and response locations must be carefully chosen to obtain information for all the frequencies and mode shapes of interest.

Theoretically, the frequency response function is a characteristic of the structure under test and does not depend on the type of excitation (impact, swept sine, or random). Therefore, a general discussion of modal analysis techniques will be given, without reference to the specific excitation source or response measurement. However, the quality of measured

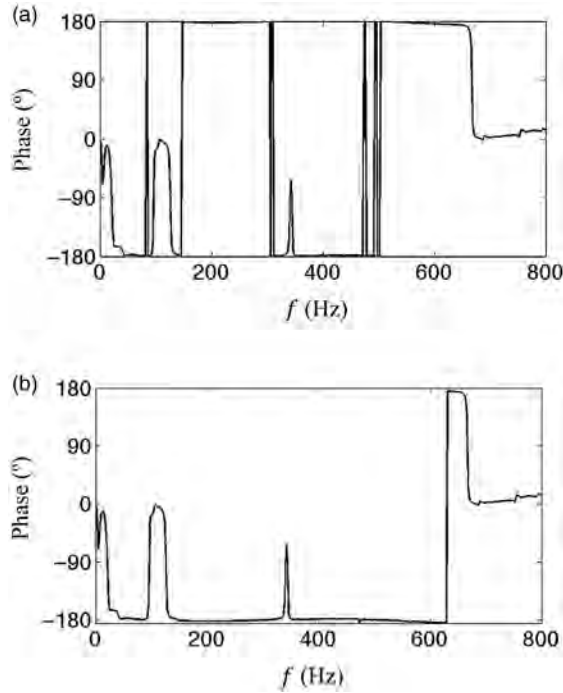


FIGURE 12.30 Phase of the frequency response function (FRF) for the plate shown schematically in Figure 12.27, measured with an accelerometer at location 18 for a hammer excitation at location 19. In (a), the nearly vertical lines going from -180° to 180° and back again are extraneous, but often occur in vibration measurements. These extraneous lines are a result of “wrapping” the phase into the interval from -180° to 180° . The phase is replotted in (b) over a range from -185° to 175° to eliminate these extraneous lines, so that the meaningful characteristics of the plot can be more easily identified.

data will depend on the source (instrumented hammer or a shaker) and response measurement (displacement, velocity, or acceleration) device used, so these should be carefully chosen. The quality of the frequency response functions also will depend on the digital analyzer settings (input range, windowing, averaging, etc.) as discussed in Section 12.3.3. Misleading or meaningless data may be obtained if improper settings are used.

The frequency response functions

$$H_{ij}(f) \stackrel{\text{def}}{=} \frac{X_i(f)}{F_j(f)} \quad (12.33)$$

for a set of measurements at N locations can be assembled into a matrix form as

$$\begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_N \end{bmatrix} = \begin{bmatrix} H_{11} & H_{12} & \cdots & H_{1N} \\ H_{21} & H_{22} & \cdots & H_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ H_{N1} & H_{N2} & \cdots & H_{NN} \end{bmatrix} \begin{bmatrix} F_1 \\ F_2 \\ \vdots \\ F_N \end{bmatrix} \quad (12.34)$$

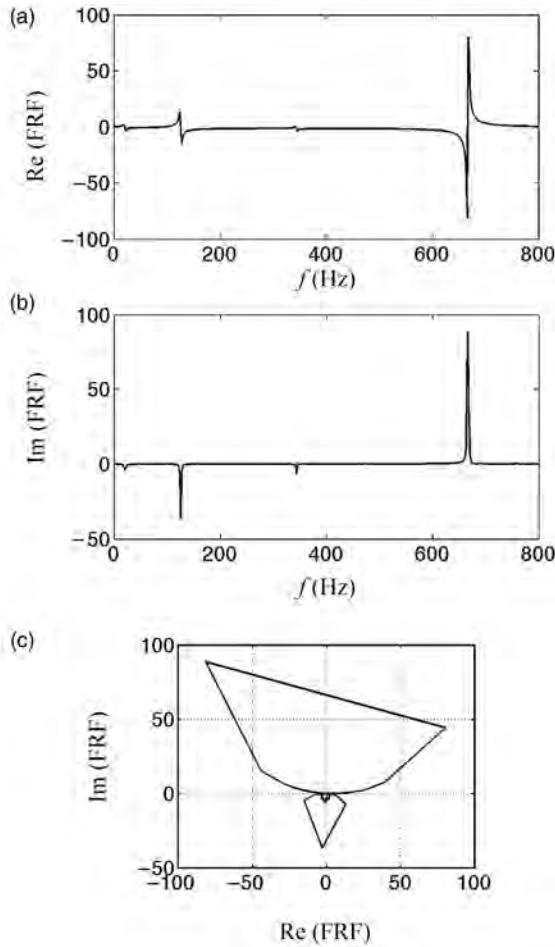


FIGURE 12.31 Plots of the (a) real part and (b) imaginary part of the frequency response function (FRF) of Figures 12.29 and 12.30 versus frequency. The Nyquist plot in (c) is a plot of the imaginary part versus the real part of the frequency response function. The curve in plot (b) is repeated as the third curve (at $x = 2$ in) in the waterfall plot of Figure 12.34.

Any one row or column of the matrix contains sufficient information to estimate the natural frequencies, mode shapes, and modal damping coefficients of the structure in the measured frequency range. Additional measurements can be made to increase the confidence level of the data.

Single-input, single-output techniques can be used to obtain the j th row of the matrix by measuring the response at location j (using an accelerometer or other measurement device) for successive hammer impacts at each grid location i , for $i = 1, 2, \dots, N$. If a shaker is used, it is often easier to move the response measurement location than it is to move the shaker excitation location, so SISO measurements can be used to obtain a column of the matrix, that is, measurements for different response locations for one shaker excitation location.

If a number of accelerometers or other measurement devices are available, then multiple measurements can be taken simultaneously for a single excitation. These measurements are referred to as *single-input/multiple-output (SIMO)*. SIMO techniques not only shorten the data acquisition time but also may reduce measurement inconsistencies caused by variations in the structure or environment over time. *Multiple-input/multiple-output (MIMO)* data may result in additional improvements.

For linear systems, the frequency response function matrix is theoretically symmetric (i.e., $H_{ji} = H_{ij}$) due to Maxwell's reciprocity theorem. The theorem states that the response at location i due to an excitation at location j is equal to the response at location j if the same excitation is instead applied at location i . In other words, the frequency response function obtained from a measurement at location A for an excitation at location B would theoretically be identical to the frequency response function obtained from a measurement at location B for an excitation at location A. For example, the frequency response function obtained for an accelerometer measurement at location 19 for a hammer excitation at location 18 would ideally be identical to the frequency response function plotted in Figures 12.29–12.31. Actual experimental measurements may not show exact symmetry if the mass of the accelerometer or stiffness added by a shaker significantly affects the response of the structure. Verification of the symmetry stated by Maxwell's reciprocity theorem can be used to identify measurement errors and nonlinearities in the response.

Some structures also have geometric symmetries that can be exploited either to reduce the measurement data required or to add redundancy to increase the confidence level of the data. However, care must be taken when exploiting structural symmetries, because vibrational mode shapes can be either symmetric or anti-symmetric about lines or planes of geometric symmetry. For example, the rectangular plate geometry in Figure 12.27 is symmetric about the line $x = 0$. As shown in Figure 12.28, the first mode, is symmetric about this line, but the third mode is anti-symmetric about this line.

12.5.1.3 Modal Parameter Estimation Modal parameter estimation is the process of obtaining estimates for natural frequencies, modal damping coefficients, and mode shapes from a set of measured input/output data. Modal parameter extraction is the key to experimental modal analysis. Many modal parameter estimation algorithms have been developed based on different mathematical models for the response of a test structure. Ideally, all methods should produce similar estimates. However, in practice, the estimates may vary due to limitations in the measurement data and the presence of environmental noise. An overview of all methods is beyond the scope of this chapter, so the reader is referred to Piersol and Paez (2010), Ewins (2000), and other texts on modal analysis and vibration testing for additional information.

In this section, basic concepts that can be used for simple modal parameter estimation will first be summarized. These concepts can also be used as a sanity check on the experimental data and on estimates produced by modal analysis software. Most techniques use FRF data for modal parameter estimation, so only such methods will be discussed in this section. Methods also exist that utilize other forms of the experimental data, such as impulse response functions data, obtained by applying a discrete inverse Fourier transform to the frequency response function data. Time domain methods, such as this, are beyond the scope of this chapter.

Single-Mode Analysis Methods Single-degree-of-freedom modal analysis methods estimate the modal parameters of each mode independently. The basic assumption for

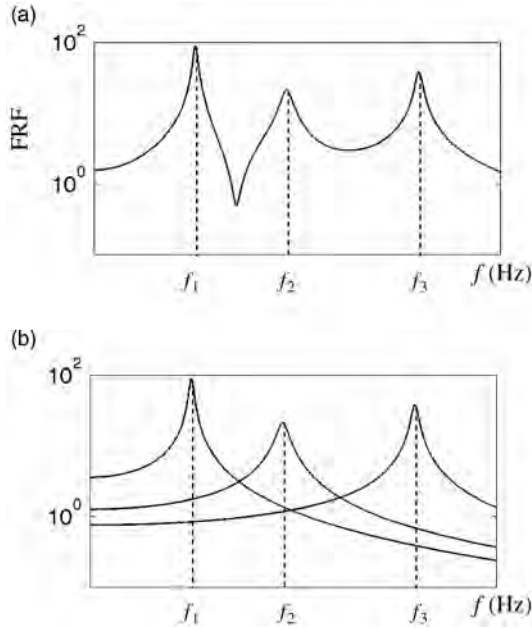


FIGURE 12.32 The frequency response function with three peaks in (a) can be represented in terms of the sum of three single-degree-of-freedom responses, each with a single peak as illustrated in (b). There is an antinode between the first two peaks because the responses corresponding to these peaks are nearly out of phase.

single-degree-of-freedom modal analysis approximations is that in the vicinity of a resonance, the response is due primarily to that single mode. A frequency response function with N peaks can be represented as the summation of N single-degree-of-freedom system peaks as illustrated in Figure 12.32. This also follows from Equation (12.31) or (12.32) which express the structure's response as a sum of contributions from each mode. Therefore, if the peaks of a frequency response function are sufficiently separated, simple single-degree-of-freedom methods described in Section 12.2 can be used to estimate the modal parameters for each mode in the measurement frequency range.

Specifically, if damping is small, then the i th natural frequency, f_i , and nondimensional damping coefficient, ζ_i , can be approximated from the i th peak of the frequency response function by extensions of Equations (12.14) and (12.15). The generalized equations are

$$f_i = \frac{f_{\text{peak}i}}{\sqrt{1 - 2\zeta_i^2}} \sim f_{\text{peak}i} \quad (12.35)$$

and

$$\zeta_i \sim \frac{f_R - f_L}{2f_{\text{peak}i}} \quad (12.36)$$

respectively. In Equations (12.35) and (12.36), $f_{\text{peak}i}$ is the frequency at the i th peak. If damping is small, then the resonance peak ($f_{\text{peak}i}$) occurs at the undamped natural

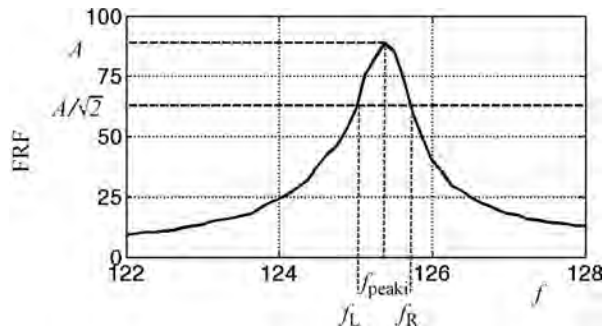


FIGURE 12.33 Amplitude of the frequency response function (FRF) for the plate shown schematically in Figure 12.27, measured with an accelerometer at location 18 for a hammer excitation at location 19. The measurement was taken using a frequency range from 100 Hz to 200 Hz with 800 lines of resolution.

frequency (f_i). If the amplitude of the peak is denoted by A , then f_L and f_R are the $1/2$ power point frequencies, where the amplitude of the frequency response function is $A/\sqrt{2}$. The i th modal damping coefficient, ζ_i , can be determined from the $1/2$ power point frequencies for the i th peak using Equation (12.36).

The variables used in Equations (12.35) and (12.36) are illustrated on the frequency response function plot shown in Figure 12.33. This frequency response function was obtained for the same excitation and response measurement locations as were used to obtain the plots for Figure 12.29. To increase the resolution of the frequency response function, a reduced measurement frequency range (from 100 Hz to 200 Hz) was used. The amplitude of the frequency response function was plotted for a frequency range 122–128 Hz in Figure 12.33 to show details of the first peak and the definitions of $f_{\text{peak}i}$, f_L , and f_R .

Alternate single-degree-of-freedom techniques exist for approximating the i th natural frequency, f_i , and nondimensional damping coefficient, ζ_i , from plots using the real and/or imaginary parts of the frequency response function. If damping is small, then the natural frequencies and modal damping coefficients predicted by the different techniques will be approximately the same. For example, the natural frequencies, f_i , can be identified by determining the frequencies at which the real part of the frequency response function (Figure 12.31a) is equal to zero. Also for small damping, a peak will occur in the imaginary part of the frequency response function (Figure 12.31b) approximately at the natural frequency, f_i , with an amplitude approximately equal to the magnitude of the i th peak of the frequency response function. If damping is small; however, there may be too few data points to determine the peak amplitudes accurately. For these cases, methods of fitting semicircles to the Nyquist plot (Figure 12.31c) have been developed so that more points around the peak are used in the estimation of f_i and ζ_i . Modal analysis software typically will offer a variety of automatic single- (and multiple-) degree-of-freedom curve fitting methods to estimate the modal analysis parameters. See Piersol and Paez (2010), Ewins (2000), and other texts on modal testing for additional details.

Mode shapes of the structure can be quickly approximated from the imaginary part of frequency response functions measured across the structure. The approximations are most accurate if damping is small and the peaks are well separated. As an illustration, seven frequency response functions were measured along the center ($y = 0$) line of the example

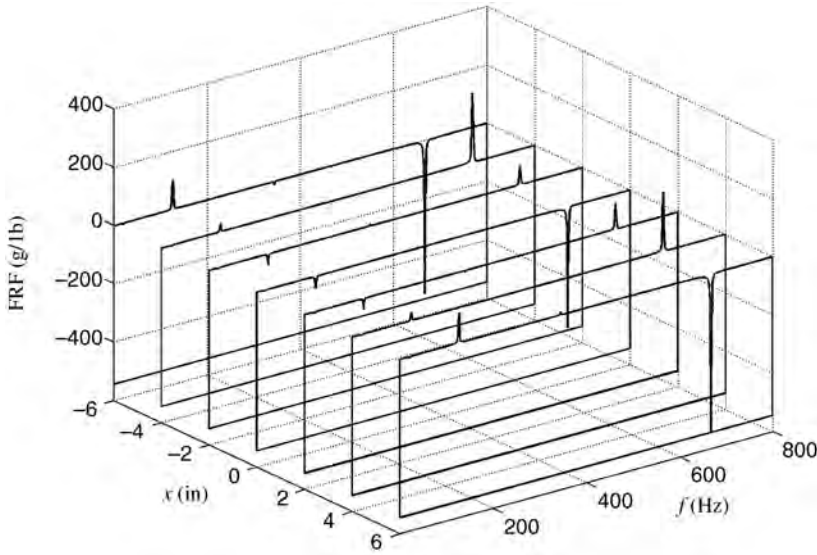


FIGURE 12.34 Waterfall plot of the imaginary part of the frequency response functions (FRFs) for the plate shown schematically in Figure 12.27, measured with an accelerometer at location 18 for seven hammer excitation locations along the $y = 0$ axis. A curve connecting the first major peak (at ~ 125 Hz) of each plot approximates the 1st mode shape, as demonstrated in Figure 12.35a. A curve connecting the second major peak (at ~ 675 Hz) of each plot approximates the 5th mode shape, as demonstrated in Figure 12.35c. Modes 2, 3, and 4 do not significantly contribute to these responses since the accelerometer is at a nodal point for these mode shapes.

plate at locations 15 through 21, with the accelerometer at the center ($x = 0$ and $y = 0$), location 18. These measurements are compared to finite element predictions of the mode shape deflections (given in Figure 12.28), plotted for the line $y = 0$.

The imaginary parts of the seven frequency response functions are displayed in the waterfall plot of Figure 12.34. The imaginary part of the frequency response function at each x location has two major peaks, at the 1st and 5th natural frequencies of the plate. By connecting the 1st major peak in each plot, an approximation to the 1st mode shape can be visualized. This is demonstrated in Figure 12.35a, where the normalized measured peak values are plotted and compared to the centerline deflections of the first mode shape predicted by finite element analysis. The measured mode shape matches the predicted mode shape well except at the center point, where the mass of the accelerometer decreases the response amplitude of the plate.

Similarly, by connecting the 2nd major peak in each plot, the 5th mode shape can be approximated. This is demonstrated in Figure 12.35c, where these measured peak values are compared to the fifth mode shape predicted by finite element analysis. Other mode shapes in the frequency range, for example the 3rd mode shape (Figure 12.35b), cannot be adequately approximated from the measurement data since the accelerometer is at a nodal point for these modes.

Note that, if the response or *excitation* location is at a nodal point for a mode, then information about that mode will be missing from the frequency response function and modal parameter extraction methods may miss this mode. Therefore, the excitation and response

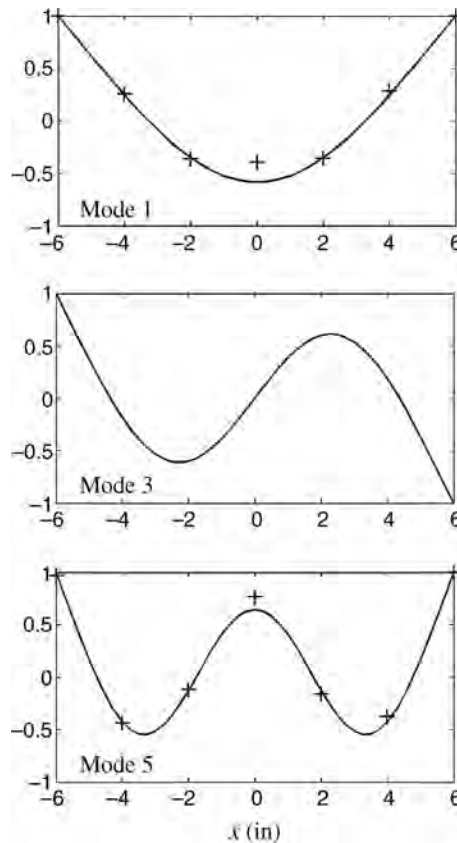


FIGURE 12.35 Profiles of the centerline ($y = 0$) deflections for the 1st, 3rd, and 5th mode shapes given in Figure 12.28. These mode shapes approximate the first three bending modes of an unsupported beam. The markers in (a) and (c) correspond to measured data demonstrating that a mode shape can be approximated by from the peaks of the imaginary part of the frequency response function. The center point has the largest deviation because the mass of the accelerometer affects the response at this location.

locations must be carefully chosen to obtain information about all frequencies of interest. In addition, to obtain more accurate values for the modal parameters at a given natural frequency, the frequency range of the measurement can be centered about this natural frequency.

In some cases, the single-degree-of-freedom modal analysis approximations are not sufficient and therefore multi-degree-of-freedom modal analysis methods must be used to obtain accurate modal parameters. In particular, the single-degree-of-freedom methods assume that the peaks of the frequency response function are well separated. However, if a structure has some closely spaced natural frequencies, this requirement may not be satisfied. In addition, large damping increases the width of the peaks in the frequency response function, which can cause significant peak overlap such that the single-degree-of-freedom methods would not be sufficiently accurate.

Multi-Mode Analysis Methods Multi-degree-of-freedom modal analysis methods estimate the modal parameters of several modes simultaneously. The methods assume that

the measured response data can be approximated by summing contributions from a finite number of modes, as illustrated schematically in Figure 12.32. Equations are used to express this summation in terms of the unknown modal parameters. Estimates for the modal parameters in a specified frequency range are then obtained by curve-fitting these equations to the measurement data.

A wide variety of modal estimation techniques have been developed based on different forms of the equations and different curve fitting techniques. For example, the equations can be expressed in the frequency domain or in the time domain. In addition, the methods can be applied to one set of data (corresponding to one load location and one response location) or to multiple sets of data obtained for a structure, simultaneously. The modal analysis software will determine what methods are available and therefore, specific methods will not be presented here. A full discussion of the different formulations is well beyond the scope of this chapter and therefore only a brief overview is given here. Results from the analysis of one set of experimental measurements will be used to illustrate the process.

At least four quantities need to be specified when using modal analysis software to extract the modal parameters: (1) the frequency range for the analysis, (2) an estimate of the number of resonance frequencies within the specified frequency range, (3) whether to include estimates of residuals from modes outside the analysis frequency range, and (4) the number of response data sets to use (and which ones).

For illustration of the modal parameter estimation process, measurements on the example plate were obtained with an accelerometer at location 10 for roving instrumented hammer excitation at locations 1–35. The resulting frequency response functions are overlaid in Figure 12.36 for the frequency range from 0 Hz to 800 Hz. This frequency range was chosen because it includes frequencies corresponding to the first five modes shown in

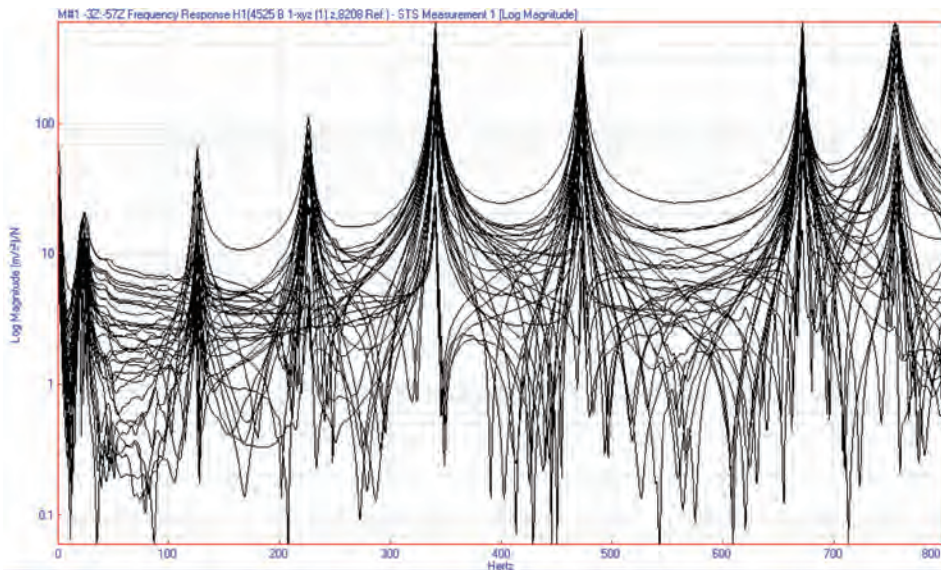


FIGURE 12.36 Amplitude of the frequency response functions (FRFs) for the plate shown schematically in Figure 12.27, measured with an accelerometer at location 10 for a hammer excitation at locations 1–35.

Figure 12.28. Because the accelerometer is offset from the centerline ($y = 0$), resonances are observed at the twisting natural frequencies as well as the bending natural frequencies. Recall that the lowest peak at 23 Hz corresponds to the rigid body mode of the plate vibrating on the foam support.

For the plate, identifying the number of resonance peaks in the frequency range of interest is an easy task because there are clearly differentiated peaks at all seven resonance frequencies that fall into this frequency range. With noisy data from a complex structure with many natural frequencies in the frequency range of interest, it may not be easy to determine a precise value to use for the number of degrees of freedom. For example, when two frequencies are close, their peaks may overlap creating one broader peak or there may be extraneous smaller peaks due to noise in the measurement. Also, peaks may be missing from a plot if the response or excitation location is at a nodal point.

Modal indicator functions can be used to help identify the resonance peak locations automatically, by analyzing multiple frequency response functions simultaneously. A plot of the modal indicator function for the plate measurements is shown on the bottom left of Figure 12.37. In the modal indicator function plot, a (gray) dot indicates a potential resonance mode and the plot shows that all seven resonance peaks were identified.

A technique to estimate the modal parameters in the frequency range of interest for the desired number of modes must be chosen. For the plate example, an orthogonal polynomial method accounting for four residuals was used. This method was applied to data from all 35 response locations over the frequency range from 10 Hz to 780 Hz

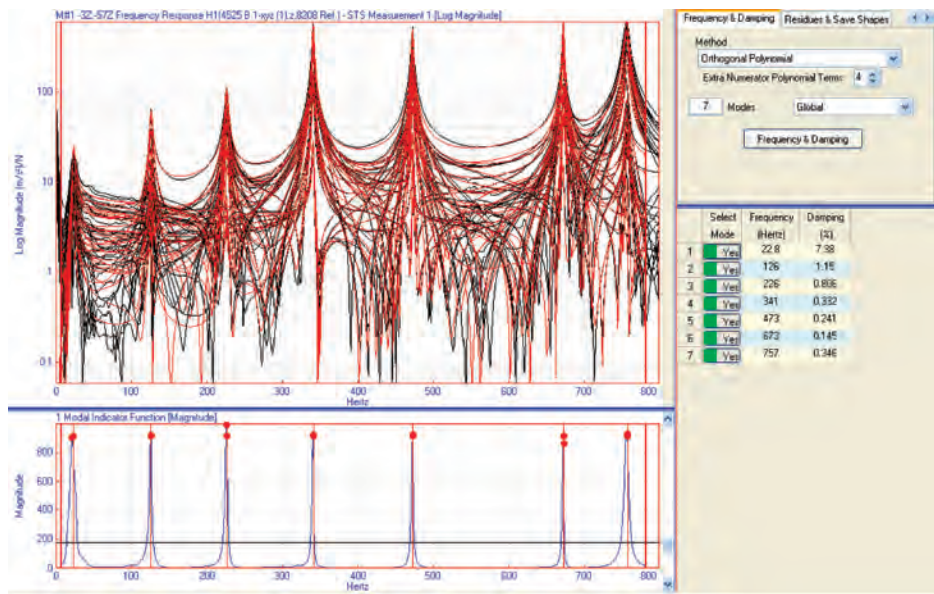


FIGURE 12.37 Modal parameter estimation results for the example plate shown schematically in Figure 12.27. In the upper left, frequency response functions (FRFs) predicted based on the estimated modal parameter are overlaid on the measured frequency response functions for an accelerometer at location 10 for a hammer excitation at locations 1–35. In the lower left, the dots at the peaks of the modal indicator function indicate potential resonance modes. On the upper right, the modal analysis method and parameters are displayed. On the lower right, the predicted modal frequencies and damping are displayed in a table.

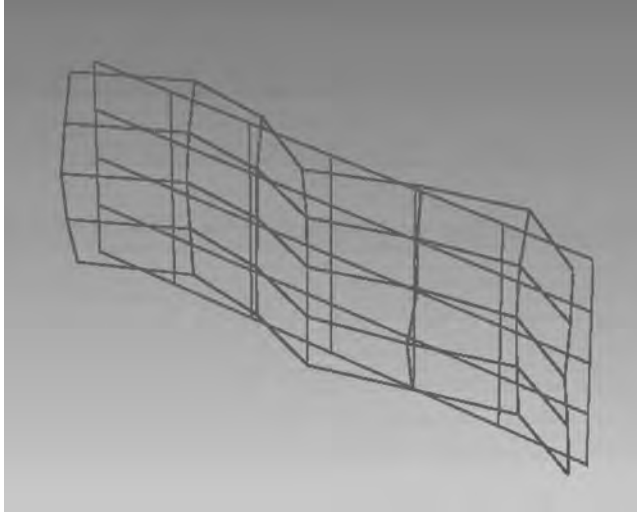


FIGURE 12.38 Mode shape estimated from the measured data at 673 Hz. This mode shape looks similar to the mode shape of mode 5 in Figure 12.28 that was predicted at 669 Hz using finite element analysis.

(shown in the top right of Figure 12.37) to estimate the modal parameters for all seven resonance frequencies. Estimated values for the frequency and damping of each of the seven modes for the plate example are shown in the table in the lower right of Figure 12.37.

Mode shapes corresponding to each resonance peak are also obtained from the analysis. The mode shapes can be plotted, as illustrated in Figure 12.38 for one mode or exported to use in verifying/updating a finite element model. The mode shape shown in Figure 12.38 estimated from the measured data at 673 Hz looks similar to the mode shape of mode 5 in Figure 12.28 that was predicted at 669 Hz using finite element analysis. Mode shape animations provide valuable information, for example, critical regions with large displacements or weak regions with large deformations can be identified from the mode shape animations.

To access the accuracy of the modal parameter estimation (i.e., the goodness of fit), frequency response functions predicted based on the estimated parameters can be compared to the measured frequency response functions. For the plate example, these two sets of curves are shown overlaid (in red and black, respectively) in the upper right plot of Figure 12.37. The good agreement is an indication of accurate modal parameter estimation.

The modal assurance criterion (MAC) is another analyzer software function that indicates the accuracy of the estimated modal parameters. The modal assurance criterion uses a correlation coefficient to compare two estimates of a mode shape obtained from experimental data. The modal assurance criterion can also be used to compare mode shapes obtained experimentally to those predicted by finite element analysis. If the two mode shapes are identical, they are perfectly correlated and the modal assurance criterion will equal 1. In practice, values >0.9 are considered good correlations.

Multiple estimates for the modal parameters can be obtained by repeating the modal parameter extraction process using various values for the number of degrees of freedom. Alternatively, the analysis can be repeated for different subsets of the response data or for

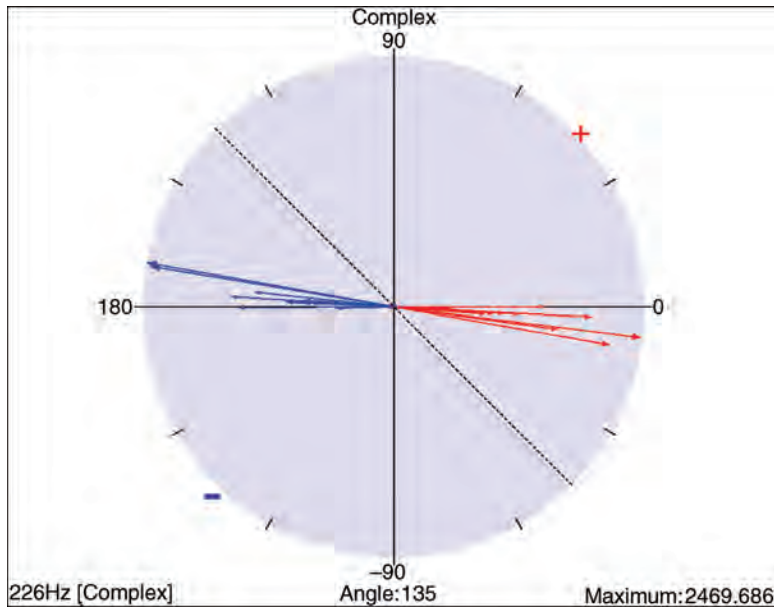


FIGURE 12.39 Complexity plot for the mode at 226 Hz, showing that the mode has low complexity (the vectors are nearly collinear).

slightly different frequency ranges that encompass the frequencies of interest. This redundant data can be used to identify modes that are consistently present for most or all choices of analysis parameters and data sets. The consistently determined modes are more likely to be genuine modes rather than artifacts of the analysis method. These modes are called “stable” and stability diagrams can be used to visualize the results of these multiple analyses and identify the genuine resonance frequencies.

A plot of the modal complexity (defined in Section 12.5.1.1) for the mode at ~ 226 Hz is shown in Figure 12.39. Each vector displays the amplitude and phase of one location (1–35) on the example plate of Figure 12.27. If the vectors are collinear, then the modal complexity is zero and the mode will vibrate as a standing wave. Damping is low in the steel plate and nonproportional damping is not present, so the vectors are nearly collinear and the modal complexity is low.

12.5.2 Operational Data Analysis

For the discussion of the modal analysis in Section 12.5.1, it was assumed that the measured response was generated by a single known applied excitation. Data that are gathered for unknown excitations, typically while the structure is in operation, are called *operational data*. Operational data may be the only option if the structure is too large or fragile to excite, or if a machine cannot be completely shut down for vibration testing. Operational loads only excite the modes of interest and therefore operational data may more quickly provide clues needed to solve existing vibration problems. In addition, operational data measurement requires less setup time, because the need for external excitation is eliminated. However, because the excitation is not measured, frequency response functions cannot be obtained explicitly from operational data.

12.5.2.1 Operational Deflection Shapes An *operational deflection shape* is typically defined as the shape of the deflection of a structure when vibrating at a single frequency. For example, an operational deflection shape can be obtained from measurements taken at various points on the structure, when the structure is excited by an unbalanced rotating shaft. Specifically, the shape is obtained from the amplitudes and phases of the peaks in the response data at the shaft rotational frequency. In order to obtain phase data, either all the measurement must be obtained simultaneously, or a reference measurement must be obtained at a fixed location for each additional frequency response function measurement. Operational deflection shapes may be useful in directly determining whether large deflections exist at critical locations of the structure during operation.

Note that operational shapes are different from mode shapes, even though the process of obtaining and displaying the shapes may be similar. In particular, a mode shape is an inherent characteristic of the structure associated with a natural frequency and independent of the applied force, whereas an operational deflection shape can be obtained for any frequency of excitation and depends on the location(s), direction(s), and amplitude(s) of the excitation. When the excitation frequency is near a natural frequency of the structure, then the operation deflection shape may be similar to the mode shape corresponding to that natural frequency.

12.5.2.2 Operational Modal Analysis *Operational modal analysis* (also called *output-only modal analysis* or *ambient modal analysis*) is the process of estimating modal parameters using operational response data only. Operational modal analysis techniques typically assume that the excitation resulting from operation is stationary white noise. Operational excitation that deviates from random, such as excitation due to a rotating shaft, will cause inaccuracies in modal parameter estimates. However, additional techniques have been developed that can be applied to operational data when the excitation also includes harmonic components (Mohanty and Rixen, 2005). These techniques either attempt to filter out the response due to the harmonic excitation or account for this extra response in the analysis. Curve fitting is still required over a modified set of frequency response functions.

Operational modal analysis can be used to estimate the natural frequencies, modal damping, and mode shapes of a structure. However, the modal mass cannot be estimated from operational data, so that frequency response functions cannot be obtained. However, techniques for estimating the modal mass have been developed when data are available from additional experiments with mass added to the structure.

Impact tests can also be performed on a machine that cannot be turned off, such as a very large power generation turbine. Measurements obtained by impacting a running machine with an instrumented hammer include responses to the operational loads superimposed onto the response to the hammer impact. Recordings of the machine's response excited by operational loads alone can be used to "negative average" out the operational responses. In this way the responses due solely to the impact excitation can be approximated. Modal analysis techniques described in Section 12.5.1 can be applied to the frequency response functions thus obtained to estimate the modal parameters.

12.6 APPLICATIONS OF VIBRATION MEASUREMENT

The basic vibration and measurement techniques described in the previous sections can be applied to a broad range of situations, whenever the dynamic response of a system is

important. Applications can be roughly separated into two categories: (1) structural system characterization and (2) machine condition monitoring.

12.6.1 Structural System Characterization

The dynamic response of a structural system or component can either be measured for a specific set of operating conditions or more generally characterized, so that the response of the structure or component to other excitations can be predicted. For the former, the vibration is typically excited by moving elements, for example, rotating shafts, either within or attached to the structure. For the latter, a known force generated by an external source, for example, a shaker, is generally used to excite the structure.

12.6.1.1 Operational Measurements Measurements of a structure's dynamic response during operation can provide a wealth of valuable information. By measuring the motion at critical points on the structure, "sweet spots," that is, speeds that minimize the amplitudes of undesirable vibrations, may be determined. Alternatively, the vibration response at many points on a structure can be measured under constant operating conditions to obtain operational deflection shapes. In this case, typically two accelerometers are used a reference accelerometer that remains at a fixed location and a second roving accelerometer, used to obtain the response at a number of locations on the structure. Operational deflection shapes can be used to identify locations where the structural deformations are largest so that these areas may be stiffened to improve the structure's performance. Alternatively, operational responses may be used to identify resonance conditions that result in high amplitude vibrations at a particular operating frequency, so this frequency can be avoided.

Noise generated by structural vibrations may be a critical factor in some applications. For example, to increase passenger comfort in automobiles, significant effort is put into reducing noise generated by road or engine induced vibrations. Therefore, in the automotive industry, vibration studies are grouped under the broader category of noise, vibration, and harshness (NVH). Vibration analysis tools and standards have been developed around the specific requirements of noise and vibration testing.

Vibrations cause oscillating stresses that can eventually lead to fatigue failure. Durability and fatigue tests can be used to determine the fatigue life of a structure or component. During testing, components are subject to dynamic excitations, either generated in situ by operating conditions or generated by an independent test structure. Accumulated damage and fatigue failures are observed over the duration of the test. Accelerated life testing can be used to reduce the test duration by increasing the test excitation amplitudes. A variety of algorithms exist to compress lifetime loading data into representative short duration loading. Component durability and lifespan are most accurately predicted when the loading conditions during the test closely reproduce the operational loading conditions. Characterizing the dynamic loads a component experiences in operation is critical, not only for durability and fatigue testing but also many other vibrations tests as well.

12.6.1.2 Modal Analysis Measurements Determination of the dynamics characteristics of a structural system or component provides additional information that can be used to investigate design or operational changes that may improve the system's performance. Modal analysis measurements of the structure's natural frequencies and mode shapes are needed to predict and understand its response to more general loadings. For example,

knowledge of the natural frequencies and mode shapes of a structure can be used to tune operating conditions to either enhance or reduce the structure's vibration amplitude.

Models are often used to investigate the effects of modifications of the structure on its dynamic response. Models are especially important when the operation loads cannot be applied directly to test the structure, for example, earthquake loads on a building or bridge and launch loads on a spacecraft or its payload. Traditionally, finite element models have been used. Modal analysis data can be used to validate and refine a finite element model and the refined model can be used to further investigate the structure dynamic response. More recently, structural dynamic modification methods have been incorporated in the modal analysis software that directly use the modal analysis or frequency response function data to predict the effect of design changes on the structural response.

When creating finite element models, many approximation are typically made and therefore, the response of the finite element model will not exactly predict the actual response of the structure. Joints, in particular, are difficult to model precisely. For example, a bearing, connecting a shaft to the main structure, may be modeled with simple linear spring and damper elements that approximate the stiffness and damping of the bearings. Modal analysis data can be used to validate a finite element model by comparing the predicted and measured frequencies and mode shapes. Measures, such as the modal correlation coefficient (MAC), are typically used to quantify the agreement. The modal correlation coefficient has a value between 0 and 1, with 1 implying exact correlation. If there are significant discrepancies between the predicted and measured frequencies and mode shapes, then the finite element model can be refined based on insight gained from the correlation analysis.

Once a sufficiently accurate finite element model has been established, then this model can be used for further predictions. The finite element model can be used to predict the response of the structure to additional loading scenarios that may occur during operation. If the structure is a component of a larger structure, then either a complete or reduced finite element model of the component may be added to a finite element model of the entire structure. This combined model can be used to determine the component's in situ response or its affect on the rest of the structure. If predicted response amplitudes are outside the design specifications for a structure, then finite element models can be used to investigate the effects of design modifications on its response to expected loading scenarios. This type of investigation can be used to identify designs with improved dynamic characteristics that satisfy the design criteria.

12.6.2 Machine Condition Monitoring

Measurements of the vibrations of a machine during operation can often be used to detect defects in machine components and monitor their severity. Often vibration measurements can be used to identify defects in their early stages, long before the defects can be sensed by other methods or before they have a significant effect on the machine's performance. This is true, even though the amplitudes of vibration initially caused by a defect may be orders of magnitude smaller than the amplitudes of other machine vibrations. The advantage of vibration measurement is that these small amplitude vibrations often generate a peak or set of peaks in the vibration spectrum of the machine that can be used to identify the defect. Displaying the vibration spectrum on a logarithmic or dB scale makes it possible to observe these lower amplitude peaks.

Vibrations induced by different components of a machine can be distinguished by the specific characteristics they create in frequency domain representations of the machine's vibrations during operation. For example, information about the peak frequencies and existence of side lobes or harmonics can help differentiate sources of the vibration. In addition, observations of changes in the vibration spectral response with operational speed may also help to distinguish between contributions from difference components.

Vibration measurements can be analyzed to identify a defective component causing a machine error, such as excessive cutter marks in machining, so that steps can be taken to reduce the error. Alternatively, vibration measurements can be used to periodically or continuously monitor a machine so that failing components can be identified and replaced efficiently, for example, before catastrophic failure occurs.

For machine monitoring, peaks in the vibration spectrum corresponding to the component(s) of interest should first be identified. Then, baseline measurements for the machine operating in good condition need to be established. These baseline measurements will depend on transducer type and location, machine operating speed, and any loading on the machine. Therefore, successive measurements need to be made under similar conditions when monitoring the machine performance. As a machine component deteriorates, the amplitude of the peaks in the vibration spectrum corresponding to this component will increase. Criteria for predicting component failure are developed based on accumulated vibration measurements. These criteria are usually defined as a limit on the maximum measured vibration amplitude at a frequency or over a range of frequencies corresponding to the component of interest. If this limit is exceeded, then the component is replaced.

A brief overview of the characteristics of a few common machine components and faults will be discussed in this section to demonstrate how the component geometry and type of defect determine the frequencies of the resulting vibration. Table 12.1 gives a list of a broader range of machine faults and their vibration characteristics. While a complete listing of all potential faults is not possible, additional information about fault detection and condition monitoring can be found in references on machine monitoring such as Rao (2000), Scheffer and Girdhar (2004), and Wowk (1991).

12.6.2.1 Unbalanced, Misaligned, or Bent Shaft No component is perfect, so any rotating shaft will produce some machine vibration. An unbalanced, misaligned, or bent shaft will produce vibration at the shaft rotational frequency, f_{shaft} (in Hz), and/or at its harmonics.

TABLE 12.1 Common Machine Faults and Their Typical Dominant Frequencies of Vibration

Nature of Fault	Frequencies of Dominant Vibration
Unbalanced shaft	f_{shaft}
Misaligned or bent shaft	f_{shaft} , $2f_{\text{shaft}}$, possibly harmonics
Damaged rolling element bearing	See Table 6.2
Damaged or worn gears	$N_{\text{teeth}} \times f_{\text{shaft}}$, N_{teeth} = number of teeth
Faulty belt drive	f_{belt} and harmonics (belt rotation frequency)
Oil-film whirl in journal bearings	$0.42f_{\text{shaft}}$ to $0.48f_{\text{shaft}}$
Mechanical looseness/rubbing	f_{shaft} , plus harmonics and possibly subharmonics

Unbalance of a shaft assembly (shaft with attached components) is the easiest machine defect to understand. For example, an unbalance is created when one blade of a fan or turbine is damaged, so that the blade mass is reduced or its center of mass is shifted. The magnitude of the unbalance is equal to me , where m is the mass of the shaft assembly and e is the eccentricity. The eccentricity is defined as the radial distance between the center of rotation and the center of mass of the shaft assembly. When the shaft is rotating with angular speed $\omega_{\text{shaft}} = 2\pi f_{\text{shaft}}$ (rad/s), this unbalance produces a centrifugal force $F_{\text{un}} = mew_{\text{shaft}}^2$ radially outward and rotating with the shaft. The component of this force in a direction perpendicular to the shaft axis will vary harmonically in time with a frequency equal to the shaft rotational frequency. The machine vibrations generated by this force will have a dominant frequency at the shaft rotation frequency, f_{shaft} .

The need to balance shaft assemblies increases as the rotational speed increases, because the unbalance force is proportional to the square of the shaft rotational speed. Both static and dynamic balancing may be needed to reduce vibrations to acceptable levels. Static balancing reduces the unbalance force by adding a small mass \bar{m} at a radius \bar{e} radially opposed to the unbalance, such that $\bar{m}\bar{e} \approx me$. Alternatively, material can be removed from a radial position at the same angle as the unbalance, for example, by drilling a small hole. Dynamic balancing reduces the moment that exists when the principal axis of inertia of the shaft assembly is not parallel to the axis of rotation. Dynamic balancing can be accomplished by adding two masses in two different planes perpendicular to the axis of rotation so as to align the principal axis of inertia with the axis of rotation. For example, to dynamically balance an automobile tire, a weight is added both to the inside and outside of the rim.

Vibrations caused by misaligned or bent shafts are often dominated by a frequency at one or two times the shaft rotation frequency (f_{shaft} or $2f_{\text{shaft}}$) though some contributions higher harmonics may also be evident. Comparisons of vibration amplitude and phase in the radial and axial directions measured at different locations may be used to distinguish between these sources of vibration. See Scheffer and Girdhar (2004) or Rao (2000) for more details.

12.6.2.2 Rolling-Element Bearing Fault Characteristics Rolling-element bearing failure is a common problem in machinery. The dominant frequency of the resulting vibration depends on the rotational speed, the bearing geometry, and the type of defect. Defects can occur on the inner or outer race, on the balls themselves, and/or on the cage that surrounds the balls to keep them equally spaced. Equations for the frequencies of common rolling-element bearing faults are given in Table 12.2 in terms of the bearing geometry and number of rolling elements, N_{ball} .

If the cage is damaged, then vibration occurs at the frequency of the cage rotation speed, also called the fundamental train frequency. This frequency is approximately equal to half the shaft rotation frequency, $\cong (1/2)f_{\text{shaft}}$, because the cage is typically rotating between a fixed outer race and an inner race rotating at the shaft frequency. A defect on the inner or outer race generates vibration at the ball pass frequency, approximately N_{ball} times the fundamental train frequency, $\cong (N_{\text{ball}}/2)f_{\text{shaft}}$. A defect on a ball, will generate vibration at the ball spin frequency, $\cong (d_{\text{ball}}/2d_{\text{pitch}})f_{\text{shaft}}$, which is a fraction of the fundamental train frequency.

The characteristics of the vibration spectrum produced by a bearing change qualitatively as the bearing defects increase (Scheffer and Girdhar, 2004). For example, the bearing frequencies may appear with sideband frequencies, modulated with the shaft frequency (bearing defect frequency $\pm f_{\text{shaft}}$) or with natural frequencies of the bearing. As

TABLE 12.2 Equations for the Characteristic Bearing Defect Frequencies in Terms of the Shaft Rotation Frequency f_{shaft} , and the Bearing Geometry Where N_{ball} is the Number of Balls, d_{ball} the Ball Diameter, d_{pitch} the Pitch Diameter (Diameter of the Bearing Measured from the Center of Two Opposing Balls), and ψ , the Contact Angle

Defect Location	Dominant Frequency
Cage	$f_C = \frac{f_{\text{shaft}}}{2} \left(1 - \frac{d_{\text{ball}}}{d_{\text{pitch}}} \cos \phi \right)$
Ball	$f_B = \frac{f_{\text{shaft}}}{2} \frac{d_{\text{ball}}}{d_{\text{pitch}}} \left(1 - \left(\frac{d_{\text{ball}}}{d_{\text{pitch}}} \right)^2 \cos \psi \right)$
Outer race	$f_{\text{OR}} = N_{\text{ball}} f_C$
Inner race	$f_{\text{IR}} = N_{\text{ball}} (f_{\text{shaft}} - f_C)$

The rotation rate of the cage, f_C , is called the fundamental train frequency.

the bearing deteriorates further, the peaks may broaden as they increase in amplitude, making it more difficult to identify specific frequency contributions.

12.6.3 Other Machine Fault Characteristics

As illustrated for rolling-element bearing defects, the frequencies of vibration generated by a component defect depend on the speed of operation and geometry of the component. A partial list of frequencies generated by other component defects and machine faults is given in Table 12.1. Such a table should be used with caution, however, because vibration characteristics are machine dependent and can be quite complex. Additional references (e.g., Randall, 2011; Piesol and Paez, 2010; Scheffer and Girdhar, 2004; Rao 2000; Goldman, 1999; Wowk, 1991) can be consulted for additional techniques to identify and analyze machine fault characteristics.

NOMENCLATURE

a	Acceleration (m/s^2)
A_i	Amplitude factor for the i th mode shape (m)
$A(f)$	(Discrete) Fourier transform of $\ddot{x}(t)$ (m/s^2)
c	Viscous damping coefficient (N/s/m)
c_i	i th viscous damping coefficient for a multi-degree-of-freedom system (N/s/m)
C	Viscous damping matrix for a multi-degree-of-freedom system (N/s/m)
d_{ball}	Ball diameter of a bearing
d_{pitch}	Pitch diameter of a bearing
e	Eccentricity (m)
f	Frequency (Hz)

Δf	Frequency resolution (Hz)
f_{belt}	Belt frequency (Hz)
f_{B}	Ball rotation frequency of a ball bearing (Hz)
f_{c}	Cage frequency of a ball bearing (Hz)
f_{d}	Frequency of damped vibration for a one-degree-of-freedom system (Hz)
f_i	i th natural frequency (Hz)
$f_{\text{IR}}, f_{\text{OR}}$	Inner race and outer race frequencies of a ball bearing (Hz)
$f_{\text{L}}, f_{\text{R}}$	Frequencies of the left and right half power points (Hz)
f_{max}	Maximum frequency for the discrete Fourier transform
$f_{\text{n}} = \frac{\omega_{\text{n}}}{2\pi}$	Natural frequency for a one-degree-of-freedom system (Hz)
f_{peak}	Frequency at a peak of a response versus frequency plot (Hz)
$f_{\text{peak}i}$	Frequency at the i th peak of a response versus frequency plot (Hz)
f_{shaft}	Frequency of rotation of a shaft (Hz)
$F(t)$	Force applied to a one-degree-of-freedom system (N)
$F(f)$	(Discrete) Fourier transform of $F(t)$ (N)
$\mathbf{F}(t)$	Force vector for a multi-degree-of-freedom system (N)
$F_i(f)$	(Discrete) Fourier transform of $F_i(t)$ (N)
F_0	Amplitude of a harmonic force applied to a one-degree-of-freedom system (N)
f_{d}	Viscous damping force for a one-degree-of-freedom system (N)
f_{s}	Spring force (N)
F_{un}	Force produced by an unbalance (N)
$H(f)$	Frequency response function (m/N)
$H_{ij}(f)$	Frequency response function for the response at location i due to an excitation at location j
i	Integer index
j	Integer index
j	$\sqrt{-1}$ in Section 12.4.2.1
k	Spring stiffness (N/m)
k_i	i th spring stiffness for a multi-degree-of-freedom system (N/m)
\mathbf{K}	Stiffness matrix for a multi-degree-of-freedom system (N/m)
m	Mass (kg)
m_i	i th mass for a multi-degree-of-freedom system (kg)
me	unbalance magnitude (kg/m)
\mathbf{M}	Mass matrix for a multi-degree-of-freedom system (kg)
n	Positive integer
N	Positive integer: number of degrees of freedom in Section 12.3.2. number of samples in Section 12.3.3
N_{ball}	Integer number of balls in a bearing
N_{teeth}	Integer number of teeth of a gear

Q	Quality factor
t	Time (s)
Δt	Time between samples for a constant sampling rate (s)
t_i	Time at the i th peak of the response for a one-degree-of-freedom system (s)
t_{\max}	Maximum time for sampled data (s)
T_d	Period of vibration for a one-degree-of-freedom system (s)
$T_i(t)$	Time dependence for the i th mode shape
v	Velocity (m/s)
v_0	Initial velocity for a one-degree-of-freedom system (m/s)
\mathbf{v}_0	Initial velocity vector for a multi-degree-of-freedom system (m/s)
$V(f)$	(Discrete) Fourier transform of $\dot{x}(t)$ (m/s)
$w(x, y, t)$	Transverse displacement of a plate (m)
x	Displacement (m)
(x, y, z)	Position vector in rectangular Cartesian coordinates (m)
x_i	Displacement of the i th degree of freedom (m)
x_0	Initial displacement for a one-degree-of-freedom system (m)
x_s	Extension of a spring (m)
$\dot{x} = \frac{dx}{dt}$	Velocity (m/s)
$\ddot{x} = \frac{d^2x}{dt^2}$	Acceleration (m/s ²)
\mathbf{x}	Displacement vector for a multi-degree-of-freedom system (m)
\mathbf{x}_0	Initial displacement vector for a multi-degree-of-freedom system (m)
$\dot{\mathbf{x}}$	Velocity vector for a multi-degree-of-freedom system (m/s)
$\ddot{\mathbf{x}}$	Acceleration vector for a multi-degree-of-freedom system (m/s ²)
X	Harmonic response amplitude for a one-degree-of-freedom system (m)
$X(f)$	(Discrete) Fourier transform of $x(t)$ (m)
$X_i(f)$	(Discrete) Fourier transform of $x_i(t)$ (m)
X_R	Reference used for dB values (m)
δ	Logarithmic decrement
ζ	Nondimensional damping ratio for a one-degree-of-freedom system
ζ_i	i th nondimensional damping ratio
θ_i	i th phase shift (rad)
θ_{err}	Angle misalignment of a measurement transducer
ϕ	Phase (rad)
$\phi_i(f)$	Phase of the (discrete) Fourier transform $X_i(f)$
$\varphi^{(i)}, \boldsymbol{\varphi}^{(i)}$	i th mode shape
ψ	Contact angle in a ball bearing
ω_d	Frequency of damped vibration for a one-degree-of-freedom system (rad/s)
ω_{di}	i th frequency of damped vibration (rad/s)

ω_i	i th natural frequency (rad/s)
ω_n	Natural frequency for a one-degree-of-freedom system (rad/s)
ω_{shaft}	Frequency of rotation of a shaft (rad/s)
AC	Alternating current
AD	Analog to digital
DC	Direct current
DFT	Discrete Fourier transform
FFT	Fast Fourier transform
FRF	Frequency response function
IEPE	Internal electronic piezoelectric
LVDT	Linear variable differential transformer
MAC	Modal assurance criterion
MIMO	Multiple input/multiple output
NVH	Noise, vibration, and harshness
SIMO	Single input/multiple output
SISO	Single input/single output
PE	Piezoelectric
$Re[]$	Real part of a complex number

REFERENCES

- Avitabile P. Experimental modal analysis: A simple non-mathematical presentation. *Sound and Vibration* 2000;35(1):20–31.
- Batel M. Operational modal analysis – Another way of doing modal testing. *Sound and Vibration* 2002;36(8):22–27.
- Brigham EO. *The fast Fourier transform and its applications*. Prentice Hall; 1988.
- Bucher I, Ewins DJ. Modal analysis and testing of rotating structures. *Phil. Trans. R. Soc. Lond. A* 2001;359:61–96.
- De Silva CW. *Vibration: Fundamentals and practice*. 2nd ed., CRC Press; 2007.
- Ewins DJ. *Modal testing: Theory, practice, and application*. 2nd ed., Research Studies Press LTD: Baldock, Hertfordshire, England; 2000.
- Friswell M, Mottershead JE. Finite element model updating in structural dynamics. In: Gladwell GML, editor. *Solid Mechanics and Its Applications*. Vol. 38, Kluwer Academic Publishers; 2010.
- Goldman S. *Vibration spectrum analysis: A practical approach*. 2nd ed., Industrial Press Inc.; 1999.
- Kelly SG. *Fundamentals of mechanical vibrations*. 2nd ed., McGraw Hill; 2000.
- Maurer A, Sauer J, Steger H. Laser Doppler vibrometry: Vibration measurement with light – simple and versatile. *Laser+Photonics* 2010;2011(5):44–47.
- McConnell KG, Varoto PS. *Vibration testing: Theory and practice*. 2nd ed., Hoboken, New Jersey: Wiley; 2008.
- Mohanty P, Rixen DJ. Identifying mode shapes and modal frequencies by operational modal analysis in the presence of harmonic excitation. *Experimental Mechanics* 2005;45(3):213–220.
- Petyt M. *Introduction to finite element vibration analysis*. 2nd ed., Cambridge University Press; 2010.

- Piersol AG, Paez TL. *Harris' shock and vibration handbook*. 6th ed., McGraw Hill; 2010.
- Randall RB. *Vibration-based condition monitoring: Industrial, automotive and aerospace applications*. Wiley; 2011.
- Rao JS. *Vibratory condition monitoring of machines*. CRC Press; 2000.
- Rao SS. *Mechanical vibrations*. 5th ed., Prentice Hall; 2011.
- Scheffer C, Girdhar P. *Practical machinery vibration analysis & predictive maintenance*. edited by Scheffer C, series editor Mackay S, Elsevier; 2004.
- Wowk V. *Machinery vibration: Measurement and analysis*. McGraw-Hill; 1991.

13

ACOUSTICAL MEASUREMENTS

BRIAN E. ANDERSON, JONATHAN D. BLOTTER, KENT L. GEE, AND
SCOTT D. SOMMERFELDT

- 13.1 Introduction
 - 13.1.1 Acoustical measurement standards
- 13.2 Fundamental measures
 - 13.2.1 Sound pressure
 - 13.2.2 Sound power
 - 13.2.3 Sound intensity
 - 13.2.4 Decibel scale
 - 13.2.5 Frequency weightings
 - 13.2.6 Octave frequency bands
- 13.3 Microphones
 - 13.3.1 Condenser microphone
 - 13.3.2 Electret condenser microphones
 - 13.3.3 Dynamic microphone
 - 13.3.4 Microphone selection
 - 13.3.5 Microphone accuracy
 - 13.3.6 Calibration
 - 13.3.7 Pistonphone calibration
 - 13.3.8 Reciprocity calibration
 - 13.3.9 Relative calibration
 - 13.3.10 Switching calibration
- 13.4 Sound pressure level measurements
 - 13.4.1 Sound level meters
 - 13.4.2 Sound pressure level metrics
- 13.5 Measurement of sound isolation
 - 13.5.1 Transmission loss
 - 13.5.2 Insertion loss
 - 13.5.3 Noise reduction
 - 13.5.4 Sound isolation of partitions
 - 13.5.5 Sound transmission class
- 13.6 Room acoustics measurements
 - 13.6.1 Reverberation

- 13.6.2 Room resonances
 - 13.6.3 Critical distance
 - 13.6.4 Echoes
 - 13.6.5 Noise criteria
 - 13.7 Community and environmental noise
 - 13.7.1 Representations of community noise data
 - 13.7.2 Noise surveys
 - 13.8 Sound intensity measurements
 - 13.8.1 Sound intensity via the “ p - p ” principle
 - 13.8.2 Intensity probes
 - 13.8.3 Systematic measurement errors
 - 13.8.4 Transducer mismatch
 - 13.8.5 Sound intensity applications
 - 13.9 Sound power measurements
 - 13.9.1 Measurement of the sound power level in a free field
 - 13.9.2 The reverberation method
 - 13.9.3 Comparison method
 - 13.10 Sound exposure measurements
 - 13.10.1 Sound exposure level
 - 13.10.2 Noise dosage
 - 13.10.3 Equipment
- References

13.1 INTRODUCTION

As noise continues to grow as a significant design concern, the importance of proper measurement techniques is becoming more crucial. The focus of this chapter is to provide a working knowledge of the basic acoustical or sound measurements and measurement techniques for the practicing engineer. The chapter will also provide a broad overview and introduction to more advanced acoustical measurement methods. There are several published documents including tutorials on the proper methods for measuring sound. This chapter is heavily referenced pointing the reader to these excellent sources of information.

The chapter first provides a discussion of the acoustical quantities that are typically measured. Definitions of the terms, measurement units, and weighting filters are presented. The hardware used to make the measurements is then explained along with principles of calibration and data acquisition. Detailed steps to the measurement of sound pressure are then given. Acoustical measurements important in the areas of sound isolation and room acoustics are then discussed. We then move to measurements of sound intensity and sound power, which are the most advanced measures of a sound field. The chapter is then concluded with a discussion of sound exposure and noise dosage measurements.

13.1.1 Acoustical Measurement Standards

An important source of additional information beyond this chapter is the national and international standards that cover acoustical instruments, measurements, and calculations.

These standards are written by committees of experts on particular topics and represent recommended best practices. Some of the information in this chapter is drawn directly from these standards.

Three organizations that have developed standards related to acoustics instrumentation and measurements are the International Electrotechnical Commission (IEC), the International Organization for Standardization (ISO), and the American National Standards Institute (ANSI). Because these organizations deal with standardization well beyond the scope of acoustics, the specific technical committees that treat acoustical measurement standards are provided. Subcommittees SC1 and SC2 within ISO TC 43 deal with noise and building acoustics, respectively. IEC TC 29 treats the development of instrumentation under the topic of electroacoustics. ANSI Accredited Standards Committees (ASC) S1 (Acoustics) and S12 (Noise) deal specifically with acoustical hardware and measurements. ISO TC 108 and ANSI ASC S2 both treat mechanical vibration and shock, topics related to acoustical measurements. Numerous other organizations related to standardization and regulation of acoustic testing and the development of test codes are provided in Lang and Nobile (1998).

Standards generally begin with an introduction, explanation of the scope, and, in some cases, areas of application. The introductory sections are then usually followed by a list of normative references and definitions that aid in understanding the standardized hardware specifications or measurement procedures. Some example standards from ISO, IEC, and ANSI are described in Table 13.1.

TABLE 13.1 Various Standards Related to Acoustical Measurements

Standard Number	Standard Name	Description
ANSI S1.1-1994 (R2004)	Acoustical Terminology	Provides definitions for a wide variety of terms, abbreviations, and symbols used in acoustics
ANSI S12.19— 1996 (R2006)	Measurement of Occupational Noise Exposure	Presents methods that can be used to measure a person's noise exposure in the work place
ANSI S12.54— 1999/ISO 3744—1994 (R2004)	Acoustics—Determination of sound power levels of noise sources using sound pres- sure—Engineering Method in an Essentially Free Field over a Reflecting Plane	Method for measuring sound pressure levels on a measurement surface enveloping a noise source in order to calculate sound power level produced by the source
ISO 2923:1996	Acoustics—Measurement of noise on board vessels	Techniques and conditions for measurement of noise on board vessels. Results may be used to compare various vessels, and for example, assess audibility of acoustical alarms
ISO 9295:1988	Measurement of high- frequency noise emitted by computer and business equipment	Details four methods for determining sound power levels in the 11.2–22.4 kHz range
IEC 61260 (1995-2008)	Electroacoustics—Octave- band and fractional-octave- band filters	Performance requirements and methods for various implementations of bandpass filters that comprise a filter set or spectrum analyzer
IEC 61672-1 (2002-2005)	Electroacoustics—Sound Level Meters—Part 1: Specifications	Performance specifications for three kinds of sound measuring instruments

13.2 FUNDAMENTAL MEASURES

This section outlines some of the fundamental measures commonly encountered in acoustical measurements. First, the three fundamental measures of sound pressure, sound power, and sound intensity will be discussed. These three terms are discussed in most texts on acoustics (see Hunt, 1992; Pierce, 1989; Beranek, 1993; Kinsler et al., 2000). We will then discuss the decibel scale, frequency weightings, and octave frequency bands. It is crucial that the reader understand these basic fundamental measures before progressing to advanced topics in acoustical measurements.

13.2.1 Sound Pressure

Sound pressure, or acoustic pressure, refers to the small deviations of pressure from the ambient generated by sound waves. To better understand this wave phenomenon, take, for example, a U-shaped tuning fork as shown in Figure 13.1a. The tuning fork is a device used to create a tone at a specific frequency that can then be used to tune various musical instruments. The tone is created by striking one of the tines with a hammer causing the forks to freely vibrate as shown in Figure 13.1b. As the tine moves outward from equilibrium, the air adjacent to the outward moving tine is compressed and accelerated outward. As the air adjacent to the tine is pushed outward, the pressure at that location increases slightly above atmospheric pressure as the density of the air molecules also increases. This local increase in air density is commonly called a condensation. After the tine passes back through equilibrium and displaces in the opposite direction, the adjacent air molecules spread out leaving a pressure slightly lower than atmospheric pressure along with a decreased density of air molecules. This local decrease in air density is commonly

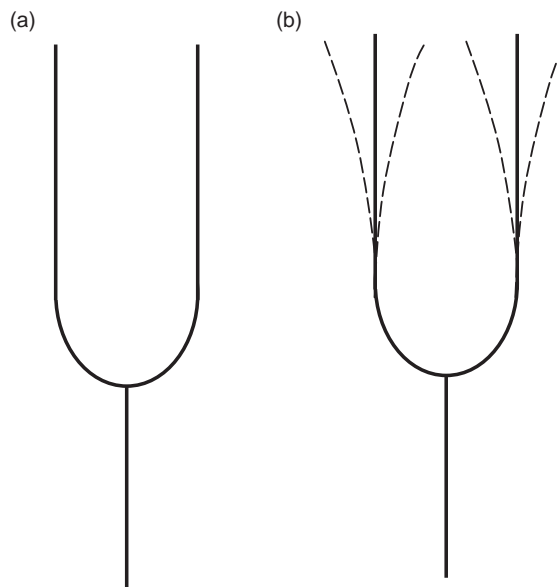


FIGURE 13.1 Illustration of the vibration of a tuning fork, (a) static position, (b) vibrating motion showing tine motion.

TABLE 13.2 Relationship of Sound Pressure and Sound Pressure Level for Various Examples

Sound Source	Sound Pressure (Pa)	Sound Pressure Level (L_p) (dB re 20 μ Pa)
Krakatoa (at 160 km)	20,000	180
M1 Garand being fired	12,000	176
Jet engine (at 30 m)	630	150
Threshold of pain	100	130
Hearing damage (short-term exposure)	20	120
Jet (at 100 m)	6–200	110–140
Jack hammer (at 1 m)	2	100
Hearing damage (8-h exposure)	0.5	88
Major road (at 10 m)	0.2–0.6	80–90
Passenger car (at 10 m)	0.02–0.2	60–80
Normal talking (at 1 m)	0.002–0.02	40–60
Very calm room	0.0002–0.0006	20–30

called a rarefaction. These deviations above and below atmospheric pressure are sound pressure fluctuations.

Everyday sound pressures may take on a very wide range of values (see Table 13.2). For example, the sound pressure from a typical jet engine is on the order of 630 Pa while the threshold of hearing is approximately 2×10^{-5} Pa at 1 kHz. Therefore, the term sound pressure level (L_p) or also referred to as SPL as defined by Equation (13.1) has been developed:

$$L_p = 20 \log \left(\frac{p}{p_{\text{ref}}} \right) \quad (13.1)$$

where p is the measured rms sound pressure, p_{ref} is the reference rms sound pressure usually taken as 20 μ Pa (2×10^{-5} Pa), which is the threshold of hearing for an undamaged ear and also corresponds closely with the reference power of 1×10^{-12} W as shown in a following section. Table 13.2 shows the relationship between sound pressure and sound pressure level for several acoustic environments. Sound pressure level measurements are the most common acoustical measurement made. The measurement can be made with a simple low-cost microphone or an expensive, comprehensive sound level meter. Procedures and methods for sound pressure measurements are discussed in Section 13.4.

13.2.2 Sound Power

Sound power is the term that describes the rate at which sound energy is radiated per unit time. The units of sound power are Joules/second or Watts. As an example of sound power, assume that there is a small spherical sound source. The sound power is a measure of the total amount of sound that the source can produce. Assuming that the source has constant sound power output and that unobstructed radiation occurs, the sound power output of the source is the same at any radius away from the source. A balloon analogy is often used to describe sound power. A balloon has the same amount of material

TABLE 13.3 Relationship Between Sound Power and Sound Power Level for Various Examples

Sound Source	Sound Power (W)	Sound Power Level (L_w) (dB re 1×10^{-12} W)
Saturn rocket	25–40 million	195
Rocket engine	1,000,000	180
Turbojet engine	10,000	160
Siren	1,000	150
Heavy truck engine/rock concert	100	140
75-Piece orchestra pipe organ	10	130
Piano/small aircraft/Jackhammer	1	120
Excavator/trumpet	0.3	115
Chain saw/blaring radio	0.1	110
Helicopter	0.01	100
Auto on highway/loud speech/shouting	0.001	90
Normal speech	1×10^{-5}	70
Refrigerator	1×10^{-7}	50
Auditory threshold	1×10^{-12}	0

surrounding the internal gas whether the balloon is partially or fully inflated. This is the same concept in that the sound power radiating from the source is constant, and is independent of distance from the source.

The range of sound powers encountered can be extremely large. For example, the sound power of a large rocket engine is on the order of 1,000,000 W, whereas the sound power of a soft whisper is approximately 0.000000001 W. Because of this large range in power values, a logarithmic scale known as the decibel scale is widely used in acoustic applications.

When sound power, W , is expressed in terms of decibels, it is called a sound power level, L_w , and is computed as

$$L_w = 10 \log \left(\frac{W}{W_{\text{ref}}} \right) \quad (13.2)$$

where W_{ref} is the standard reference power usually taken to be 1×10^{-12} W or 1 pW. Table 13.3 provides an example of the range of sound power and sound power levels.

Sound power may be measured in various ways depending on the environment in which the measurements are to be made (free-field, diffuse field, hemi anechoic, etc.). The two basic techniques of measuring sound power consist of using an array of pressure microphones placed around the source or using a sound intensity probe. Other comparative techniques also exist where measurements are compared with calibrated sources. The international standards that outline the processes for measuring sound power are ISO 3741, 3742, 3744, and 3745.

13.2.3 Sound Intensity

Sound intensity is defined as the amount of sound power per unit area of a given wave front. Relating this back to the balloon analogy, sound intensity would be a measure of

the balloon wall thickness at a certain location. The thickness of the balloon material will depend on how much the balloon is inflated, with the thickness decreasing as the balloon is further inflated. By analogy, sound intensity decreases as the distance from the source is increased, and vice versa. As previously mentioned, sound power and sound intensity are related by an area. For a spherical wave front, the relationship is expressed as

$$I = \frac{W}{4\pi r^2} \quad (13.3)$$

where I represents sound intensity, $4\pi r^2$ is the surface area of the sphere, and r is the radius of the sphere. From Equation (13.3), it should be noted that the decay of the sound intensity is directly proportional to the square of the radius of the sphere. This relationship is often referred to as the inverse square law.

Sound intensity values are directly related to sound power values by an area term and hence also have a very wide range of values. Again, it is therefore common to express the sound intensity using a decibel scale referred to as a sound intensity level. Sound intensity level is expressed as shown in Equation (13.4), where L_I is the sound intensity level, I is the measured intensity at some distance from the source, and I_{ref} is the reference sound intensity usually taken as $1 \times 10^{-12} \text{ W/m}^2$:

$$L_I = 10 \log \left(\frac{I}{I_{\text{ref}}} \right) \quad (13.4)$$

Sound intensity is the time-averaged product of the pressure and particle velocity. The pressure can simply be measured with a standard microphone. However, the particle velocity is somewhat more difficult to measure. One common method of measurement is to place two similar microphones a known distance apart. By doing this, the pressure gradient can be measured. Then using Euler's equation, the pressure gradient can be used to compute the particle velocity. This topic is expounded upon in the section of this chapter on sound intensity.

13.2.4 Decibel Scale

The decibel scale is a unitless logarithmic scale that is widely used in many fields of science including acoustics, electronics, signals, and communication. The decibel scale is used to describe a ratio of values, which among others may be sound power, sound intensity, sound pressure, and voltage. One main advantage of a logarithmic scale is that a very wide range of values can be reduced to a much smaller range, which provides the user with a better feel for the particular values of interest.

In the field of acoustics, the decibel (dB), which is 1/10 of a bel, is almost exclusively used. Decibels are used over bels to simply increase the sensitivity of the values, similar to using inches instead of feet or centimeters instead of meters when a pencil-sized object is measured. The term bel (in honor of Alexander Graham Bell) refers to the basic logarithm of the ratio of two values. The definition of the base 10 logarithm is shown in Equation (13.5)

$$10^{\log p} = p \quad (13.5)$$

The log of p is the value to which 10 must be raised in order to obtain the value p . The decibel equation can then be expressed as shown in Equation (13.6)

$$N(\text{dB}) = 10 \log \frac{A}{A_{\text{ref}}} \quad (13.6)$$

Two examples using the decibel scale are now provided to better illustrate the significance of the relationship.

EXAMPLE 13.1: *An increase in sound power of a ratio of 2:1 results in what gain in dB?*

Solution

$$N(\text{dB}) = 10 \log 2/1 = 10 \times (0.3010) = 3.01 \text{ dB}$$

Therefore, a doubling of the sound power results in an increase of approximately 3 dB.

EXAMPLE 13.2: *The noise level at a factory property line caused by 12 identical air compressors running simultaneously is 60 dB (assuming they are incoherent noise sources). If the maximum sound pressure level permitted at this location is 57 dB how many compressors can run simultaneously?*

Solution

For 1 compressor: $L_{p1} = 20 \log(P_{\text{rms1}}/P_{\text{ref}})$.

For 12 compressors: $P_{\text{rms12}}^2 = 12P_{\text{rms1}}^2$, therefore $P_{\text{rms12}} = \sqrt{12}P_{\text{rms1}}$, $L_{p12} = 20 \log(\sqrt{12}P_{\text{rms1}}/P_{\text{ref}}) = 20 \log(P_{\text{rms1}}/P_{\text{ref}}) + 20 \log \sqrt{12}$.

Therefore, $L_{p12} - L_{p1} = 20 \log \sqrt{12} = 10.79$ and $L_{p1} = L_{p12} - 10.79 = 60 - 10.79 = 49.21 \text{ dB}$.

Now for x compressors

$$L_{p\text{max}} = 20 \log \sqrt{x} + L_{p1}$$

Therefore, $\log \sqrt{x} = (L_{p\text{max}} - L_{p1})/20 = (57 - 49.21)/20 = 0.3895$ $x = (10^{(0.3895)^2})^2 = 6.01$.

Therefore, six compressors, at most, can be operated in order to meet the 57 dB limit.

13.2.5 Frequency Weightings

Besides the conventional decibel scale, there are other scales have been developed that filter or weight the sound according to how human ears respond to amplitudes at different frequencies. This chapter continues by presenting and discussing several of these filtered scales.

Fletcher and Munson (1933) presented what is typically called the equal loudness contours (see Figure 13.2). This set of contours represents how human ears perceive tones of equal loudness over the frequency range from 20 Hz to 15 kHz. Apparent loudness is given in units of phons. The magnitude for the phon unit comes from the intensity level in dB of a 1 kHz tone with the same perceived loudness as other tones at different

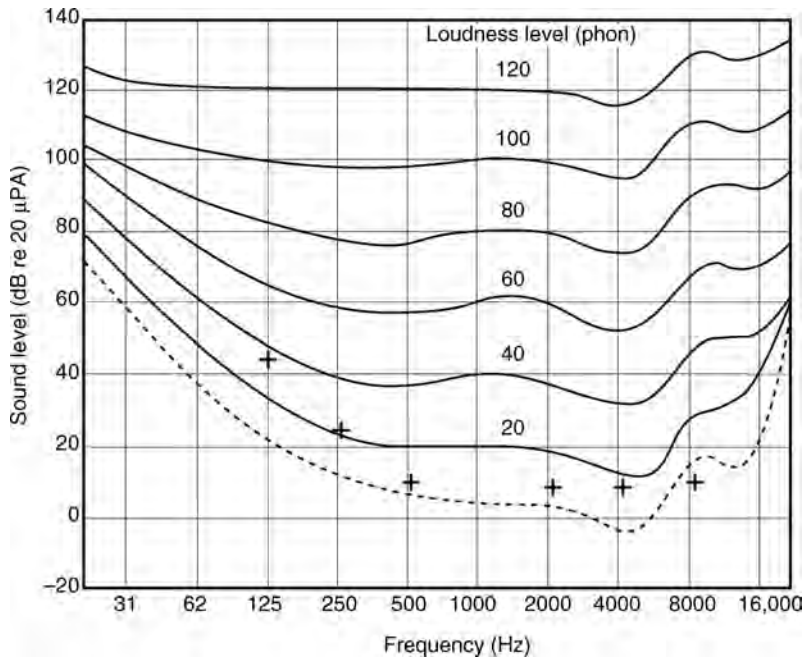


FIGURE 13.2 Equal loudness level contours for sinusoidal sounds. Dashed contour represents the threshold of hearing. (After Robinson and Dadson, 1957; Davis and Silverman, 1978).

frequencies and intensity levels. Therefore, a dB value and a phon value are the same only for a 1 kHz tone. These contours were determined experimentally in an anechoic chamber. The experiment consisted of playing pure tones at a specific intensity level, along with a 1 kHz reference signal. The intensity of the reference signal was increased until the two tones were perceived to be of the same loudness by the listener. Other work, by Robinson and Dadson (1957) and Davis and Silverman (1978), on the development of equal loudness contours has been generated since Fletcher and Munson. The most recent standards can be found in ISO 226:2003.

Weighting filters have been developed to adjust sound pressure levels measured by microphones to mimic how our ears respond according to these equal loudness contours. The first weighting filter to be discussed is called A-weighting and usually carries the associated symbol dBA or dB(A). This weighting was developed to nominally apply the same filter to an incident sound that the human ear applies for tones that lie on the 40-phon equal loudness contour. The 40-phon contour represents relatively quiet tones. It should be noted that the A-weighting curve roughly corresponds to the inverse of the 40-phon equal loudness contour developed by Fletcher and Munson. The A-weighting curve tends to amplify signals in the 1–6 kHz region. Sound outside this frequency range is reduced by the A-weighting. The A-weighting filter is shown in Figure 13.3.

The A-weighting filter is very popular and is used in many applications. Some of these include environmental noise, noise control, construction noise, community noise standards, and so on. The usage of the A-weighting filter for such a wide range of sound pressure levels is in spite of the fact that the A-weighting filter is only appropriate for relatively quiet pure tones (the 40-phon contour).

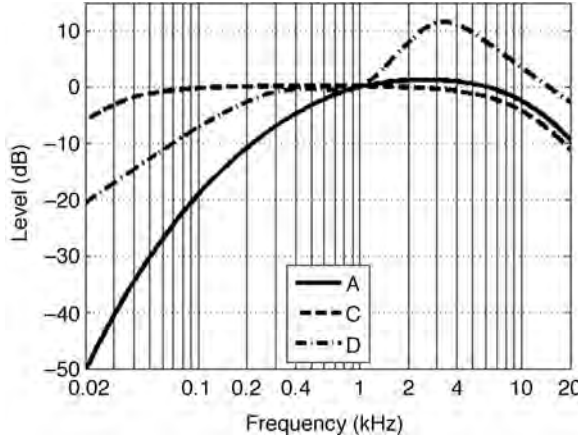


FIGURE 13.3 Common frequency weighting curves based on equal loudness contours.

The dB(B) and dB(C) scales were developed along similar methods of the dB(A) scale except that they were developed for tones with higher sound pressure levels. The dB(B) scale is seldom used but was developed for tones between the low sound pressure level of the A-weighted scale and the high sound pressure level of the C-weighted scale. The dB (C) scale is used for the high sound pressure levels (based off of the 90 phon curve) and as shown in Figure 13.3 is nearly flat over a large frequency range. The dB(D) scale has been developed for aircraft noise and is depicted in Figure 13.3. The D-weighting curve heavily penalizes high frequencies to which the ear is most sensitive. The equations for the weighting scales are shown in Equation (13.7) where $s = j\omega$, $j = \sqrt{-1}$ and ω is the angular frequency.

$$\begin{aligned}
 G_A(s) &= \frac{k_A \times s^4}{(s + 129.4)^2 (s + 676.7) (s + 4636) (s + 76,655)^2} \\
 G_B(s) &= \frac{k_B \times s^3}{(s + 129.4)^2 (s + 995.9) (s + 76,655)^2} \\
 G_C(s) &= \frac{k_C \times s^2}{(s + 129.4)^2 (s + 76,655)^2} \\
 G_D(s) &= \frac{k_D \times s \times (s^2 + 6532s + 4.0975 \times 10^7)}{(s + 1776.3) (s + 7288.5) (s^2 + 21,514s + 3.8836 \times 10^8)}
 \end{aligned} \tag{13.7}$$

where $k_A \approx 7.39705 \times 10^9$, $k_B \approx 5.99185 \times 10^9$, $k_C \approx 5.91797 \times 10^9$, and $k_D \approx 91104.32$.

Another unit of measure for perceived loudness is the sone. The sone was developed to relate a perceived doubling of loudness in phons to a sones number that doubles. Experiments found that when volunteers were asked to adjust the loudness of a tone until it was perceived to be twice as loud as the original tone that the doubling corresponded to an increase of 10 phons. One sone is arbitrarily equal to 40 phons, and a doubling of the loudness (in sones) corresponds to a doubling of the perceived loudness. The relationship between sones and phons is given in Table 13.4.

TABLE 13.4 Table of Phons to Sones Conversion

Loudness Level (Phones)	Loudness (Sones)
20	0.12
30	0.4
40	1
50	2
60	4
70	8
80	16
90	32
100	64

13.2.6 Octave Frequency Bands

In reference to music theory, the term octave corresponds to a doubling or a halving of frequency. For example, 200 Hz is one octave above 100 Hz and is one octave below 400 Hz. 400 Hz would be considered to be two octaves above 100 Hz and so on.

In acoustics, the term “octave band” is commonly used to represent a range of frequencies where the highest frequency is twice the lowest frequency. These octave bands are commonly used to represent what is going on in the acoustic signal over a band or range of frequencies as compared to discrete frequencies. Many sound level meters will output sound pressure level data in standardized octave bands.

Octave bands are typically specified by a center frequency f_C . The upper frequency limit, f_U , and the lower frequency limit, f_L , of the band are computed as shown in Equation (13.8):

$$f_L = \frac{f_C}{\sqrt{2}}, f_U = \sqrt{2}f_C, \text{ where } f_C = \sqrt{f_L f_U} \quad (13.8)$$

Furthermore, the bandwidth for a given octave can be computed by Equation (13.9).

$$\text{Bandwidth} = f_U - f_L = \frac{f_C}{\sqrt{2}} \quad (13.9)$$

In the field of acoustics, standard octave bands have been defined. The lower frequency, center frequency, and upper frequency for the standard acoustical octave bands are shown in Table 13.5.

In many applications, a higher resolution on the frequency axis is desired. In these cases, the octave band is split into smaller groups. A common way to do this is to split up the octave band into three smaller bands called 1/3-octave bands. The 1/3-octave band scale is commonly used for environmental, noise control, and building acoustics among others. The lower, center, and upper frequencies for the 1/3-octave bands are shown in Table 13.6 and calculated using Equation (13.10).

$$f_L = \frac{f_C}{\sqrt[3]{2}}, f_U = \sqrt[3]{2}f_C, \text{ where } f_C = \sqrt{f_L f_U} \quad \text{Bandwidth} = f_C \left(\sqrt[3]{2} - \frac{1}{\sqrt[3]{2}} \right) \quad (13.10)$$

TABLE 13.5 Octave Band Lower, Center, and Upper Band Frequencies (Hz)

Lower Band Frequency f_L	Center Band Frequency f_C	Upper Band Frequency f_U
22.4	31.5	45
45	63	90
90	125	180
180	250	355
355	500	710
710	1,000	1,400
1,400	2,000	2,800
2,800	4,000	5,600
5,600	8,000	11,200
11,200	16,000	22,400

TABLE 13.6 1/3-Octave Band Lower, Center, and Upper Band Frequencies (Hz)

Lower Band Frequency f_L	Center Band Frequency f_C	Upper Band Frequency f_U
18.0	20.0	24.4
22.4	25.0	28.0
28.0	31.5 ^a	35.5
35.5	40	45
45	50	56
56	63 ^a	71
71	80	90
90	100	112
112	125 ^a	140
140	160	180
180	200	224
224	250 ^a	280
280	315	355
355	400	450
450	500 ^a	560
560	630	710
710	800	900
900	1,000 ^a	1,120
1,120	1,250	1,400
1,400	1,600	1,800
1,800	2,000 ^a	2,240
2,240	2,500	2,800
2,800	3,150	3,550
3,550	4,000 ^a	4,500
4,500	5,000	5,600
5,600	6,300	7,100
7,100	8,000 ^a	9,000
9,000	10,000	11,200
11,200	12,500	14,000
14,000	16,000 ^a	18,000
18,000	20,000	22,400

^aOctave band center frequencies.

13.3 MICROPHONES

This section will discuss various types of microphones. It will then cover important considerations in microphone selection, including the sound field in which the microphone will be used and how the physical characteristics of a microphone affect measurements. We will then discuss microphone accuracy specifications. The section will then conclude by discussing various calibration techniques.

A microphone is a device that converts acoustical energy in the form of a compression wave into electrical energy in the form of a time-varying voltage. There are several types of microphones and some of these will be discussed in this section. The most common types of microphones use a thin membrane referred to as a diaphragm to sense the incident wave energy. The compression and expansion of sound waves impinging on the diaphragm causes the diaphragm to vibrate. The mechanical motion of the diaphragm is coupled to a transducer, which then produces a varying voltage proportional to the diaphragm deflection. The coupling mechanism between the diaphragm and the transducer is what distinguishes different types of microphones.

The most common microphones as based on the transducing components are the condenser, electret, and dynamic microphones. This section provides a brief overview of these different microphones and the reader is also referred to the vast amount of information available on the Internet.

13.3.1 Condenser Microphone

Typical elements of a condenser microphone (also called a capacitor microphone or electrostatic microphone) are shown in Figure 13.4. One conducting plate, often referred to as the back plate, is fixed to the microphone casing, although it is electrically isolated from the casing. A metal or metallized plastic diaphragm is placed a small distance away from the back plate and acts as the second (and movable) conducting plate of a dynamic

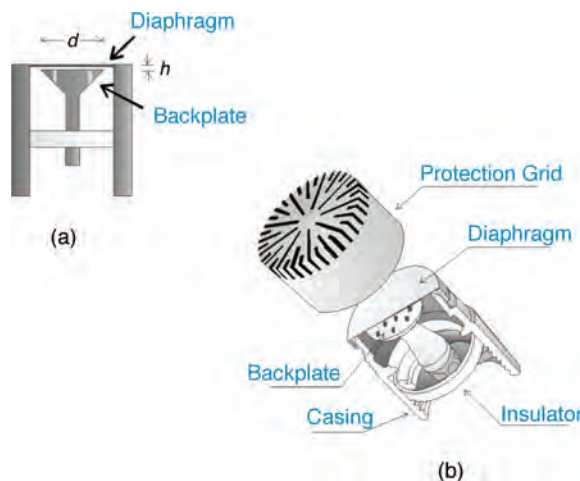


FIGURE 13.4 Cross section of a condenser microphone (Courtesy of G.R.A.S. Sound and Vibration).

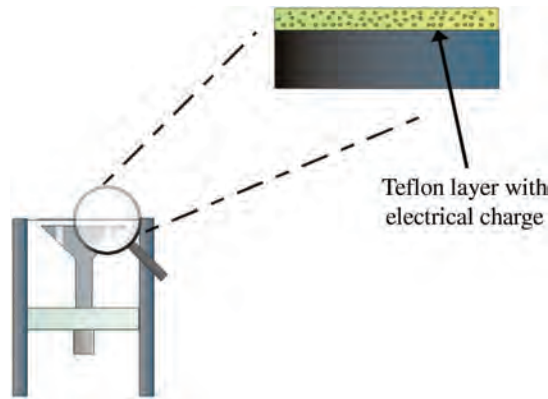


FIGURE 13.5 Illustration of the electrical charge applied to the diaphragm of an electret microphone (Courtesy of G.R.A.S. Sound and Vibration).

capacitor. When the diaphragm vibrates an oscillating voltage appears on the electrical leads attached to the plates. The principal advantage of this type of microphone is that it possesses a uniform frequency response over a wide band of frequencies. The principal disadvantage of these microphones is the need to have a DC bias voltage (e.g., 200 V) across the plates (in order to produce an electrical output signal and to linearize the response of the microphone) and a preamplifier in close proximity to the plates. Condenser microphone designs have typically incorporated a preamplifier in the microphone casing. Condenser microphones are used in research, to produce high-fidelity recordings, and in hearing aids and portable sound equipment.

13.3.2 Electret Condenser Microphones

Technically a subset of condenser microphones, electret microphones (sometimes called a prepolarized microphone) are now in wide use because they offer the advantages of a condenser microphone without the disadvantage of needing a large polarizing DC voltage. Figure 13.5 depicts an illustration of an electret microphone with an electrically charged Teflon diaphragm. The diaphragm consists of an electret plastic foil that has been permanently electrically charged and overlaid with a thin metallic layer. The electret condenser microphone typically employs a preamplifier. It is slightly less sensitive than the condenser microphone and the “permanent” electric charge has a finite life span (e.g., 5 years, but this depends on the design). One should note here that sometimes certain preamplifiers, used in conjunction with electret microphones, require a polarization voltage (e.g., 24 VDC drawing approximately 4 mA of current).

13.3.3 Dynamic Microphone

A dynamic microphone (also called an electrodynamic microphone or a moving-coil microphone) has a coil of wire attached to its diaphragm as shown in Figure 13.6. The transduction in a dynamic microphone is induced when the diaphragm moves its attached coil between the poles of a magnet, inducing an electromotive force (emf) and a

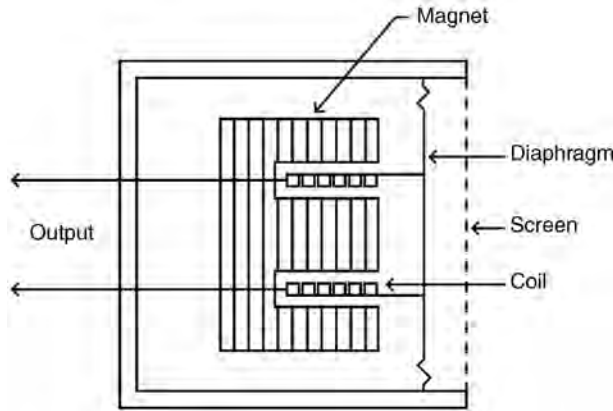


FIGURE 13.6 Illustration of the parts of a dynamic microphone.

corresponding electrical voltage signal (this is known as Faraday's law of induction). Dynamic microphones are capable of a relatively high gain before feedback. They are rugged, and capable of a broad frequency response over a wide dynamic range. Because they are able to withstand the high-intensity sound levels often associated with popular music, they are widely used for live performances and for recording sessions. The principal disadvantage of dynamic microphones is that their frequency response is not as uniform as for condenser microphones.

13.3.4 Microphone Selection

Most microphones are designed to work in one of three sound fields. These sound fields are typically classified as free field, diffuse field, and pressure field. The different microphone designs have been developed to minimize the effect of the microphone in the particular field of measurement. Because the correction for the influence of the microphone in the sound field can be built into the microphone body, care should be taken when selecting microphones for a particular sound field. Using the wrong microphone in a particular field may result in errors of several decibels.

Free-field conditions exist when the sound waves propagate in a specific direction without the effects of any reflections. These conditions could exist in an anechoic chamber, outdoors, or in other places where reflecting surfaces are not present (or provide minimal reflection). The microphone is typically placed in this field with the diaphragm perpendicular to the sound wave as shown in Figure 13.7. When the sound wave is not perpendicular to the diaphragm, corrections can be made if the angle of incidence is known.

A diffuse field, or random field, is created when sound waves from many different directions pass through a point. Statistically, in a diffuse field, sound waves have equal probability of arriving from any direction, and the sound field is considered to be spatially uniform in amplitude for a particular frequency (but not in general equal in amplitude across all frequencies). These sound waves may or may not be generated from the same source but typically multiple reflections are part of the sound field as well. These microphones, sometimes referred to as random incidence microphones (depicted in Figure 13.7

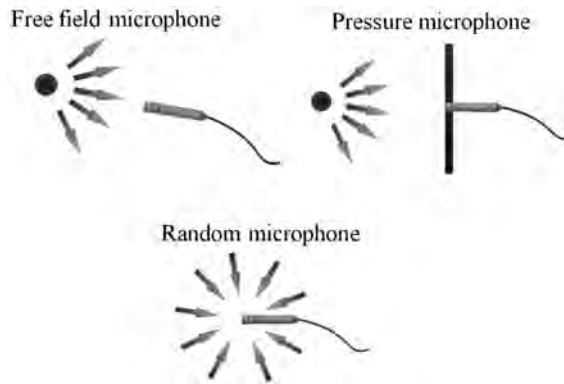


FIGURE 13.7 Illustrations of how free-field, pressure, and random incidence microphones are intended to be used (Image courtesy of G.R.A.S. Sound and Vibration).

as a “Random microphone”), are designed to be direction insensitive. This is achieved by designing their response to compensate for the fact that the presence of the microphone perturbs the sound field. This type of sound field would exist with a steady-state noise source in a closed room with reflecting walls.

Pressure-field conditions exist in enclosures where the dimensions are much smaller than the wavelength of the sound. This is typically achieved with a single source and the resulting field has the same magnitude and phase at all locations. The microphone, sometimes termed a pressure microphone, is typically placed such that the end of the microphone is flush with the enclosure surface (as depicted in Figure 13.7). These microphones are often used to measure the sound waves inside a duct or on a surface. As with the other microphones, pressure-field microphones are designed to account for the changes they make in the sound field.

13.3.5 Microphone Accuracy

Measurement microphones are also rated on a standard scale as given by ANSI S1.12-1967 (R1977). This particular standard rates the microphone on its calibration and type of application. This standard is summarized below and was taken from the Measurement Microphone primer by Brüel and Kjær, which can be found on their website (<http://www.bksv.com/library/primers.aspx> last viewed on February 2, 2011).

Type L: Precisely calibrated reference-standard with a closely specified outer diameter (to enable use in couplers).

Type XL: As above, but with no specified outside diameter.

Type M: For measuring sound-pressure magnitudes of the order 0.1 N/m^2 , or higher. Better high-frequency and high sound pressure performance than for Types L and XL. Copes with relatively large ambient-pressure changes.

Type H: For applications in which diffraction errors in the measurements must be small or in which the sound-pressure magnitude is of the order of 0.5 N/m^2 , or higher. Copes with relatively large ambient-pressure changes.

13.3.6 Calibration

Proper calibration is a vital process when making acoustic measurements. ANSI S1.10-1966 (R1986) gives the standard method for the calibration of microphones, although there are many different methods of microphone calibration. In this section, we provide an overview of the following types of microphone calibration: pistonphone, reciprocity, relative, and switching calibrations. Good resources for further reading on calibration is provided by ANSI S1.12-1967—R1977, ANSI S1.10-1966—R1986 (Burnett and Nedzel-nitsky, 1987; Torr and Jarvis, 1989; Kinsler et al., 2000, pp. 428–430; Maclean, 1940, pp. 140–146; Bobber, 1990).

13.3.7 Pistonphone Calibration

A pistonphone calibrator has an enclosed volume in which the microphone to be calibrated is inserted with an airtight seal. In the pistonphone, a mechanical cam drives a piston, or a pair of pistons, in a sinusoidal motion to create a pressure field at a steady frequency and at steady amplitude. The pressure in the enclosure can be accurately determined by knowing the internal dimensions of the enclosure, the ratio of the specific heats for the gas in the coupler (1.402 for air at 20°C and 1 atm), the atmospheric pressure, the cross section area of the piston(s), and the peak motion of the piston(s). The microphone recordings may then be calibrated by knowing this pressure field inside the pistonphone either electronically within a sound level meter, or in the postprocessing phase of data analysis.

Pistonphones are commercially available from most major microphone manufacturing companies. Given that the sound field is generated with a rotating cam and a moving piston, most pistonphones calibrate at low frequencies at or near 250 Hz (though some operate at 1 kHz). There is a little more variation on the level of pressure generated by a pistonphone but most commercial pistonphones calibrate at levels from 114 dB re 20 μ Pa to 134 dB re 20 μ Pa. One of the main sources of error with a pistonphone calibration is an accurate barometric pressure reading. However, depending on the barometer, pistonphones can be very robust and can be used to calibrate Class 0 and Class 1 systems following IEC standard 942, 1988 Sound Calibrators.

13.3.8 Reciprocity Calibration

Reciprocity calibration has become a very common primary calibration process. The calibration can be performed using an anechoic chamber or in an environment that can be considered to be a free-field environment. The calibration process consists of using three transducers (1–3). The first is the microphone to be calibrated, the second any reversible transducer that may act as both a transmitter or a receiver, and the third a source. The three transducers are organized into three pairs (3-2, 3-1, and 2-1). Each of the three pairs is tested separately. During the first test, the source is placed a distance, r , from the reversible transducer and the open-circuit voltage response is measured at the reversible transducer's terminals, V_2 . The distance r must be sufficiently large so that these measurements are made in the acoustic far field of the transducers used as sources. In the second test, the source outputs the same signal as in the first test and the microphone is put in the place of the reversible transducer. The open-circuit voltage response at the

microphone's terminals, V_1 , is then measured. Finally in the third test, the reversible transducer is used in place of the source and is now operated as a source, while the current through the reversible transducer is measured, I_2 . The open-circuit voltage response is now measured at the microphone's terminals, V'_1 . The open-circuit receiving sensitivity of the microphone, M_{O1} , is then

$$M_{O1} = \sqrt{\frac{2r}{\rho_0 f} \frac{|V_1||V'_1|}{|V_2||I_2|}} \quad (13.11)$$

where ρ_0 is the ambient air density.

13.3.9 Relative Calibration

In cases where one cannot afford expensive precision microphones, one may use lower quality microphones and calibrate them against a microphone with a known sensitivity, M_{KS} . In this method, sometimes referred to as a substitution technique, the precision microphone is placed in a repeatable broadband sound field and its open-circuit voltage response as a function of frequency is then measured, V_{KS} . The microphone of unknown calibration is then put in place of the precision microphone and its open-circuit voltage response is measured, V_{UM} . The sensitivity of the unknown microphone, M_{UM} , may then be determined as

$$M_{UM} = M_{KS} \frac{V_{UM}}{V_{KS}} \quad (13.12)$$

13.3.10 Switching Calibration

In some sound fields it may be desirable to employ multiple microphones. One need not possess multiple precision microphones if a switching calibration is employed, however, the microphones must be stable over the period in which the measurements are made. In the switching calibration technique, the microphones are placed in their desired locations in a sound field and a complex, frequency-dependent transfer function is measured between the two microphones, $H_{I,12}$. The microphones' positions are then interchanged and a second transfer function is measured, $H_{II,12}$. The calibration transfer function, $H_{CAL,12}$, between these two microphones is then

$$H_{CAL,12} = \sqrt{H_{I,12} \times H_{II,12}} \quad (13.13)$$

This calibration transfer function can then be used to determine the open-circuit voltage response of one of the two microphones used for the calibration if the other one is used by dividing the measured open-circuit voltage response by $H_{CAL,12}$.

Another example where the switching calibration technique is useful is in absorption coefficient measurements of acoustic ceiling tiles using a plane wave tube. The absorption

properties are measured in a plane wave tube through a measurement of a complex frequency response between two microphones. This complex frequency response must be calibrated by the measured frequency response by dividing the measured frequency response by $H_{\text{CAL},12}$.

13.4 SOUND PRESSURE LEVEL MEASUREMENTS

Since there is an extremely large range of pressure measurements the results are often presented as a logarithmic measure with units of decibels (reported in terms of dB) as previously discussed. This chapter will discuss the basics of sound level meters and then describe various types of typical quantities measured in field tests that are then used for community noise standards, and to determine worker sound exposure, as will be described in more detail later in this chapter.

13.4.1 Sound Level Meters

Sound pressure level measurements are typically made with a sound pressure level meter or sound level meter, SLM. Many manufacturers provide SLMs that provide a fast and simple method of making sound pressure level measurements. A SLM typically consists of a condenser microphone, preamplifier, data acquisition hardware and software, a display to present the results, and a storage system to save the recorded data. The microphone is typically mounted on a rod extending out past the data acquisition box to minimize sound reflections from the user and system during the test. A photograph of a typical SLM is shown in Figure 13.8.

There is a large range of quality and price on SLMs. A significant cost of the system is associated with the quality of microphone. Another significant cost factor is the capability of the system hardware and software. For example, storage space, real time spectral



FIGURE 13.8 Photograph of a typical sound level meter.

TABLE 13.7 Sound Pressure Level Measurement Standards

1. ISO 2204: 1979—Acoustics—Guide to International Standards on the measurement of airborne acoustical noise and evaluation of its effects on human beings. This standard provides a very good introduction to various acoustic terms and measuring methods
2. ANSI S1.4-1983 (R2006)/ANSI S1.4A-1985 (R2006)—American National Standard Specification for Sound Level Meters
3. IEC 61672 : 2003—International Standard for Sound Level Meter Performance

analysis, averaging, sample rate, automatic data logging, filtering, and weighting are some of the capabilities to consider when selecting a SLM.

To help the user select the appropriate SLM, a classification known as types or classes of sound level meters has been set. The two main classes of SLM classifications are Class 1 and Class 2. The basic difference between the two classes is the tolerance limit with which the specifications are met. Class 1 meters are designed to be used for laboratory and field applications where the highest accuracy is required. Class 2 meters can be considerably less expensive but do not provide the same high accuracy and precision, especially in the high frequency range. In general, specifications for Class 1 and Class 2 SLMs have the same design goals and differ mainly in the tolerance limits and the range of operational temperatures. Tolerance limits for Class 2 specifications are greater than, or equal to, those for Class 1 specifications. The details of these classifications can be found in International Standard IEC 61672-1. There are also Class 0 and Class 3 transducers. The Class 0 transducers are used in laboratory applications for the calibration of other meters. Class 3 transducers are used primarily in field studies where accuracy is not of extreme importance. ANSI S1.4-183 is another standard which discusses classifications 0-2. Some manufacturers will designate their microphone, preamp or filter as a certain type or to meet a particular standard. In order to conform, the complete system must be reviewed. The whole system is what must meet the standard, not just one component.

National and international standards exist for making sound pressure level measurements. The ANSI and the IEC are two organizations responsible for developing and maintaining these standards. The standards state both the measurement process which should be followed and the specifications of the equipment required to make the measurements. Several of the standards for sound pressure measurements are listed in Table 13.7.

13.4.2 Sound Pressure Level Metrics

There are numerous ways of representing measured noise. One way is to simply display the A-weighted sound pressure level (L_A) as a function of time. In addition to A-weighted sound pressure level, there are many other single-number metrics that are used to describe noise.

Perhaps, the most common metric is the equivalent continuous sound level (L_{eq}). The L_{eq} is commonly determined from A-weighted sound pressures but it may in general be expressed using other types of weightings (though the weighting used should always be specified, for example, the c-weighted sound pressure level would be L_C). Further it is the level of the steady sound that has the same time-averaged energy as the noise event.

For the time interval, T , which runs between T_1 and T_2 , L_{eq} is calculated as

$$L_{eq} = 10 \log_{10} \left\{ \frac{1}{4 \times 10^{-10}} \frac{1}{T} \int_{T_1}^{T_2} p_A^2(t) dt \right\} \quad (13.14)$$

where p_A is the instantaneous A-weighted sound pressure (a time domain A-weighting filter must be applied to an instantaneous pressure versus time waveform to obtain $p_A[t]$). A running, live update on a sound level meter of L_{eq} is often called a running L_{eq} . The $1/4 \times 10^{-10}$ is the reciprocal of the reference sound pressure squared $(20 \mu\text{Pa})^2$. The L_{eq} gives equal weighting to all noise summed over T irrespective of the time at which it is present. Common values for T include 1 s, 1 min, 1 h, 8 h, and 24 h.

If one has a set of A-weighted, L_{eq} measurements, each taken over a 1-h period of time [expressed as $L_{A,1h}(n)$], for the 24 h of a given day, then we may average these values to arrive at the average $L_{eq,24h}$ for that day

$$L_{eq,24h} = 10 \log_{10} \left\{ \frac{1}{4 \times 10^{-10}} \frac{1}{24} \sum_{n=1}^{24} 10^{0.1L_{A,1h}(n)} \right\} \quad (13.15)$$

L_{eq} measurements are therefore useful to determine average noise exposures over a given period of time.

In order to obtain live readings of sound pressure levels, for example, with a hand held sound level meter in the field, exponential-time-weightings have been developed. Various exponential-time-weightings have been standardized including slow, fast, and impulse. Often a slow exponential time-weighting is called a slow detector for short. Slow and fast exponential-time-weighted sound pressure levels, $L_{A,tw}$, are given by

$$L_{A,tw} = 10 \log_{10} \left\{ \frac{1}{4 \times 10^{-10}} \frac{1}{\tau} \int_{T_1}^{T_2} p_A^2(\xi) e^{-(t-\xi)/\tau} d\xi \right\} \quad (13.16)$$

where $\tau = 1$ s for slow, and $\tau = 0.125$ s for fast weightings. Thus, a slow detector has a longer “memory” than does a fast detector since it gives a higher weighting to previous sound pressures than a fast detector does. The impulse detector is similar in nature to slow and fast detectors, but it has a different time constant depending on whether the sound pressures are rising ($\tau = 0.035$) or falling ($\tau = 1.5$). Thus, the impulse detector responds well to impulsive changes in sound pressure and holds that level fairly constant on the meter so that it can be noted by the operator.

Figure 13.9a displays pressure versus time for an example sound recording that has had a time domain A-weighting filter applied to it. Figure 13.9b shows the instantaneous sound pressure levels using a slow detector and a fast detector. Note the differences in the trailing slopes of the slow and fast detectors when the sound pressures drop suddenly, indicating the respective memories of each type of detector. Figure 13.9c shows a running L_{eq} , and L_{eq} measurements at intervals of 0.125 and 1 s. Recall that, in general, an L_{eq} measurement represents a summation without any time weighting applied.

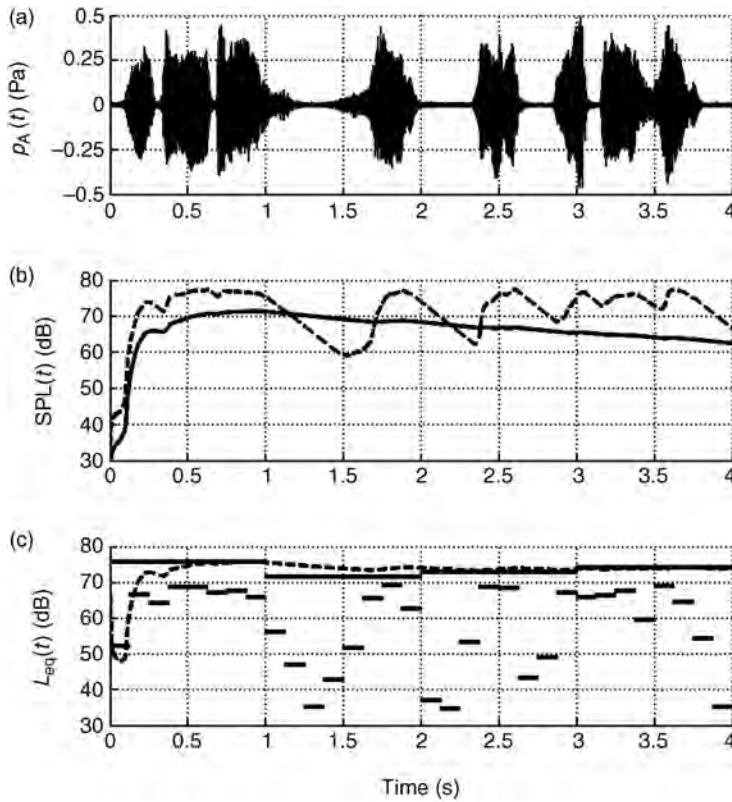


FIGURE 13.9 (a) A sample microphone recording displayed in A-weighted sound pressure versus time. (b) Instantaneous sound pressure levels with a slow (solid line) and a fast detector (dashed line) for the waveform in (a). (c) Equivalent continuous sound level measurements, L_{eq} , for the waveform in (a) (dashed line represents the running L_{eq} , the solid lines of long and short duration are L_{eq} measurements over the time intervals of 1 s and 1/8 s respectively).

13.5 MEASUREMENT OF SOUND ISOLATION

There are numerous applications where it is desired to measure the isolation that is achieved by placing an acoustic treatment between a source and a receiver location. Such applications include noise barriers, enclosures, partitions or walls, and mufflers among others. While the concepts presented here are applicable to all of these applications, the development will focus specifically on sound isolation in a room or cabin and transmission between two such adjacent spaces. Several metrics have been developed to quantify the effectiveness of such treatments.

13.5.1 Transmission Loss

Conceptually, the ideal metric to quantify the effectiveness of an acoustic treatment is often the *transmission loss*. The transmission loss is defined as

$$TL = 10 \log_{10} \left(\frac{W_i}{W_t} \right) \quad (13.17)$$

where W_i is the total acoustic power incident on the treatment of interest and W_t is the total acoustic power transmitted through the treatment. In practice, this metric can often be difficult to measure, as it requires one to be able to measure the incident power with the treatment in place. A common measurement error encountered occurs when the user measures the sound power radiated by the source without the treatment in place, then puts the treatment in place and measures the sound power radiated through the treatment, and calls the difference between the two measurements the transmission loss. This approach would be correct if the sound power radiated by the source does not change with a changing acoustic load. However, it is often the case that inserting the acoustic treatment also alters the acoustic load on the source and as a result alters the acoustic power radiated from the source. Thus, for an accurate measurement of transmission loss, the incident power must be measured with the acoustic treatment in place.

13.5.2 Insertion Loss

As a result of this difficulty, alternative metrics have also been developed. One of those is the *insertion loss*. The insertion loss is defined as

$$IL = 20 \log_{10} \left(\frac{p_w}{p_{w/O}} \right) \quad (13.18)$$

where $p_{w/O}$ is the pressure at a field point without the treatment in place, and p_w is the pressure at the same field point with the treatment in place.

13.5.3 Noise Reduction

A third metric used when the application of interest is transmission of noise from one reverberant room to a second reverberant room is the *noise reduction*, defined as

$$NR = 10 \log_{10} \left(\frac{I_1}{I_2} \right) = L_{p1} - L_{p2} \quad (13.19)$$

Here, I_1 and I_2 are the acoustic intensities in the source and receiving rooms, respectively, and L_{p1} and L_{p2} are the sound pressure levels in the source and receiving rooms. For Equation (13.19) to be accurate, the sound field in the two rooms must be diffuse—a condition that is approximated in a reverberation chamber. For a more accurate measurement, a spatially averaged sound pressure response should be obtained using an array of microphones (as described in ISO 3741, for example) to reduce the effects of nonuniform distribution of sound energy.

For the case of two coupled reverberant rooms, the noise reduction in Equation (13.19) can be related to the transmission loss in Equation (13.17). The sound power incident on the partition between the rooms can be expressed as $W_i = I_1 S$, where S is the area of the partition. The power transmitted (and subsequently absorbed) in the receiving room can be expressed as $W_t = I_2 A$, where A is the total sound absorption of the receiving room. This leads to

$$TL = NR + 10 \log_{10} \left(\frac{S}{A} \right) \quad (13.20)$$

13.5.4 Sound Isolation of Partitions

A single leaf partition is a partition where both faces of the partition are connected together in a manner such that the cross-section moves as a rigid body. Solid panels would be the simplest example of such a partition. At low frequencies, the transmission loss associated with a single leaf partition placed between two acoustic spaces is governed by the mass per unit area of the partition. The mass law predicts the attenuation of a rigid single leaf partition as

$$TL = 20 \log_{10}(f \rho_S) - 47 \text{ dB} \quad (13.21)$$

where f is the frequency (Hz), and ρ_S is the surface density of the solid partition (kg/m^2). The mass law predicts that the transmission loss increases by 6 dB with a doubling of frequency, or with a doubling of the surface density. This leads to the general rule that low frequencies are harder to isolate than higher frequencies, and if further isolation is needed at low frequencies, it often requires the use of higher surface densities to achieve the desired objective.

In practice, single leaf partitions also exhibit structural wave effects. This causes the transmission loss to deviate from the mass law, due to the interaction of the stiffness, mass, and damping properties of the partition. The critical frequency corresponds to the frequency where the phase speed of bending waves in the partition (which is frequency dependent) matches the phase speed in the surrounding acoustic fluid. Above this critical frequency, the coincidence effect is observed, where waves incident from a particular angle pass through the partition with little attenuation, due to a matching of the phase speeds. This results in a coincidence dip that is observed as reduced transmission loss, typically from frequencies slightly below the critical frequency up to frequencies an octave or more above the critical frequency.

Double leaf partitions are constructed using two single leaf partitions with an acoustic cavity between them. The acoustic cavity may also be filled with an absorptive material. Double leaf partitions have the ability to provide additional transmission loss. These designs offer the potential of increasing the transmission loss at most frequencies. However, they do introduce additional resonance effects, including the coincidence frequencies of both partitions and the mass-air-mass resonances that are established with the partitions providing mass effects and the cavity providing stiffness effects. At these resonances, the transmission loss is typically reduced to levels around that corresponding to the mass law.

13.5.5 Sound Transmission Class

The sound transmission class (STC) is a single number rating system that attempts to characterize the sound isolation associated with an acoustic space. It represents the most common single number metric used in North America for this purpose, and is generally used to rate the sound insulation properties of windows, doors, and partitions. The STC is obtained using the transmission loss values associated with the 1/3-octave bands from 125 to 4000 Hz. The STC is obtained as follows. The measured 1/3-octave band values are compared with a reference contour that is obtained from three straight lines. The first line covers the range from 125 to 400 Hz, and increases by 15 dB over this range. The second line covers the mid-frequency range from 400 to 1250 Hz and increases by 5 dB over this range from the value at 400 Hz. The third line for frequencies above 1250 Hz is

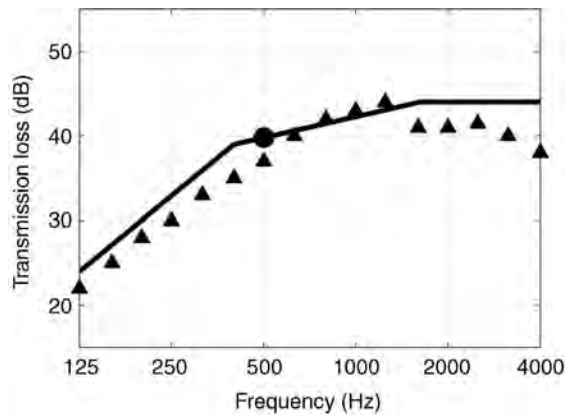


FIGURE 13.10 An example of an STC measurement. Here, the STC = 40.

a horizontal line matching the value of the second line at 1250 Hz. The following guidelines are used to match this reference contour to the measured 1/3-octave band data.

1. No individual transmission loss value can lay more than 8 dB below the reference contour after it is fitted to the data.
2. The total deficiency (obtained as the sum of all deviations below the reference contour) cannot exceed 32 dB. The STC reference contour is fitted to the data to be as high as possible while still meeting these two criteria. The STC value then corresponds to the transmission loss of the STC curve at 500 Hz. An example for determining STC is provided in Figure 13.10, where the STC of this acoustical treatment is 40.

13.6 ROOM ACOUSTICS MEASUREMENTS

13.6.1 Reverberation

One of the first subjective characteristics that a person notices when entering any given room is whether the room is acoustically *live* or *dead*. The degree to which a room is live or dead greatly determines the ability of two people to hold a successfully understood conversation. If a room is too live, it may be difficult to understand speech except at short distances, whereas in a very dead room speech can generally be understood irrespective of distance. This perceptual characteristic is entirely dependent on the amount of reflections off of the walls in the room and the strength of those reflections, which we term room reverberation. For a given sized room, the degree to which the room is live or dead depends on the absorptive qualities of the materials placed on the surfaces of the walls, ceiling, and floor.

Reverberation is quantified through the measurement of an impulse response. An impulsive source, such as a starter pistol or a balloon pop, is used to excite the reverberation in the room and the response of the room due to this impulse may be measured at any selected microphone location. This impulse response may, in general, depend on both the

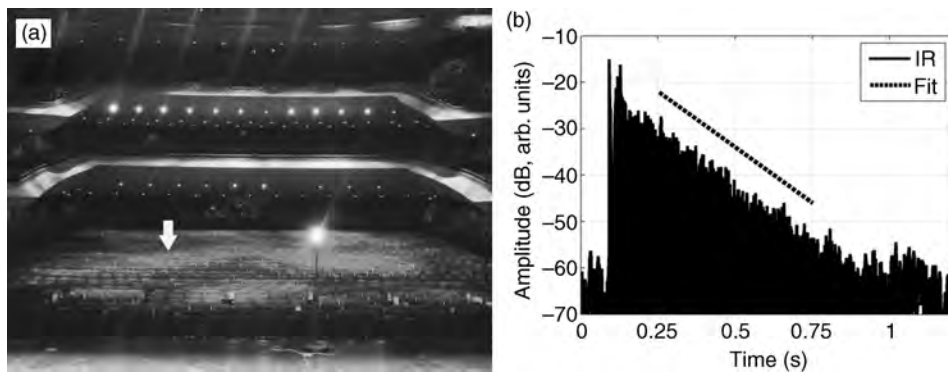


FIGURE 13.11 (a) Photograph of the Eisenhower Auditorium on the campus of The Pennsylvania State University. (b) Impulse response measurement with the source on stage and the microphone located at the seat location indicated by the arrow. The dotted line indicates the average slope of the impulse response's decay.

locations of the source and the microphone. In a classroom, for example, one may place the impulsive source at the location where the instructor would teach from and then place the microphone at a given seat location.

Reverberation in a room is assumed to decay at a constant rate of dB over time. The reverberation time, RT, is defined as the time it takes for the sound pressure level in a room to decay 60 dB. Figure 13.11a shows a voltage versus time impulse response as measured by a microphone in the Eisenhower Auditorium on the campus of The Pennsylvania State University (2500 seat capacity). The impulsive source was located on the stage while the microphone was placed at an audience seat location. Figure 13.11b shows the magnitude of the impulse response versus time on a dB scale. The initial sound arrival occurs 91 ms into the recording. Note that from approximately the 200 to the 900 ms mark that the sound is decaying at a constant rate. After the 900 ms mark, the sound from the impulse is now into the noise floor (we know it is the noise floor because it matches the level before the initial sound arrival and because the sound level flattens out after that time). The initial arrivals of sound commonly do not follow the steady decay rate. To obtain the RT, we must select a portion of the impulse response with a constant decay. If we conservatively choose the time frame from 250 to 750 ms, we can avoid the initial sound arrivals and any influence of the noise floor. The dashed line in Figure 13.11b displays a line that matches the slope of the decay rate. Over this 500-ms interval, the sound decays 24 dB. Thus, the RT is 1.25 s, since it will take that long for the impulsive sound to decay 60 dB.

The RT can be calculated (in metric units) for a given room if the room's volume, V , total surface area, S , and average absorption coefficient, $\bar{\alpha}$, of the room's surfaces are known:

$$RT = 0.161 \frac{V}{-S \ln(1 - \bar{\alpha})} \quad (13.22)$$

This formula is called the Norris-Eyring reverberation time (see Pierce, 1989, pp. 263–265). The quantity $\bar{\alpha}$ may be determined through an area weighted average of the

TABLE 13.8 List of Absorption Coefficients for Common Room Surface Materials

Material	Absorption Coefficients					
	125	250	500	1000	2000	4000
Acoustic tile (suspended, 2 cm thick)	0.76	0.93	0.83	0.99	0.99	0.94
Brick (unpainted)	0.03	0.03	0.03	0.04	0.05	0.07
Concrete (coarse)	0.36	0.44	0.31	0.29	0.39	0.25
Concrete (painted)	0.10	0.05	0.06	0.07	0.09	0.08
Drapery (heavyweight, half area)	0.14	0.35	0.55	0.72	0.70	0.65
Drapery (lightweight, flat)	0.03	0.04	0.11	0.17	0.24	0.35
Glass (typical window)	0.35	0.25	0.18	0.12	0.07	0.04
Heavy carpet (on concrete)	0.02	0.06	0.14	0.37	0.66	0.65
Linoleum (on concrete)	0.02	0.03	0.03	0.03	0.03	0.02
Plaster (gypsum on brick)	0.01	0.02	0.02	0.03	0.04	0.05
Plaster (gypsum on concrete)	0.12	0.09	0.07	0.05	0.05	0.04
Plywood paneling (1 cm thick)	0.28	0.22	0.17	0.09	0.10	0.11

absorption coefficients of the room surfaces:

$$\bar{\alpha} = \sum_i \frac{S_i \alpha_i}{S} \quad (13.23)$$

where, for the i th surface, S_i is the surface area and α_i is the absorption coefficient. Table 13.8 gives a listing of common room surface materials and their corresponding absorption coefficients. Table 13.9 gives a listing of absorption data for seating in terms of αS in square feet per person or per seat. The absorption coefficients listed in these tables are broken down in terms of how absorbent a particular surface is, on average, over a given octave frequency band. These data represent collected and averaged values from the tables listed by Egan (Table 13.8 adapted from Egan, 1972, pp. 32–34) and also Beranek (Table 13.9 adapted from Beranek, 1972, pp. 300–301) and originally come from

TABLE 13.9 List of Absorption Coefficients for Various Seating Types and People (Data for Seating in Terms of αS in Square Feet per Person or per Seat)

Material	Absorption					
	125	250	500	1000	2000	4000
<i>Empty seats</i>						
Upholstered back (leather)	2.0	2.5	3.0	3.0	3.0	2.5
Theater (heavily upholstered)	3.5	3.5	3.5	3.5	3.5	3.5
Wooden chairs	0.1	0.15	0.2	0.35	0.5	0.6
<i>People in seats (add to above values)</i>						
Upholstered back (leather)	0.7	0.6	0.5	1.3	1.6	2.0
Theater (heavily upholstered)	0.7	0.6	0.6	1.0	1.0	1.0
Wooden chairs	4.0	7.5	11.0	13.0	13.5	11.0
<i>Other values (not added to above values)</i>						
People standing	2.0	3.5	4.7	4.5	5.0	4.0
High school students (in chairs)	2.2	3.0	3.3	4.0	4.4	4.5
Elementary school students (in chairs)	1.8	2.3	2.8	3.2	3.5	4.0

TABLE 13.10 List of Optimal Reverberation Times for Various Room Types

Room Type	Room Size (m ³)			
	30	300	3,000	30,000
Office—speech	0.4	0.6	—	—
Classroom—speech	0.6	0.9	1.0	—
Workroom—speech	0.8	1.2	1.5	—
Rehearsal room—speech	0.8	0.9	1.0	—
Studio—music	0.4	0.6	1.0	—
Chamber music	—	1.0	1.2	—
Classical music	—	—	1.5	1.5
Modern music	—	—	1.5	1.5
Opera	—	—	1.4	1.7
Organ music	—	1.3	1.8	2.2
Romantic music	—	—	2.1	2.1
Room in home—speech	0.5	0.8	—	—
Room in home—music	0.7	1.2	—	—

Source: Data from Strong and Plitnik (2007), p. 197.

many sources. Beranek asserts that numerous inconsistencies in the data given by various authors make these values listed only approximate values.

The optimal RT for a given room depends on the primary intended use for that room. Perceptual studies were conducted to determine the optimal RT for various types of rooms. When the room is primarily used for speech, such as for a classroom, lower reverberation times are desired so that speech may be understood clearly above any reflections and echoes. On the contrary, when a room is used for musical performances, longer reverberation times are desired to give a sense of liveliness of the concert hall. Table 13.10 gives a listing of optimal reverberation times for various types of rooms. Thus if one measures an average RT for a given type of room, one may then use Equations (13.22) and (13.23) to determine how they may change the room surfaces to increase or decrease the RT as needed.

13.6.2 Room Resonances

The Fourier transform of a room's impulse response yields its frequency response. The frequency response of a room provides information about modal resonance frequencies of that room. For smaller rooms, the modes can cause undesirable coloration of the acoustic response at low frequencies (as is experienced when one sings in a typical shower). Thus, the frequency response may be used to identify problematic resonance frequencies. However, in order to reduce the strength of a given modal frequency, one must also know the spatial distribution of that mode.

The resonance frequencies of a rectangular room, of dimensions L_X by L_Y by L_Z , are given by

$$f_{l,m,n} = \frac{c}{2} \sqrt{\left(\frac{l}{L_X}\right)^2 + \left(\frac{m}{L_Y}\right)^2 + \left(\frac{n}{L_Z}\right)^2} \quad (13.24)$$

where l , m , and n are integers (0, 1, 2, 3, . . .). There are three types of room modes: axial, tangential, and oblique. Axial modes have amplitude dependence only along one of

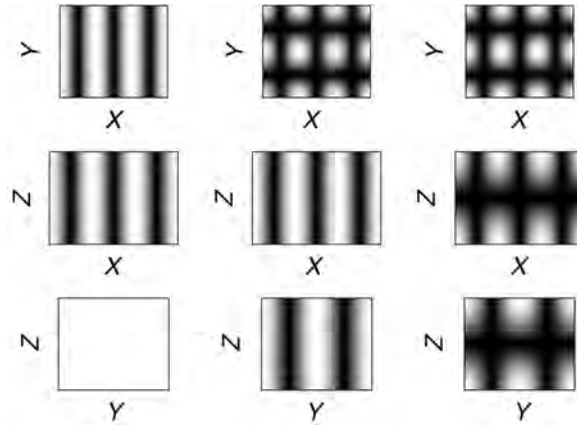


FIGURE 13.12 Illustrations of the spatial variations in axial, tangential, and oblique room modes. Axial mode ($l=3, m=0, n=0$) is shown in the first column, tangential mode ($l=3, m=2, n=0$) is shown in the second column, and oblique mode ($l=3, m=2, n=1$) is shown in the third column.

the axes and they occur when two of the integers are equal to zero. Tangential modes have amplitude dependence in only two dimensions and they occur when one of the integers is equal to zero. Finally, oblique modes have amplitude dependence in all three dimensions and they occur when all integers are non zero. Figure 13.12 gives an example of the spatial distribution of the sound pressure for each of these three types of modes. In the figure, note the lack of Y and Z dependence in the axial mode, and the lack of Z dependence in the tangential mode. If one wishes to characterize a certain modal frequency in a room, both the microphone and the source used in this measurement must be placed away from nodal lines/planes (locations in the room where a mode does not respond).

Each resonance frequency has a finite sized bandwidth associated with it, which depends on the absorption in the room. Increasing absorption decreases the room mode's amplitude and makes the modal frequency response occur over a broader range of frequencies for that mode (so that it is not so apparent). For larger sized rooms, modal frequencies occur within a lower frequency range than for a smaller sized room. Above a frequency called the Schroeder frequency, f_s , modal frequencies overlap to the degree that no one mode dominates the acoustic response of the room. In general, above f_s one need not worry about undesirable coloration of the acoustic response due to a given room resonance. The f_s , valid for most rooms (except for very elongated rooms), is defined in terms of the RT and the volume of the room:

$$f_s = \sqrt{\frac{c^3}{4\ln(10)}} \sqrt{\frac{RT}{V}} = 2090 \sqrt{\frac{RT}{V}} \quad (13.25)$$

where c is the speed of sound (343 m/s).

13.6.3 Critical Distance

Above f_s a room is considered to be spatially uniform in its acoustic response, meaning that any reverberation in a room will be uniformly distributed. However, when two people

are standing close enough to one another, successful communication can still take place. It turns out that there exists a distance at which the direct sound energy from a source equals the reverberant energy (whose energy is created by that source). This distance, for omnidirectional sources (sources without a preferred direction), is called the critical distance, r_C , and is defined as

$$r_C = \sqrt{\frac{\bar{\alpha}S}{16\pi(1 - \bar{\alpha})}} \quad (13.26)$$

This distance increases for directional sources (see Pierce, 1989, pp. 267–270).

13.6.4 Echoes

An echo is a delayed repetition of sound that results in a degradation of speech communication. It has been determined, through perceptual studies, that a reflection of sound which arrives 50 ms later than the direct sound arrival, or any other prominent reflection arrival, is perceived as an echo. This later arrival of sound must also be of sufficient amplitude above the usual constant decay of reverberant energy to be perceived as an echo. Further, this late reflection may only be perceived as an echo if there are no other prominent reflections that precede it by 50 ms. Thus, one way to mask a troublesome echo is to provide other prominent reflections within the 50 ms preceding the troublesome echo. Another way to reduce the echo is to locate the room surface providing the offending echo and treat it with highly sound absorbent materials to reduce the amount of sound energy that is reflected.

13.6.5 Noise Criteria

In the absence of distinct tonal noise, one may define noise criteria, NC, rating for a given room. The NC, given in dB, for a room gives a measure of the perceived background noise since it depends on an overall, A-weighted, sound pressure level measurement, SPL_A :

$$NC = 1.24(SPL_A - 13) \quad (13.27)$$

This SPL_A measurement is made with only background noise present. Table 13.11 gives acceptable NC values for various types of rooms. In some cases, such as a library or a

TABLE 13.11 List of Acceptable Noise Criteria Ratings for Various Room Types

Room Type	Noise Criteria Rating (dB)
Studio	15–20
Concert hall	15–20
Theater	20–25
Auditorium	25–30
Bedroom	25–30
Living room	30–35
Business offices	30–35
Restaurant	35–45

Source: Data from Strong and Plitnik (2007), p. 197)

restaurant, spectrally uniform background noise may be desirable to mask the speech of others. An expansive treatment of additional background noise metrics is given by Blazier (1998).

13.7 COMMUNITY AND ENVIRONMENTAL NOISE

Measurement and analysis of the impact of noise on individuals, communities, and the environment are major subfields within acoustics. These topics are inherently fraught with debate, because of the subjectivity of human perception and in the challenge of defining acceptable limits. This section summarizes some of the measurements and calculations commonly carried out in community and environmental noise assessment.

13.7.1 Representations of Community Noise Data

There are numerous ways of representing community noise. Because of the time scales involved, most involve level-based measurements and the tracking of events of significance, rather than detailed time waveform recordings. One possibility is to display the sound pressure level as a function of time, in addition to a calculation of the equivalent sound level (see Equation (13.14)) over the sampling period. The running level measurements can be more helpful in identifying individual noise events than the equivalent level. As an example, the running A-weighted sound pressure level (L_A) and the $L_{A,eq}$ as a function of time is displayed in Figure 13.13. The data, recorded using a Larson Davis 824 Sound Level Meter, include the emptying of several trash dumpsters during the early morning at an apartment complex in Provo, UT. Before the arrival of the garbage truck, major noise sources (>60 dBA) were due to intermittent traffic from the nearby street. After the arrival of the truck (around 8 min into the data shown), the most significant noise events were due to the dumpsters being shaken by the hydraulic arms on the truck before being noisily set back down.

Other representations of community noise are statistical in nature. Using the same garbage truck example, a cumulative distribution function can be calculated. As shown in Figure 13.14, this function shows the percentage of time that the recorded noise levels exceed a given value. From the cumulative distribution function, X-percentile-exceeded noise levels (L_X) can be calculated. In Figure 13.14, lines representing the L_{90} , L_{50} , and L_{10} of 54.4, 60.4, and 75.3 dBA are shown.

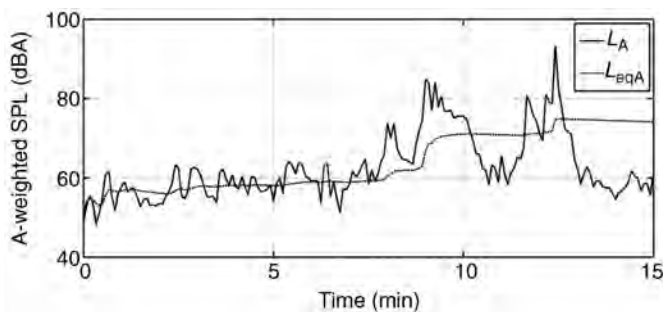


FIGURE 13.13 A-weighted sound pressure level and L_{eq} as a function of time before, during, and after the garbage truck arrival.

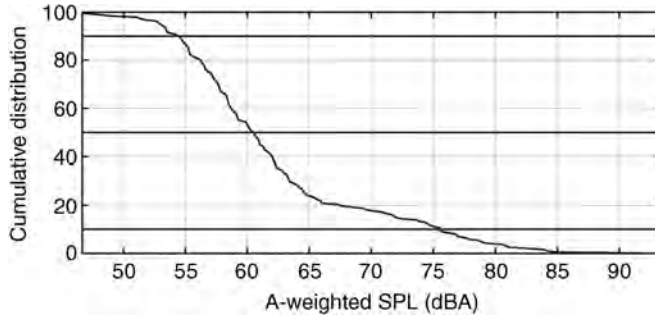


FIGURE 13.14 Cumulative distribution of the time series in Figure 13.13, showing what fraction of time the sound level exceeds a given sound pressure level.

There are various single-number metrics that are used to describe community noise. These metrics have been the result of attempts to correlate subjective response with objective, albeit empirical, measures. Some of the commonly used metrics are as follows:

- Equivalent continuous sound level (L_{eq}). Although the L_{eq} was defined previously in Equation (13.14) and (13.15), further comments regarding common averaging times are merited. These include hourly levels, day levels (7 A.M.–10 P.M.), evening levels (7–10 P.M.), and night levels (10 P.M.–7 A.M.).
- Day–Night Level (DNL or L_{dn}). The L_{eq} obtained for a 24-h period after a 10-dBA penalty is added to the night levels (10 P.M.–7 A.M.). For individual L_{eq} calculations carried out over 1-h intervals (L_{1h}), L_{dn} may be expressed as

$$L_{dn} = 10 \log_{10} \left\{ \frac{1}{24} \left[\sum_{i=0100}^{0700} 10^{0.1(L_{1h}(i)+10)} + \sum_{i=0800}^{2200} 10^{0.1L_{1h}(i)} + \sum_{i=2300}^{2400} 10^{0.1(L_{1h}(i)+10)} \right] \right\} \quad (13.28)$$

The DNL is important in, for example, land planning around airports and in predicting community annoyance due to transportation noise. The predicted community response in terms of percent of people highly annoyed (see Fidell et al., 1991) as a function of DNL is shown in Figure 13.15.

- *Community Noise Equivalent Level (CNEL)*. CNEL is calculated using the L_{eq} obtained for a 24-h period after 5 dBA is added to the evening levels (7–10 P.M.) and 10 dBA is added to the night levels. It can be calculated similar to L_{dn} , with the appropriate penalty given during the evening (between 2000 and 2200 h).
- *Effective Perceived Noise Level (EPNL)*. This metric was designed for characterizing aircraft noise impact and is used by the Federal Aviation Administration (see FAR Part 36 Sec. A.36) in the certification of commercial aircraft. The metric accounts for (a) the nonuniform response of the human ear as a function of frequency (i.e., the perceived noise level), (b) the additional annoyance due to significant tonal components of the spectrum (the tone-corrected perceived noise level), and (c) the change in perceived noisiness due to the duration of the flyover event. Too involved to be repeated here, calculation procedures for EPNL may be found in FAR Part 36 Sec. A.36.4 or Raney and Cawthorn (1998).

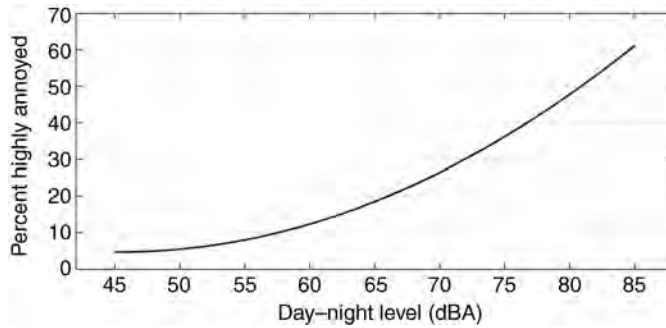


FIGURE 13.15 Percentage of people highly annoyed for a given DNL.

13.7.2 Noise Surveys

In order to document existing community and environmental noise challenges and assist in, for example, land usage planning, properly conducted noise surveys are important. Adequacy of the noise survey parameters will change according to the specific issue and other legal or regulatory requirements. However, careful consideration of number of measurement locations, survey time and duration, appropriate measures, and other measurement parameters is important. In residential surveys, measurements are often made a prescribed distance from a property line. The survey may need to be made during different times of day or seasons and under different meteorological conditions, some of which favor sound propagation to the receiver. A log of individual noise events, such as aircraft flyovers, can form an important part of the survey.

Whatever the specifics of the measurements, an environmental survey report should include not only the data but also information essential to their interpretation. This includes (see Bies and Hansen, 2009, p. 185)

- Reference to the appropriate regulatory document(s)
- Measurement dates, times, and measurement locations
- Local meteorology (wind, temperature, humidity, precipitation) during the survey
- Noise source descriptors
- Instrumentation (type and orientation of meter or microphone and acquisition system, presence of windscreen, etc.)
- Types of noise data recorded
- Impact of extraneous noise sources

13.8 SOUND INTENSITY MEASUREMENTS

13.8.1 Sound Intensity via the “ p - p ” Principle

Although sound pressure measurements are immensely useful in that they yield the sound field magnitude at the measurement location, they may not fully describe the sound

energy transmission for many types of sound fields. The acoustic intensity can be a very important measure in that it is potentially a three-dimensional vector measurement of the net sound energy flux.

Acoustic intensity was defined previously in Equation (13.3) as the sound power per unit area. The instantaneous intensity along direction x , with units of W/m^2 may be written as

$$I_{i,x}(t) = p(t)u_x(t) \quad (13.29)$$

where $p(t)$ and $u_x(t)$ are the time dependent pressure and particle velocity in the x direction. Of more practical measurement interest is the time-averaged intensity, which can be written as

$$I_x(t) = \frac{1}{T} \int_0^T p(t)u_x(t)dt \quad (13.30)$$

The measurement of $p(t)$ is obtained in straightforward fashion with a microphone; however, direct particle velocity measurements are more difficult. Particle velocity measurements have been obtained using ultrasonic transducers or hotwire techniques, and a commercial “ p - u ” probe exists (see Jacobsen and de Bree, 2005). However, the more common technique of obtaining $u_x(t)$ involves processing the pressure signals from closely spaced, well-matched microphones, that is, the “ p - p ” principle (see Fahy, 1995). This is represented schematically in Figure 13.16. The pressure at the center of the two microphones spaced Δx apart is obtained by averaging the pressure signals as

$$p(t) \approx \frac{p_1(t) + p_2(t)}{2} \quad (13.31)$$

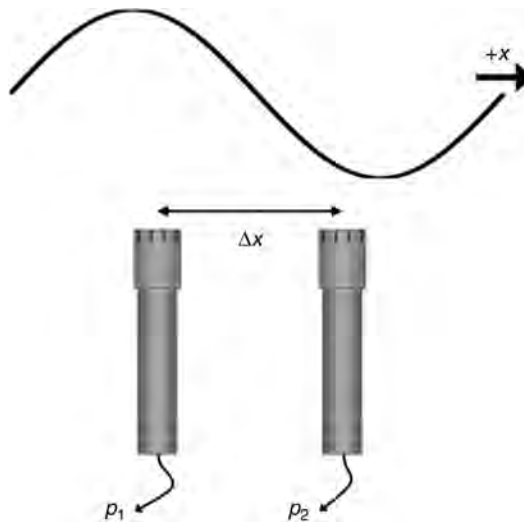


FIGURE 13.16 Layout of two microphones used to obtain particle velocity and pressure estimates needed for calculating acoustic intensity.

and the particle velocity is estimated through Euler's equation

$$\frac{\partial p}{\partial x} = -\rho_0 \frac{\partial u_x}{\partial t} \quad (13.32)$$

which results in

$$u_x(t) \approx \frac{1}{\rho \Delta x} \int_{-\infty}^t [p_2(\tau) - p_1(\tau)] d\tau \quad (13.33)$$

For the case of a time-stationary signal, Fahy (1995) shows that the mean intensity in the x direction may be written as

$$I_x = -\frac{1}{\rho \Delta x T} \int_0^T p_1(t) \int_{-\infty}^t p_2(\tau) d\tau dt \quad (13.34)$$

In other words, the average intensity in a time-stationary field is obtained from averaging the product of one microphone's signal and the integrated signal from a second, closely spaced microphone.

Given the prevalence of real-time spectral analyzers, it is helpful to define intensity in terms of the Fourier transforms of the pressure signals. In the frequency domain, the complex intensity is defined as

$$I_C(f) = I(f) + jJ(f) \quad (13.35)$$

In Equation (13.35), $I(f)$ is the mean active intensity, and represents the in-phase components of the pressure and velocity signals, whereas $J(f)$ is the amplitude of the "reactive" intensity and represents the in-quadrature portions of the signals (see Fahy, 1995). The time average of the reactive intensity is zero; for additional information on its calculation and significance (see Fahy 1995; Jacobsen 1991; Mann et al., 1987). For the two microphones in Figure 13.16, the active intensity spectrum [the spectral form of Equation (13.34)] may be written as

$$I_x(f) = \frac{1}{2\pi f \rho \Delta x} \text{Im}\{G_{p_2 p_1}(f)\} \quad (13.36)$$

where $G_{p_2 p_1}(f)$ is the single-sided time-averaged cross spectral density between the two microphones. In the case of a multidimensional intensity probe, calculation of the intensity vector components is geometry-specific, but involves weighting cross-spectral components between various pairs. A method for deriving the calculation process that results in minimized least-square error has been proposed by Pascal and Li (2008).

13.8.2 Intensity Probes

There are several one and multidimensional intensity probes available. These come in various configurations. One- and three-dimensional sound intensity probes are shown in Figure 13.17. Alternate three-dimensional probes are displayed in Figure 13.18, which use four microphones located at the vertices of a regular tetrahedron. The probe at the left



FIGURE 13.17 One (two microphones) and three-dimensional (three orthogonal pairs of microphones) sound intensity probes, with the microphones in the face-to-face configuration and employing a solid spacer.

is the Ono Sokki Tetra-phone[®] and the probe at the right was developed for rocket noise measurements by Brigham Young University and GRAS Sound and Vibration (see Gee et al., 2009, 2010). Other possible configurations exist, such as those described by Pascal and Li (2008).

13.8.3 Systematic Measurement Errors

It is important to understand the possible sources of error in the intensity measurement technique. Some errors that arise include scattering from probe or holder surfaces and environmental effects, such as wind. These errors are field and measurement specific and are not treated further in this section. However, other systematic errors that arise are inherent in the finite-sum averaging and finite-difference that takes place. When the microphone

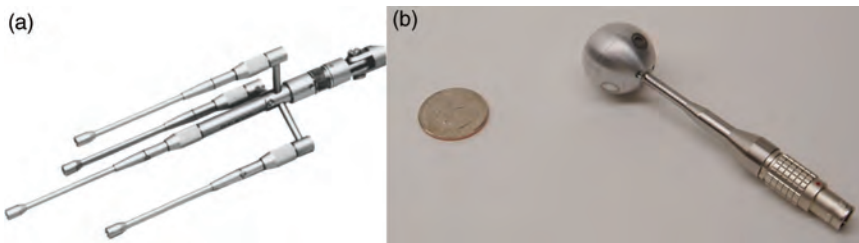


FIGURE 13.18 (a) Ono Sokki Tetra-phone, which uses four microphones in a side-by-side configuration and arranged in a regular tetrahedron. (b) Spherical intensity probe (along with a U.S. quarter for scale), developed by Brigham Young University and GRAS Sound and Vibration for rocket noise measurements. The microphones are arranged in a tetrahedral configuration on the surface of the sphere.

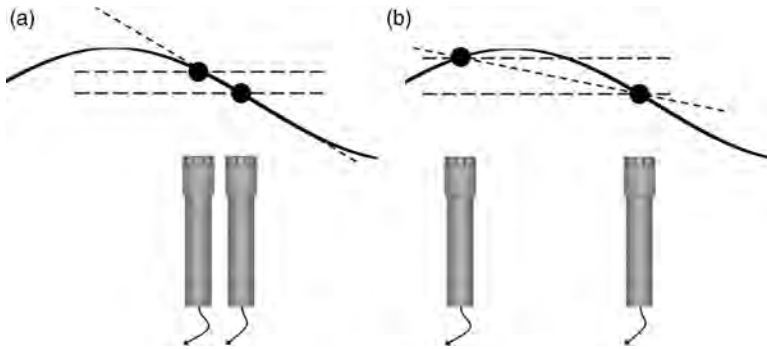


FIGURE 13.19 Diagram showing the process of finite differencing and summation. (a) for small separation distance, the two pressure estimates (black dots) can be used to accurately represent the pressure at the midpoint between the microphones. (b) the separation distance has increased and although finite-differencing results in a near-accurate slope estimate, the pressure average obtained will underestimate the true pressure.

separation distance is small relative to a wavelength, the two microphones can be used to adequately represent the average pressure and the pressure gradient between the two microphones. However, as frequency and/or separation distance is increased such that the distance relative to a wavelength increases, errors result. Although there are errors associated with both the finite-differencing and summation, it is the latter that primarily causes the estimate of the intensity to be too low, as is illustrated in Figure 13.19.

The error in intensity magnitude for different sensor spacings (Δx) is shown in Figure 13.20. The magnitude errors grow as a function of frequency and microphone

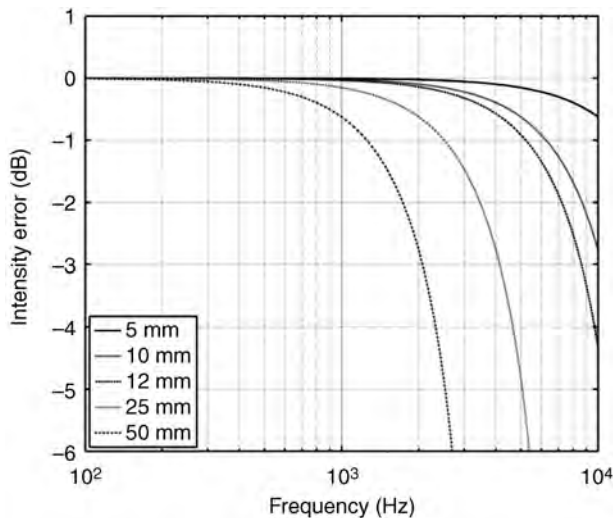


FIGURE 13.20 Error in intensity magnitude for ideal intensity probes with different microphone spacings. This underestimation of acoustic intensity at high frequencies is directly related to an underestimation in the acoustic pressure, based on the separation of the microphones.

spacing. Jacobsen et al. (1998) have found that, in practice, probes that employ solid spacers with the microphones in the face-to-face configuration have a resonance phenomenon that partially counteracts the finite summation bias error and extends the probe's useable bandwidth. It is also worth noting that the phase errors introduced by the “ p - p ” measurement principle are zero for the case of the plane progressive wave and planar interference fields, provided that the sensor separation distance is less than a half wavelength, at which point spatial aliasing begins to occur.

13.8.4 Transducer Mismatch

Another source of error is transducer or instrumentation system mismatch. Any amplitude or phase mismatch will result in errors in the estimation of p or u , which translate into intensity amplitude and/or direction errors. The intensity magnitude error for a progressive plane wave propagating along the probe axis with channel phase mismatch of 0.1° is shown in Figure 13.21. The figure shows the increased error as frequency and sensor spacing decreases, as the acoustic phase difference between channels becomes progressively smaller and the phase error becomes relatively more important. In general, phase errors between measurement channels are frequency dependent and often grow as the low-frequency limit of the microphone is approached. For this reason, commercial intensity probes are sold with solid spacers of various lengths to provide accurate measurements at different frequency ranges.

In a field where both microphones are exposed to the same pressure, amplitude and phase, the time-averaged intensity should be zero. This can be seen from the fact that the imaginary part of the cross-spectrum between the two signals in Equation (13.36) is zero. However, for a real probe with slightly mismatched microphones measuring a uniform sound field in a small cavity, the phase error between the microphones results in a residual or “false” intensity measurement. For a given pressure amplitude, p , phase error, φ_{err} , and

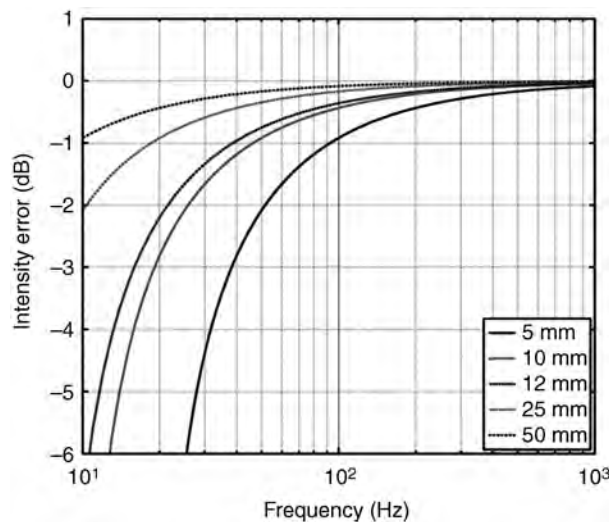


FIGURE 13.21 Intensity magnitude error for plane wave propagation along the probe axis with 0.1° phase error between channels and various sensor spacings.

separation distance relative to wavelength, $k\Delta x$, the residual intensity error, I_{res} , can be expressed as

$$\frac{I_{\text{res}}}{p^2/(\rho c)} = \frac{\varphi_{\text{err}}}{k\Delta x} \quad (13.37)$$

The pressure-residual intensity index, which is a convenient way of expressing the phase error, is calculated as

$$\delta_{pI_{\text{res}}} = 10 \log_{10} \left[\frac{p^2/(\rho c)}{I_{\text{res}}} \right] \quad (13.38)$$

For example, a 0.1° phase error at 100 Hz for a 25-mm microphone spacing results in a pressure-residual intensity index of 14.2 dB. For the 50 Hz one-third-octave band, ANSI standard S1.9-1996 calls for a minimum pressure-residual intensity index of 7.0 dB for a Class 2 measurement system and 13.0 dB for a Class 1 system. Other specifications for probe and calibration and classification are provided within the standard.

13.8.5 Sound Intensity Applications

As discussed below, one application of sound intensity measurements is the measurement of sound power. Other applications include sound source characterization and localization and examination of energy flow. Near-field intensity vectors for the radiated noise from a small solid rocket motor at three different frequencies are shown in Figure 13.22 (see Gee et al., 2010). The intensity vectors reveal the directionality and magnitude of the sources. Another example is the graphical display of an intensity measurement system developed for NASA Glenn Research Center (see Dix et al., 2003). The system involves an intensity probe attached to a computer-controlled motion system that can be rapidly moved around

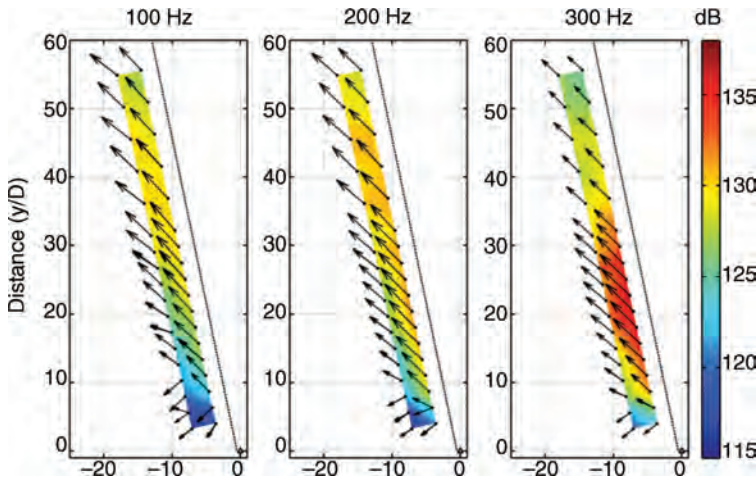


FIGURE 13.22 Measurement of vector sound intensity made near a rocket plume at three different frequencies. The vectors reveal the directionality and magnitude of the energy flow.

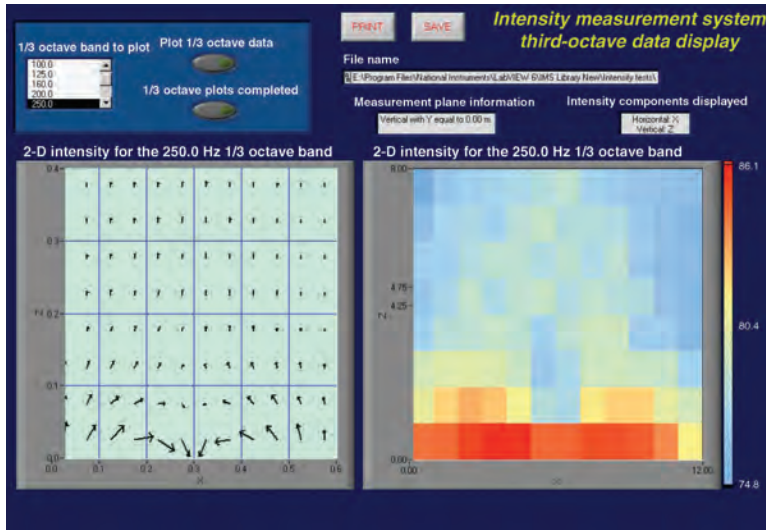


FIGURE 13.23 Graphical display of an automated sound intensity measurement system developed for NASA Glenn Research Center.

to produce results such as those shown in Figure 13.23. The sources in this case comprise three loudspeakers receiving the same source signal but with the middle speaker out of phase with the other two. The intensity vectors clearly reveal the source/sink nature of the near-field energy flow, which the intensity magnitude alone (shown to the right) does not.

13.9 SOUND POWER MEASUREMENTS

When considering acoustic radiation from a source, sound power measurements are often preferable to sound pressure measurements. The sound power corresponds to the amount of acoustic energy radiated per unit time, and can also be expressed as the integral of the acoustic intensity over the area through which the acoustic energy is radiated. It is expressed in units of watts (W). Unlike sound pressure, the measurement of sound power is independent of the distance from the source where the measurement is made. Furthermore, for many sources the sound power radiated is nearly independent of the physical location of the source. Thus, if the sound power is measured outdoors or inside a room the value will be nearly identical; however, if the sound pressure radiated from that same source is measured in those two environments, significantly different results would be expected.

The magnitude of the sound power varies tremendously for normally encountered sources. For this reason, it is convenient to express the sound power on a logarithmic scale and is referred to as the sound power level. The sound power level, L_W , is given by

$$L_W = 10 \log_{10} \left(\frac{W}{W_{\text{ref}}} \right) \quad (13.39)$$

where W is the sound power of the source in watts and W_{ref} is the reference sound power. For airborne applications, the reference sound power is usually taken to be 10^{-12} W.

Most sound power level measurements are obtained using pressure measurements, based on the relationship between the radiated power and the radiated pressure. For these measurements, the measurement can be categorized as a free field or a reverberant field condition.

13.9.1 Measurement of the Sound Power Level in a Free Field

The configuration corresponding to measurement of the sound power level in a free field is generally obtained in an anechoic chamber. An anechoic chamber has the property that there is no appreciable reflection from any of the walls of the chamber, generally obtained by constructing the walls with sound absorbing wedges. For a free field, the sound power can be obtained by integrating the sound intensity over an imaginary spherical surface of area $S = 4\pi r^2$ enclosing the source. If the measurement is in the far field, the intensity is directly related to the acoustic pressure. Thus

$$\begin{aligned} W &= \int_S \vec{I} \cdot d\vec{S} \\ &= \int_S \frac{p^2}{\rho c^2} dS \\ &= \frac{p_{\text{ave}}^2}{\rho c^2} S \end{aligned} \quad (13.40)$$

Using standard reference values (20 μPa for pressure and 10^{-12} W for power) allows the sound power level to be obtained from the *average* sound pressure level as

$$L_W = \bar{L}_p + 10 \log_{10} \left(\frac{S}{S_0} \right) \quad (13.41)$$

where S_0 is the reference area of 1 m^2 . The average sound pressure level is obtained using

$$\bar{L}_p = 10 \log_{10} \left(\frac{1}{N} \sum_{i=1}^N 10^{0.1 L_{pi}} \right) \quad (13.42)$$

Here, L_{pi} is the sound pressure level from the i th measurement location (in dB), usually measured in third-octave or octave bands and typically A-weighted, and N is the total number of measurement locations. The ISO standards outline the number of points that should be included in the measurement. There are also a number of standard measurement surfaces that are used.

For free-field measurements, the measurement surface is generally taken as an imaginary sphere. As the first step, an imaginary box (called the reference box) is constructed that is as small as possible but completely encloses the source of interest. The characteristic distance, d_0 , associated with this box is taken as the distance from the projection of the center of this box on the floor to one of the upper corners of the box. The radius of the measurement sphere, R , should then be at least $2d_0$. If the size of the source is large, relative to the available measurement space, then the measurements can be made on the surface of a large imaginary box surface. The measurement locations are chosen to represent equal area on the sphere or box.

For many sources, it is more practical to measure the sound power in a hemianechoic room or outdoors. The procedure for this measurement parallels that described above for

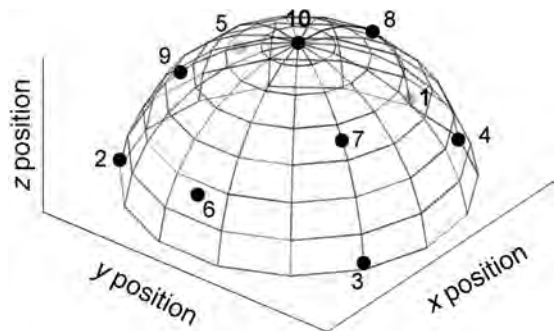


FIGURE 13.24 Basic sound power measurement positions that should be used for a hemispherical measurement surface. The two measurement locations on the backside of the hemisphere are displayed here in gray.

anechoic conditions, except that the measurement surface corresponds to either a hemisphere or a five-sided box. Several standards are available to guide the measurement of sound power for these conditions. Precision (or grade 1) methods yield reproducibility with a standard deviation of less than 1 dB, while engineering (or grade 2) methods and survey (or grade 3) methods yield reproducibility with a standard deviation of less than 2 dB or 3 dB, respectively.

Figure 13.24 shows the basic measurement positions that should be used for a hemispherical measurement surface. The engineering method described in ISO 3744 normally

TABLE 13.12 Microphone Locations According to ISO 3745

No.	x/R	y/R	z/R
1	-0.99	0	0.15
2	0.50	-0.86	0.15
3	0.50	0.86	0.15
4	-0.45	0.77	0.45
5	-0.45	-0.77	0.45
6	0.89	0	0.45
7	0.33	0.57	0.75
8	-0.66	0	0.75
9	0.33	-0.57	0.75
10	0	0	1.0
11	0.99	0	-0.15
12	-0.50	0.86	-0.15
13	-0.50	-0.86	-0.15
14	0.45	-0.77	-0.45
15	0.45	0.77	-0.45
16	-0.89	0	-0.75
17	-0.33	-0.57	-0.75
18	0.66	0	-0.75
19	-0.33	0.57	-0.75
20	0	0	-1.0

Positions 1–10 are used for a hemispherical measurement surface of radius R , and positions 1–20 are used for a spherical measurement surface (the z -axis is the vertical axis).

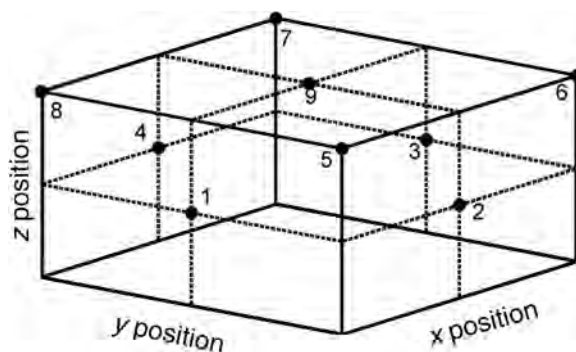


FIGURE 13.25 The nine measurement positions used for sound power measurements if a box surface is implemented in a hemianechoic environment.

uses the 10 positions shown in this figure, while the survey method described in ISO 3746 only uses positions 4, 5, 6, and 10. The one restriction is that if the difference between any two sound pressure level measurements (in dB) exceeds the number of measurement locations, the number of locations must be increased. If a fully anechoic condition exists, so that a spherical measurement surface is used, the measurements in the lower hemisphere correspond to those on the upper hemisphere. The coordinates for all 20-measurement positions are given in Table 13.12. Figure 13.25 shows the nine-measurement positions used if a box surface is implemented in a hemianechoic environment.

13.9.2 The Reverberation Method

The sound power of an unknown source can also be determined in a reverberant environment. The concept behind this measurement is that the acoustic power radiated in this environment must equal the power absorbed by the surfaces of the test room. If the test room is a reverberation chamber, the sound is considered to be diffuse. For such a field, the sound power can be obtained directly from the sound pressure level in the room if the absorption of the room is known, which can be determined through measuring the reverberation time in the room. Since the room is never truly diffuse, the sound pressure level is obtained through the use of multiple pressure measurements, as outlined in the ISO 3741 standard.

Since the room response is never truly diffuse, the measurement of the sound power is generally more problematic with a narrowband sound source than a broadband source using the reverberation method. Thus, the reverberation method is more accurate for use with broadband noise, with the sound power being measured in octave bands or 1/3-octave bands.

In its simplest form, the sound power is obtained as

$$L_W = \bar{L}_p + 10 \log_{10} \left(\frac{A}{A_0} \right) - 6 \text{ dB} \quad (13.43)$$

where \bar{L}_p is the mean sound pressure level measured in the room (in dB), A is the total absorption in the room (m^2) obtained from measuring the reverberation time, and A_0 is the reference absorption, taken as 1 m^2 .

In practice, several corrections can lead to a more accurate measurement. Since the energy in the room is not truly diffuse, the standard requires the sound pressure

measurements to be at least 1.0 m away from the boundary. However, the energy density is greatest near the walls, so it has been found that the term $10 \log_{10}[1 + Sc/(8Vf)]$ can correct for this underestimate. To correct for air absorption in the room, the term $4.34A/S$ is added. Corrections for meteorological conditions can also be added, resulting in the sound power being obtained as

$$L_W = \bar{L}_p + 10 \log_{10} \left(\frac{A}{A_0} \right) + 10 \log_{10} \left(1 + \frac{Sc}{8Vf} \right) + 4.34 \frac{A}{S} - 6 + C_1 + C_2 \quad (13.44)$$

Here, S is the total surface area of the measurement room (m^2), V is the volume of the measurement room (m^3), f is the center frequency of the frequency band of interest (Hz), c is the speed of sound given by $c = 20.05 \sqrt{273 + \theta} \text{ m/s}$ (where θ is the temperature [$^{\circ}\text{C}$]), $C_1 = -10 \log_{10} \left(\frac{B \sqrt{T_0}}{B_0 \sqrt{T}} \right)$ (where B is the static atmospheric pressure, $B_0 = 1.013 \times 10^5 \text{ Pa}$, T is the temperature in K, and $T_0 = 313.15 \text{ K}$), and $C_2 = -15 \log_{10} \left(\frac{BT_1}{B_0 T} \right)$ (where $T_1 = 296.15 \text{ K}$).

13.9.3 Comparison Method

The sound power can also be determined through the use of the comparison method. This method is simple and accurate, but does require the use of a known reference power source. The reference power source, radiating reference sound power $L_{W,\text{ref}}$, is placed in the measurement volume and the sound pressure level, $L_{p,\text{ref}}$, associated with the reference power source is measured. The unknown source is then operated and the sound pressure level, L_p , is measured. The unknown sound power is then obtained as

$$L_W = L_{W,\text{ref}} + L_p - L_{p,\text{ref}} \quad (13.45)$$

For accurate results, the test room should be hard-walled with no major absorbing surfaces. The volume must exceed 40 times the volume of the reference box, a minimum of three microphones must be used (which remain in the same locations for both measurements), and the sound power should be processed in octave bands or 1/3-octave bands. If the unknown source is not movable, the reference sound source is located on top of the unknown source.

13.10 SOUND EXPOSURE MEASUREMENTS

Two acoustical quantities that are important to determining hearing risk and appropriate conservation procedures are the sound exposure and noise dosage. Additional information on these measurements can be found in Kardous (2007) and Marsh and Richings (1998).

13.10.1 Sound Exposure Level

The sound exposure, with W -type weighting filter applied, is defined for an event occurring over time T as

$$E_W = \int_0^T p_W^2(t) dt \quad (13.46)$$

where $p_w(t)$ is the instantaneous weighted pressure. The most common weighting applied in this situation is A-weighting. Without a weighting filter applied, this quantity is related to the total acoustic energy incident on a receiver. From the sound exposure in Equation (13.46), the W-weighted sound exposure level (SEL) is calculated as

$$\text{SEL}_W = 10 \log_{10} \left(\frac{E_W}{E_{\text{ref}}} \right) \quad (13.47)$$

where E_{ref} is the reference sound exposure of $400 \mu\text{Pa}^2/\text{s}$.

Relationships between sound exposure, the SEL and the L_{eq} defined previously in Equation (13.14) can be stated. First, the SEL can be calculated from the L_{eq} of the same weighting as

$$\text{SEL} = L_{\text{eq}} + 10 \log_{10}(T) \quad (13.48)$$

Alternatively

$$\text{SEL} = p_{\text{ref}}^2 \times T \times 10^{0.1L_{\text{eq}}} \quad (13.49)$$

Although the international standard unit for sound exposure is the squared-pascal second, a second unit, the squared pascal hour (Pa^2/h), is also used. This is a convenient unit for measuring workplace noise exposure, particularly since an A-weighted sound exposure of $1 \text{ Pa}^2/\text{h}$ is nearly equal to a constant level of 85 dB over an 8-h working day.

13.10.2 Noise Dosage

The noise dose is a metric used to quantify noise exposure in the workplace. According to ANSI S1.25, the noise dose for an 8-h workday and exchange rate, Q , is defined as

$$D_Q = \frac{100}{8} \int_0^T 10^{[(L_A(t) - L_C)/q]} dt \quad (13.50)$$

where t is measured in hours and T is the measurement duration. In Equation (13.50), $L_A(t)$ is the slow or fast A-weighted sound level in decibels, provided that the level is greater than a selected threshold level, or is $-\infty$ if below the threshold. The level, L_C , is the criterion sound level that defines 100% noise dosage and $q = Q/\log(2)$, where Q is the exchange rate of 3, 4, or 5 dB. The exchange rate is the change in sound level that corresponds to a doubling or halving of the sound exposure duration while maintaining a constant percentage of criterion exposure. For Occupational Safety and Health Administration (OSHA) regulations, the threshold level is 80 dBA, the criterion level is 90 dBA, and the exchange rate is 5 dB. The exchange rate for the National Institute of Occupational Safety and Health (NIOSH) and most international standards is given as 3 dB.

An alternate method of calculating noise dosage is defined for a workday with noise exposure at different levels for discrete intervals. In this case

$$D = 100 \times \sum_i \frac{t_i}{T_i} \quad (13.51)$$

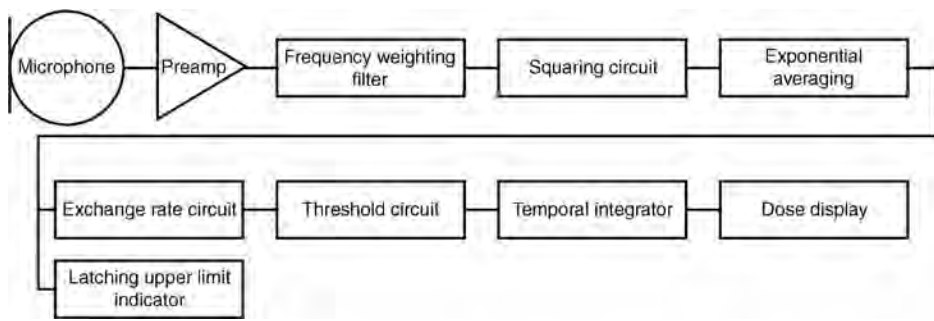


FIGURE 13.26 Elements of a noise dosimeter.

where t_i is the exposure time in hours and T_i is the allowable (100% dosage) duration for exposure at that level. Values for T_i can be calculated for an 8-h workday as

$$T = \frac{8}{2^{[(L-L_c)/Q]}} \quad (13.52)$$

(see Appendix A of OSHA 29CFR 1910.95 regulation).

13.10.3 Equipment

Personal sound exposure meters and noise dosimeters are used to assess noise exposure in the work place. The essential elements of the dosimeter are shown in Figure 13.26 and a typical class-2 dosimeter is displayed in Figure 13.27. ANSI standard S1.25 indicates that a dosimeter should allow for A or C weighting, and fast and slow exponential averaging.



FIGURE 13.27 A Class 2 Larson Davis noise dosimeter with microphone inserted in a calibrator (Photo courtesy of Chuck Kardous, NIOSH).



FIGURE 13.28 Noise dosimeter being worn (Photo courtesy of Churck Kardous, NIOSH).

It should also permit a criterion level of 95, 85, 84, 80, or variable dB, a threshold level of 90, 80, or variable dB, and an adjustable exchange rate of 3, 4, or 5 dB. The standard also specifies that the dosimeter should properly measure impulsive noises up to 140 dB, the maximum allowable peak sound level without hearing protection. NIOSH has shown that many dosimeters have issues with these high-level measurements (Kardous and Willson, 2004).

The noise dosimeter may be used to monitor the noise environment in an area or worn as a personal dosimeter. In this case, the ANSI S12.19 standard specifies that the microphone should be located on the middle-top of a worker's shoulder and should be oriented approximately parallel to the shoulder's plane. This is shown in Figure 13.28. The amount that the person affects the noise level measured depends on, for example, the noise spectrum, absorption due to clothing, angle of incidence, and microphone orientation. The increase may be as much as 5 dB, but an increase in level of around 1 dB is typical.

REFERENCES

- ANSI S1.12-1967—R1977. Specifications for laboratory standard microphones, New York, American National Standards Institute.
- ANSI S1.10-1966—R1986 Method for the calibration of microphones, New York, American National Standards Institute.
- Beranek LL. 1993. "Acoustics," (reprinted by the Acoustical Society of America).

- Bies DA, Hansen CH. *Engineering Noise Control: Theory and Practice*. 4th ed. London: Spon Press; 2009.
- Blazier Jr. WE., Noise control criteria for heating, ventilating, and air-conditioning systems. *Handbook of Acoustical Measurements and Noise Control. Chapter 43*. 3rd ed. Harris CM, editor. Melville (NY): Acoustical Society of America; 1998.
- Bobber R. *Underwater Electroacoustic Measurements*. Peninsula Pub; 1990.
- Burnett ED, Nedzeltnitsky V. Free-field reciprocity calibration of microphones. *Journal of Research of the National Bureau of Standards-A* 1987;92(2):129–151.
- Davis H, Silverman SF. *Hearing and Deafness. Chapter 2*. 4th ed. Holt, Rinehart, & Winston; 1978.
- Dix G, Gee KL, Sommerfeldt SD. Design and implementation of an automated intensity scanning system at the Acoustical Testing Lab of NASA Glenn Research Center. In Holger DK, Maling Jr. GC, editors. *Proceedings of Noise-Con 03*. New York: Noise Control Foundation, Poughkeepsie; 2003; paper nc03_202.
- Egan MD. *Concepts in Architectural Acoustics*. McGraw-Hill; 1972.
- Fahy FJ. *Sound Intensity*. 2nd ed. London: E&FN Spon; 1995.
- Fidell S, Barber DS, Schultz TJ. Updating a dosage effect relationship for the prevalence of annoyance due to general transportation noise. *Journal of the Acoustical Society of America* 1991;89:221–233.
- Fletcher H, Munson WA. Loudness, its definition, measurement and calculation. *Journal of the Acoustical Society of America* 1933;5(2):82–108.
- Gee KL, Giraud JH, Blotter JD, Sommerfeldt SD. Energy-based acoustical measurements of rocket noise. *AIAA* 2009;2009–3165, (May issue).
- Gee KL, Giraud JH, Blotter JD, Sommerfeldt SD. Near-field acoustic intensity measurements of a small solid rocket motor. *Journal of the Acoustical Society of America* 2010;128:EL69–EL74.
- Hunt FV. *Origins in Acoustics, The Science of Sound from Antiquity to the Age of Newton*. (reprinted by the Acoustical Society of America). 1992.
- Jacobsen F. A note on instantaneous and time-averaged active and reactive sound intensity. *Journal of Sound and Vibration* 1991;147:489–496.
- Jacobsen F, Cutanda V, Juhl PM. A numerical and experimental investigation of the performance of sound intensity probes at high frequencies. *Journal of the Acoustical Society of America* 1998;103:953–961.
- Jacobsen F, de Bree H-E. A comparison of two different sound intensity measurement principles. *Journal of the Acoustical Society of America* 2005;118:1510–1517.
- Kardous CA. Noise dosimeters. In M. J. Crocker, editor. *Handbook of Noise and Vibration Control. Chapter 39*. Hoboken (NJ): Wiley; 2007.
- Kardous CA, Willson RD. Limitation of integrating impulsive noise when using dosimeters. *Journal of Occupational and Environmental Hygiene* 2004;1:456–462.
- Kinsler LE, Frey AR, Coppens AB, Sanders JV. *Fundamentals of Acoustics*. 4th ed. John Wiley & Sons, Inc.; 2000.
- Lang WM, Nobile MA. In Harris CM, editor. *Handbook of Acoustical Measurements and Noise Control. Chapter 15*. 3rd ed. Melville (NY): Acoustical Society of America; 1998.
- Maclean WR. Absolute measurement of sound without a primary standard. *Journal of the Acoustical Society of America* 1940;12(13.1):140–146.
- Mann III JA, Tichy J, Romano AJ. Instantaneous and time-averaged energy transfer in acoustic fields. *Journal of the Acoustical Society of America* 1987;82:17–30.
- Marsh AH, Richings WV. Measurement of sound exposure and noise dose. In Harris CM, editor. *Handbook of Acoustical Measurements and Noise Control. Chapter 12*. 3rd ed. Melville (NY): Acoustical Society of America; 1998.

- Pascal J, Li J. A systematic method to obtain 3D finite-difference formulations for acoustic intensity and other energy quantities. *Journal of Sound and Vibration* 2008;310:1093–1111.
- Pierce AD. *Acoustics: An Introduction to Its Physical Principles and Applications*. (reprinted by the Acoustical Society of America). 1989.
- Raney JP, Cawthorn JM. Aircraft noise. In Harris CM, editor. *Handbook of Acoustical Measurements and Noise Control*. 3rd ed. Woodbury (NY): Acoustical Society of America; 1998.
- Robinson DW, Dadson RS. Threshold of hearing and equal-loudness relations for pure tones, and the loudness function. *Journal of the Acoustical Society of America* 1957;29(12):1284–1288.
- Strong WJ, Plitnik GR. *Music Speech Audio*. 3rd ed. BYU Academic Publishing; 2007.
- Torr GR, Jarvis DR. A comparison of national standards of sound pressure. *Metrologia* 1989;26:253–256.

14

TEMPERATURE MEASUREMENT

PETER R. N. CHILDS

Summary

- 14.1 Introduction
 - 14.1.1 The measurement process
 - 14.1.2 Calibration
 - 14.2 Selection
 - 14.3 Invasive temperature measurement
 - 14.3.1 Liquid-in-glass thermometers
 - 14.3.2 Manometric thermometry
 - 14.3.3 Bimetallic thermometers
 - 14.3.4 Thermocouples
 - 14.3.5 Resistance temperature devices
 - 14.3.6 Semiconductor devices
 - 14.3.7 Diode thermometers
 - 14.3.8 Noise thermometry
 - 14.3.9 Pyrometric cones
 - 14.4 Semi-invasive methods
 - 14.4.1 Peak temperature-indicating devices
 - 14.4.2 Temperature-sensitive paints
 - 14.4.3 Thermographic phosphors
 - 14.4.4 Thermochromic liquid crystals
 - 14.5 Noninvasive methods
 - 14.5.1 Infrared thermometry
 - 14.5.2 Thermal imaging
 - 14.6 Conclusions
- Nomenclature
- References

SUMMARY

The requirement to measure temperature arises in process control, production, environmental observation, and laboratory research. The range of techniques available for measuring temperature is extensive. Many phenomena are dependent on temperature and this can be exploited in instrumentation. A given application may permit direct contact between a measuring device or system and the medium of interest. Alternatively, remote observation of, for example, infrared radiation or fluorescence may be possible using an optical system. This chapter outlines various forms of invasive, semi-invasive, and non-invasive temperature measurement devices and systems. In addition, this chapter describes the fundamental definitions of temperature and the issues that need to be considered in determining the temperature of a given medium, be it a solid body, surface, liquid, or gas.

14.1 INTRODUCTION

Temperature can be defined qualitatively as a measure of hotness of a body. Temperature is the property that determines whether a system is in thermal equilibrium with other systems. If the temperature of two bodies in thermal contact with each other is same, then there will be no net transfer of thermal energy. Quantitatively, temperature can be defined from the second law of thermodynamics in terms of the rate of change of entropy with energy.

To allow the assignment of numerical values to bodies at different temperatures, some form of temperature scale is necessary. The SI unit of the thermodynamic temperature scale is the kelvin, with symbol K. This is defined in terms of the interval between absolute zero, 0 K, and triple point of pure water, 273.16 K. The kelvin is defined as the fraction $1/273.16$ of the temperature of a system exhibiting the triple point of water. Other temperature units are also in common use including the Celsius, Fahrenheit, and Rankine temperature scales. Conversion equations for these are given in Equations (14.1)–(14.3).

$$t = T - 273.15 \quad (14.1)$$

$$T_{\circ F} = 1.8t + 32 \quad (14.2)$$

$$T_{\circ R} = T_{\circ F} + 459.67 \quad (14.3)$$

where, t is the temperature in degrees Celsius ($^{\circ}\text{C}$), T the absolute temperature (K), $T_{\circ F}$ the temperature in degrees Fahrenheit ($^{\circ}\text{F}$), and $T_{\circ R}$ the temperature in Rankine scale ($^{\circ}\text{R}$).

The thermodynamic temperature scale is defined by means of theoretically perfect heat engines. These are not practically realizable, and the International Temperature Scale of 1990, see Preston-Thomas (1990), denoted by ITS-90, was developed as a practical best approximation using available technologies to the thermodynamic temperature scale. Its range of application extends from 0.65 K up to the highest temperature practically measurable using Planck's law of thermal radiation. The ITS-90 is believed to represent thermodynamic temperature to within ± 2 mK from 2 K to 273 K, ± 3.5 mK at 730 K, and ± 7 mK at 900 K (one standard deviation limits; see Mangum and Furukawa, 1990). The ITS-90 is constructed using a number of overlapping temperature ranges. This leads to some ambiguity, albeit small, in the true value of a temperature across an overlapping region but allows greater flexibility in the use of the scale. The ranges are defined between

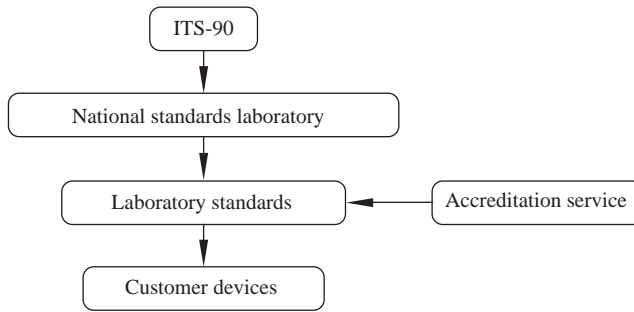


FIGURE 14.1 Traceability between customer devices and the ITS-90.

repeatable conditions using a variety of specified materials at their melting, freezing, and triple points. The ITS-90 was developed under the auspices of the Metric Treaty and the associated consultative committees, BIPM (Bureau International des Poids et Mesures), CIPM (Comité International des Poids et Mesures), and CCT (Comité Consultatif de Thermométrie) and was adopted by the International Committee of Weights and Measures in 1989.

The principal role of the ITS-90 in practical measurement is to allow a means of traceability between the measurement that is actually made and the temperature scale defined by the ITS-90. Thermometers or systems used in defining the ITS-90 are unlikely to be used to measure the temperature of a system of interest. Instead a thermometer can be used that has been calibrated against another device calibrated using guidelines specified in ITS-90. Calibration is the process of relating the output value from a measurement system to a known input value. The schematic given in Figure 14.1 illustrates this process whereby there is a transfer of calibration between practical and specialist devices. The chain between the thermometer in use and the ITS-90 may in practice have more links than those shown in the diagram; each link tending to increase the uncertainty associated with the measurement.

14.1.1 The Measurement Process

A measurement provides information about a property and gives magnitude to that property. Temperature can be measured by observing physical phenomena that is temperature dependent. This may involve inserting a probe containing a transducer into the medium of interest and relating the observed effect on the transducer to temperature. A measurement system may comprise a transducer to convert a temperature-dependent phenomena into a signal that is a function of that temperature, a method to transmit the signal from the transducer, some form of signal processing, a display and method of recording the data. In a given application some of these functions such as data recording may be undertaken by the sensor system itself or by a human operator. A calibration is used to convert the measured quantity into a value of temperature. The significance of temperature can then be considered. The related subject of heat flux measurement is reviewed by Childs et al. (1999).

Whenever a measurement is made it is unlikely that the measured value will equal the true value; the actual or exact value of the measured variable. The difference between the measured value and the true value is the error. There are a number of reasons why a

measured value will not equal the true value and many of these can be inferred by considering the measurement process. The insertion of a transducer, or sometimes even thermal interactions between a remote sensor and an application, will result in disturbance to the temperature distribution in an application. The magnitude of this disturbance will depend on the heat transfer processes involved. Natural instabilities associated with the transducer and signal-measuring devices also contribute to a deviation between the true and measured values. Further deviations between the measured and true values are due to the processing of the signal and data and uncertainties arising from the calibration process.

14.1.2 Calibration

Calibration can be achieved in a number of ways. The sensor to be calibrated can be placed in an environment at a known temperature such as one of the fixed points, for example, the triple point of water. The output from the sensor can be recorded and, within the uncertainties associated with the temperature of the fixed point, taken as an indication of that temperature. The sensor can then be exposed to a different temperature using a different fixed point, for example, the melting point of gallium, and the process repeated. For intermediate values of temperature between the fixed points, the special interpolation equations documented in the ITS-90 can be used. The process described so far is appropriate for producing a sensor with very low uncertainty capability. This, however, is an expensive undertaking and is rarely necessary for the majority of scientific and industrial applications. Instead, a sensor that has been calibrated in this way and is stable is used to calibrate another sensor by comparison. The calibrated and uncalibrated sensors are placed in thermal contact in an enclosure such as a bath of liquid or in a specially constructed heating block or furnace. The medium is heated and the output of both the calibrated and the uncalibrated sensor is recorded. The output from the uncalibrated sensor can then be related to the temperature indicated by the calibrated sensor. In this way a transfer between values of temperatures set up in the ITS-90 and sensors used in practice can be achieved as illustrated schematically in Figure 14.1. It is normally recommended that there should be no more than three or four links between a practical measurement and the ITS-90. The details of the calibration process depend partly on the type of sensor concerned typically involving established procedures for record keeping, inspection and conditioning of the sensor, generic checks, intercomparison either with another sensor or with a medium at a known temperature, analysis of the data, quantification of uncertainties, and completion of records.

Traceability involves ensuring that appropriate procedures have been followed, and in the current industrial setting, this is typically associated with quality standards. Two approaches can be adopted. The first involves following a set of standards and obtaining the appropriate approval and registration certificates for that quality system. The second approach is to accept the principles of the process commended within quality standards and to follow them closely to be able to state that the spirit of the standard has been followed. This latter approach negates the need for registration to a standards system but does require a level of trust.

The difference between a measurement and the true value of temperature is known as the error. The term uncertainty can be used to quantify confidence in the measurement indicated by a sensor. For instance, a sensor may be supplied with a 95% confidence interval uncertainty of $\pm 1.5^\circ\text{C}$. This means that, provided the device does not disturb the temperature distribution in the medium of interest, measurements made will be within 1.5°C of the true value, for 95% of the measurements made. The term uncertainty is used in

preference to accuracy as the latter should generally be reserved for qualitative use. The subject of measurement uncertainty is addressed comprehensively by Figliola and Beasley (2005).

14.2 SELECTION

The selection of a specific sensor or system for temperature measurement can require consideration of a number of aspects including

- uncertainty,
- temperature range,
- thermal disturbance,
- level of contact,
- size of the sensor,
- transient response,
- sensor protection,
- availability,
- cost.

For many applications there is often more than one method that will provide an adequate determination of temperature. A key consideration is application, whether the measurement is of a liquid, gas, solid body, or surface.

The measurement of temperature of a solid material can be achieved in several ways. A transducer can be embedded in the material by forming a hole, inserting the device and, in the case of an electrically based transducer, feeding the wires out through the drill hole. This method can also be used for liquid-in-glass or bimetallic thermometers where the stem of device is inserted to a specific depth bringing the transducer into thermal contact with the region of interest. In the case of a liquid-in-glass thermometer, the device may have to be removed from the application in order to reveal the scale to enable the temperature to be read against the scale. It is important that the transducer is in good thermal contact with the solid body in order to ensure that the transducer takes up the local temperature. Thermal contact can be improved in some cases by filling the hole or pocket with high thermal conductivity oil, or other liquid, which will provide a much better conduction path than if a gas is filling the space between the probe and the hole walls. Errors arising from the effects of unwanted heat conduction along the connecting wires, or the presence of the probe, should be minimized, subject to the requirements of the measurement.

Surface temperatures can be measured by attaching a sensor to the surface, drilling or forming a channel on the underside of the surface and attaching a sensor, by applying a temperature sensitive paint to the surface, or by observing the surface remotely using, say, an infrared thermometer. For approximate assessment of surface temperature, bimetallic thermometers are available with flat bases, spring clips, and permanent magnets. However, if a substantial temperature difference exists between the surface and the immediate surroundings, uncertainties can arise. Thermocouples, platinum resistance thermometers, or thermistors equipped with flat or suitably formed pads can also be considered. To obtain the best uncertainty with applied probes it is usually advantageous if they are small, and perhaps also recessed in shallow grooves. An alternative method of surface

temperature measurement is the noninvasive techniques of infrared radiation or thermal imaging. These techniques may be more expensive to implement. At the other end of the scale, there is frequently a need to obtain a quick and uncomplicated way of checking the temperature of a surface. Temperature indicating paints can be applied very readily to provide an indication of the temperatures that are being encountered. The measurement of the temperature of a surface that is in thermal contact with a gas and also exchanging thermal radiation with its surroundings is subject to a series of processes that could contribute to deviations between the measured temperature and that of the undisturbed object.

The measurement of liquid temperatures usually involves using a transducer that is immersed in the liquid. If a simple visual indication is required a liquid-in-glass thermometer can be considered. Extremely low uncertainty is possible with an appropriate, calibrated instrument. If a larger indicator is required a bimetallic thermometer might be suitable. Bimetallic thermometers are robust and self-contained. However, in many cases, an electrical signal is required for processing, recording, or controlling purposes. Electrically based devices include thermocouples and resistance thermometers. The actual device will be dependent on the temperature range and the precision required. If the temperatures are changing with time, the response characteristics of the sensor and system must be adequate. Thermocouples or resistance temperature detectors will probably be

TABLE 14.1 Approximate Temperature Measurement Range Capabilities for Different Methods

Method	Minimum Temperature (°C)	Maximum Temperature (°C)
Gas thermometer	about -269	700
Liquid-in-glass thermometer	-200	600
Bimetallic thermometer	-73	540
Thermocouple	-270	2300
Electrical resistance device	-260	1064
Thermistors	-100	700
Semiconductor devices	-272	300
Fiber optic probes	-200	2000
Capacitance	-272	-170
Noise	-273	1500
Thermochromic liquid crystals	-40	283
Thermographic phosphors	-250	2000
Heat-sensitive paints	300	1300
Infrared thermometer	-40	2000
Schlieren	0	2000
Shadowgraph	0	2000
Interferometry	0	2000
Line reversal	727	2527
Absorption spectroscopy	20	2500
Emission spectroscopy	20	2700
Rayleigh scattering	20	2500
Raman scattering	20	2227
CARS	20	2000
LIF	0	2700
Acoustic thermography	-26.9	2000

the best choice if a particularly rapid response is needed. Infrared thermometers can be used to measure the surface temperature of molten materials such as metals and glass.

Many of the considerations that influence the measurement of liquid temperatures also apply to measuring the temperature of a body of gas. If the gas is moving then there is potential for heat transfer due to conduction along the sensing probe, connection wires, and supports, which can lead to significant errors, unless accounted for. In addition, if the gas is at a different temperature to the surrounding surfaces then thermal radiation exchanges will occur between the sensor and surfaces, again leading to a deviation between the indicated temperature and the gas temperature unless accounted for. High gas velocities can give rise to dynamic heating with an immersed sensor and due allowance must be made for this effect. For simple, isolated measurements, liquid-in-glass or bimetallic thermometers may well be suitable. If electrical signals are required, the choice again lies between thermocouples, platinum resistance thermometers, thermistors, or transistor instruments. If protection has to be added, the increase in thermal mass and thermal resistance will tend to degrade the response of the sensor more seriously in a gas than in a liquid. For some gas temperature measurements, generally associated with combustion processes, it may be undesirable or impossible to use conventional sensors and cooled devices or a noninvasive technique may be suitable.

An indication of the temperature range of a variety of methods is given in Table 14.1.

14.3 INVASIVE TEMPERATURE MEASUREMENT

One option for measuring temperature is to locate the sensor in the medium of interest or on its surface. The sensor will, due to its presence, alter the distribution of temperature within the medium. This is due to differences in the thermal conductivity and heat capacity of the sensing device and the medium of interest. As a result, the installation of sensor in or on a solid or fluid can be classified as invasive. A wide range of invasive sensors are available including liquid-in-glass thermometers, manometric thermometers, bimetallic thermometers, thermocouples, resistance temperature devices such as PRTs and thermistors, noise thermometers, and pyrometric cones. These are described in Sections 14.3.1–14.3.9.

14.3.1 Liquid-in-Glass Thermometers

Liquid-in-glass thermometers exploit the higher volumetric expansion of liquids with temperature in comparison with that of solids. A liquid-in-glass thermometer typically consists of a reservoir and capillary tube containing a thermometric liquid, supported in a stem with a temperature-indicating scale. On heating, the volume of the liquid increases relative to that of the container and the liquid expands up the capillary tube. Liquid-in-glass thermometers can be used from approximately -196°C to 650°C , although no single instrument is capable of measuring temperature across the whole range because of the limitations of the thermometric liquids. Liquid-in-glass thermometers do not require an external power supply and can be relatively inexpensive. As glass is generally chemically stable, these thermometers can be used in a wide variety of chemical environments. Their disadvantages include fragility and the lack of remote logging capability.

In the case of a solid stem thermometer the bulb reservoir is usually a thin glass container with 0.35–0.45 mm thick walls holding a thermometric liquid such as mercury,

ethanol, pentane, toluene, or xylene. The choice of the liquid depends on the desired temperature range

- mercury, -35°C to 510°C ,
- ethanol, -80°C to 60°C ,
- pentane, -200°C to 30°C ,
- toluene, -80°C to 100°C ,
- xylene, -80°C to 50°C .

The space above the thermometric liquid can be evacuated or filled with an inert gas. As the temperature of the liquid in the bulb rises, the liquid will expand and some of it will be forced up the capillary. The temperature of the bulb is indicated by the position of the top of the meniscus, in the case of a mercury thermometer, against markings engraved on the stem.

Liquid-in-glass thermometers can be calibrated at a number of fixed points and a scale subsequently applied to the stem supporting the capillary tube to indicate the value of temperature. The uncertainty of industrial glass thermometers depends on the device concerned with values ranging from $\pm 0.01^{\circ}\text{C}$ to $\pm 4^{\circ}\text{C}$ (see for example BS 1041 (1985)). The use of liquid-in-glass thermometers is reviewed by Ween (1968), Wise (1976), Nicholas and White (1994), Nicholas (1999), and Childs (2001a,b).

Mercury-in-glass thermometers are increasingly being replaced by resistance-based temperature devices, which have cost and environmental advantages, and by infra-red devices, giving a digital readout and noninvasive measurement, or by thermally sensitive paint devices, which give a visible color-based indication of temperature.

14.3.2 Manometric Thermometry

Methods of temperature measurement based on the measurement of pressure are known as manometric thermometry. There are two principal categories: gas thermometry and vapor pressure thermometry.

Gas thermometry is based on the ideal gas law:

$$pV = n\Re T \quad (14.4)$$

where p is the pressure (N/m^2); V the volume (m^3); n the number of moles of the gas ($n = m/M$ (m = mass, M = molar mass)); \Re the universal gas constant (8.314510 J/mol K (Cohen and Taylor, 1987, 1999), and T the temperature (K).

By assuming values for the quantity of mass and for the gas constant, the temperature is obtained by measuring pressure and/or volume. The basic components of a gas thermometer are enclosures to hold the gas sample of interest, under carefully controlled conditions, and a flow circuit to allow the pressure to be measured. Gas thermometers can be used for applications from a few K to 1000 K. The principal application has been in cryogenics, and the gas enclosure, commonly known as the bulb, is usually located within a cryostat. Figure 14.2 illustrates the principal components of a constant volume gas thermometer. Gas thermometry tends to be a specialist activity and is usually confined to standards laboratories and cryogenic applications.

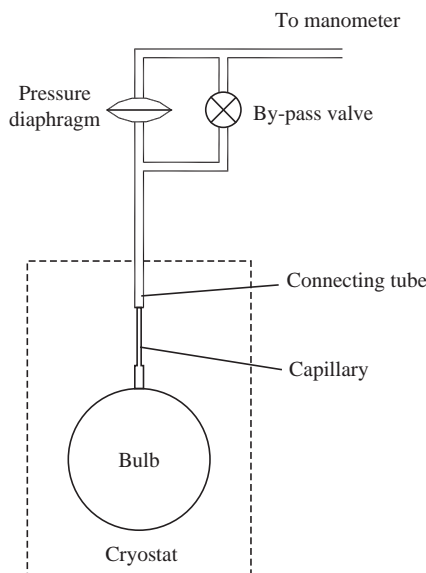


FIGURE 14.2 Principal constant volume gas thermometer components.

Direct use of Equation 14.4 requires knowledge of the gas constant. To reduce uncertainty arising from uncertainty in the gas constant, a number of methods have been devised that eliminate the need for knowing the gas constant, operating on the principle of maintaining either constant pressure, constant volume, and/or constant bulb temperature. Such techniques include

- absolute PV isotherm thermometry,
- constant volume gas thermometry,
- constant pressure gas thermometry,
- constant bulb temperature gas thermometry,
- two bulb gas thermometry.

Uncertainty in the measurement of temperature using gas thermometry is a function of the care taken and the temperature range. Pavese and Steur (1987), for example, report an uncertainty of 0.5 mK for the temperature range 0.5–30 K.

Real gases do not behave exactly according to the ideal gas equation and the nonideal nature of real gases can be modeled using the virial equation

$$p = \left(\frac{\Re T}{V} + \frac{B(T)}{V^2} + \frac{C(T)}{V^3} + \frac{D(T)}{V^4} + \dots \right) \quad (14.5)$$

where $B(T)$, $C(T)$, $D(T)$, etc., are the second, third, fourth virial coefficients, and so on.

The saturation vapor of a pure substance above its liquid phase varies with temperature and is known with low uncertainty for a number of cryogenic liquids such as helium 3 and 4, hydrogen, neon, nitrogen, oxygen, argon, methane, and carbon dioxide. Measurement of the vapor pressure can therefore be used to determine the temperature and this technique provides a method with good sensitivity and requires relatively simple equipment.

Equations quantifying the relationship between pressure and temperature for a variety of cryogenic liquids are given in Pavese (1999), and vapor pressure thermometry is described in detail by Pavese and Molinar (1992).

14.3.3 Bimetallic Thermometers

A bimetallic thermometer consists of two strips of materials, typically two different types of metal that are bonded together. When heated, the assembly will deform due to a mismatch of the coefficient of linear expansion between the two materials. If one end is fixed, then a needle mounted on the other end can be used to indicate the temperature against a calibrated scale. To maximize the bending of the assembly and hence the sensitivity of the device, materials with significantly different coefficients of expansion can be used. Table 14.2 lists some of the materials typically used in bimetallic thermometers.

A bimetallic strip can be coiled in a spiral or helical configuration in order to provide increased sensitivity for a given volume and this form is commonly used in dial thermometers. Dial bimetallic thermometers tend to be reasonable rugged devices and have the advantage that they do not need an independent power supply. A variety of options are typically available for attaching the instrument to an application including magnetic bases and clips.

The uncertainty of typical commercial bimetallic thermometers is 1–2% of the full-scale deflection with an operating range of -70°C to 600°C . The theoretical limit of operation however is from about -270°C to the elastic limit of available materials.

The general equations for defining the curvature of a bimetallic strip are developed in Timoshenko (1953) and Childs (2001a,b). An advanced design device with a sensitivity of $0.0035^{\circ}\text{C}/\text{mm}$ and a repeatability of 0.027°C was demonstrated by Huston (1962). The merits of bimetallic thermometers include that they can be easily read, can be used as an indicator of temperature or as an actuator, are relatively inexpensive and do not require an independent power supply. Their disadvantages include that they are subject to drift, the measurements are usually relatively uncertain in comparison to, say, thermocouples and industrial PRTs and they cannot provide a remote indication of temperature.

TABLE 14.2 Properties of Selected Materials Used in Bimetallic Elements.

Material	Density (kg/m^3)	Young's Modulus (GPa)	Heat Capacity ($\text{J}/\text{kg K}$)	Coefficient of Thermal Expansion ($10^{-6}/\text{K}$)	Thermal Conductivity ($\text{W}/\text{m K}$)
Al	2700	61–71	896	24	237
Brass	8500	110.6	820	19	106
Cu	8954	129.8	383.1	17	386
Cr	7100	279	518	6.5	94
Au	19300	78.5	129	14.1	318
Fe	7870	211.4	444	12.1	80.4
Ni	8906	199.5	446	13.3	90
Ag	10524	82.7	234.0	19.1	419
Sn	7304	49.9	226.5	23.5	64
Ti	4500	120.2	523	8.9	21.9
W	19350	411	134.4	4.5	163
Invar (Fe64/Ni36)	8000	140–150	480	1.7–2.0	13

Source: After Stephenson et al. (1999), Data from Meijer and van Herwaarden (1994) and Goodfellow (2000).

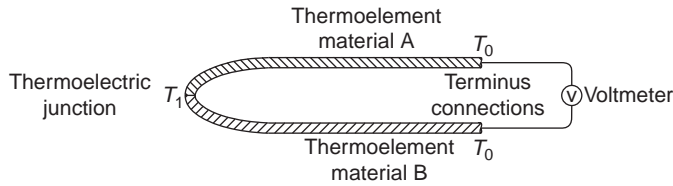


FIGURE 14.3 Simple practical thermocouple circuit. T_1 = the temperature at the thermoelectric junction. T_0 = temperature at the terminus connections (Childs, 2001a,b).

14.3.4 Thermocouples

In its simplest practical form a thermocouple can consist of two dissimilar wires connected together at one end with a voltage measurement device across the free ends, as indicated in Figure 14.3. A net electromotive force (emf) due to the Seebeck effect will be indicated by the voltmeter, which is a function of the temperature difference between the joint and the voltmeter connections. Thermocouples are widely used in industry and research with applications from -272°C to 2000°C . The merits of thermocouples are their relatively low cost, small size, rugged nature, versatility, reasonable stability, reproducibility, reasonable uncertainty, and fast speed of response. Although PRTs are more accurate and stable and thermistors are more sensitive, thermocouples are often a more economical solution than PRTs and their temperature range is greater than thermistors. The main disadvantage of thermocouples is their relatively weak signal, which makes the reading sensitive to electrical noise. For example, the output is about 4.1 mV at 100°C for a type K thermocouple. Other disadvantages include a nonlinear output that requires amplification, and calibrations can vary with contamination of the thermocouple materials, cold working, and temperature gradients.

The fundamental physical phenomenon exploited in thermocouples, discovered by Johann Seebeck, is that heat flowing in a conductor produces a movement of electrons and thus an electromotive force (emf). Seebeck demonstrated that a small current flowed through the circuit shown in Figure 14.4 when the temperature of the two junctions was different (Seebeck, 1823). The emf produced is proportional to the temperature difference

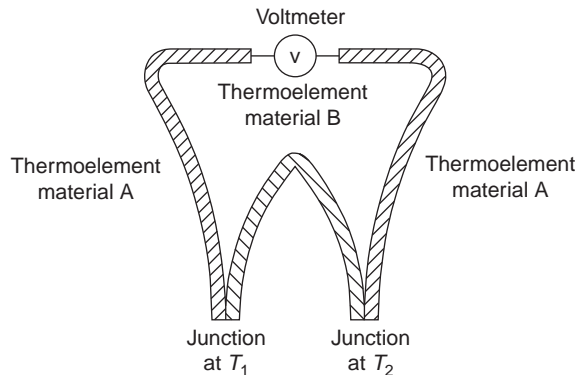


FIGURE 14.4 A current will flow through the above circuit that is proportional to the temperature difference between the two junctions (Childs, 2001a,b).

and is called the Seebeck emf or thermoelectric potential. As well as being a function of the net temperature difference, the magnitude of the emf produced is also a function of the materials used.

14.3.4.1 Thermocouple Analysis Thermocouple circuits can consist of the simple forms illustrated in Figures 14.3 and 14.4 or much more complex systems resulting from considerations of data logging and long distances between the point of interest and emf measurement. In this case, the use of thermocouple wire for the whole length may be prohibitively expensive and alternative materials, known as extension and compensation cable, can be used to convey the emf. In order to ensure correct assessment of the signal produced, the circuit must be analyzed. Analysis of a given thermocouple circuit can be undertaken using fundamental physical relationships (see for example Pollock, 1991; Barnard, 1972; Blatt et al., 1976; Bourassa, 1978; Mott and Jones, 1958). Alternatively, thermocouple behavior can be modeled in most applications using a number of laws or an algebraic technique. The thermocouple laws allow a quick common sense approach to be taken to practical temperature measurement and have been developed by a number of authors (e.g., Roeser, 1940; Doebelin, 1990; Childs, 2001a,b; ESDU, 2007a).

14.3.4.2 The Law of Interior Temperatures The thermal emf of a thermocouple with the junction at T_1 and terminus connections at T_0 is unaffected by temperature elsewhere in the circuit provided the properties of the two thermoelements used are homogenous. This requirement means that the physical properties of the wires must be constant with length. If the wire is stretched or strained in a region or the chemical makeup of the wire varies along its length, this will affect the thermoelectric output and could invalidate the law. The law of interior temperatures is illustrated in Figure 14.5. Provided the wire is uniform and homogenous on both sides of the hot spot, then no net emf is generated by the hot spot. The thermocouple will respond only to the temperature difference between the thermoelectric junction and the terminus connections. This result is particularly significant as it means that the emf from a thermocouple is not dependent on intermediate temperatures along a thermoelement.

14.3.4.3 The Law of Inserted Materials If a third homogenous material C is inserted into either thermoelement A or B, then, as long as the two new thermoelectric junctions are at the same temperature, the net emf of the circuit is unchanged irrespective of the

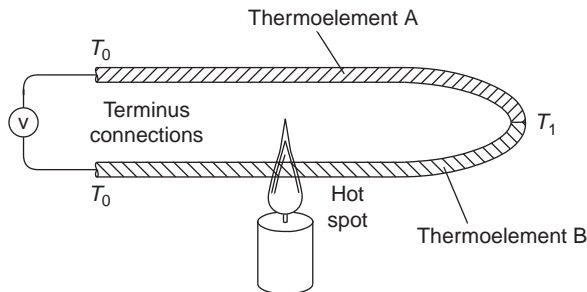


FIGURE 14.5 Illustration of the law of interior temperatures. The thermocouple is unaffected by hot spots along the thermoelement, and the reading is only a function of T_1 and T_0 (Childs, 2001a,b).

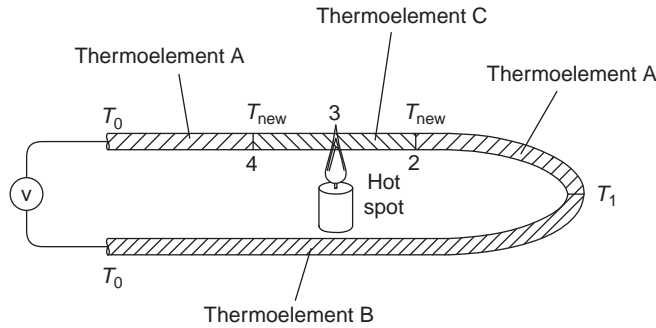


FIGURE 14.6 Illustration of the law of inserted materials. The thermocouple is unaffected by the presence of the inserted material and any local hot spot.

temperature in material C away from the thermoelectric junctions. This law is illustrated in Figure 14.6 where a third material is inserted into the thermoelement A and then heated locally. Provided the thermoelectric junctions between C and A are both at the same temperature, the net emf of the thermocouple is unaffected by the presence of the inserted material and any local hot spot as the emf excursion between 2 and 3 is cancelled by that between 3 and 4.

14.3.4.4 The Law of Intermediate Materials If material C is inserted between A and B, the temperature of C at any point away from the junctions A–C and B–C is not significant. This law is illustrated in Figure 14.7. Here an intermediate thermoelectric material is inserted between the two thermoelements. As there is no thermal gradient across the new thermoelectric junctions then the presence of the inserted material does not contribute to the net emf produced by the thermocouple. This law is of great practical significance as it allows us to model the implications of manufacturing techniques used to form thermoelectric junctions. Provided there is no thermal gradient across the thermoelectric junction it does not matter if the thermoelements are joined by a third material such as solder or if local thermoelectric properties are changed at the junction by, for instance, welding.

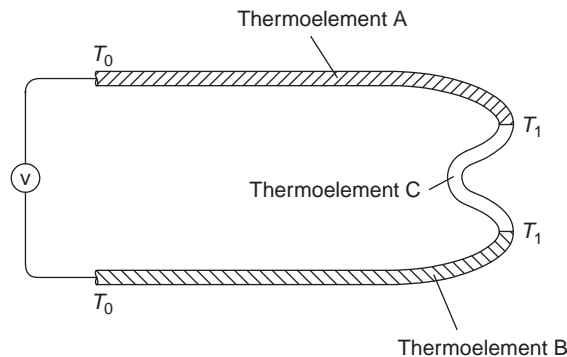


FIGURE 14.7 Illustration of the law of intermediate materials. As there is no thermal gradient across the new thermoelectric junctions the presence of the inserted material does not contribute to the net emf produced by the thermocouple.

14.3.4.5 Modeling Fundamental Thermoelectric Phenomena An assessment of the fundamental thermoelectric phenomena can be made. These are the Seebeck, Peltier, and Thompson effects but in practical thermocouple circuits, the contribution of the Peltier and Thompson effects is insignificant. The Seebeck effect is the generation of emf in a conductor whenever there is heat transfer and is a consequence of electron movements when heat transfer occurs. An emf will be generated in a material whenever there is a temperature difference in the material and its magnitude will be a function of the temperature difference and the type of material. The Seebeck coefficient is a measure of how the electrons are coupled to the metal lattice and grain structure. It is sensitive to changes in the chemical and physical structure of the solid and will alter if the material is contaminated, oxidized, strained, or heat treated. Data for thermocouple materials normally uses the relative Seebeck coefficient, stating the Seebeck coefficient relative to platinum.

The Seebeck coefficient is defined by Equation (14.6) and is a transport property of all electrically conducting materials.

$$S(T) = \lim_{\Delta T \rightarrow 0} \frac{\Delta E}{\Delta T} = \frac{dE}{dT} \quad (14.6)$$

where S is the Seebeck coefficient ($\mu\text{V/K}$); ΔT the temperature difference across a segment of the conductor (K); and ΔE the absolute Seebeck emf (μV).

The Seebeck coefficient cannot be measured directly. Instead it must be determined by measuring the Thomson coefficient and using a thermodynamic relationship. The Seebeck coefficient varies with temperature so it must be mathematically defined by the gradient dE/dT at a specific temperature.

Rearranging Equation (14.6) allows modeling of the net thermoelectric emf generated by a practical thermocouple circuit:

$$dE = S(T)dT \quad (14.7)$$

If the circuit comprises two materials, A and B, then

$$E = \int_{T_1}^{T_2} S_A dT - \int_{T_1}^{T_2} S_B dT = \int_{T_1}^{T_2} (S_A - S_B) dT \quad (14.8)$$

where E is the thermoelectric emf (μV); S_A the Seebeck coefficient for material A ($\mu\text{V/K}$) and S_B the Seebeck coefficient for material B ($\mu\text{V/K}$).

In Equation (14.8), a difference in the Seebeck coefficients in the two thermoelements appears. It is this difference that is of practical interest in thermocouple thermometry, and it is called the relative Seebeck coefficient. The relative Seebeck coefficient is normally determined with respect to a reference material such as platinum.

$$S_{\text{APt}} = S_A - S_{\text{Pt}} \quad (14.9)$$

$$S_{\text{BPt}} = S_B - S_{\text{Pt}} \quad (14.10)$$

$$S_{\text{AB}} = S_A - S_B = S_{\text{APt}} - S_{\text{BPt}} \quad (14.11)$$

TABLE 14.3 Values of the Seebeck Coefficient for Various Materials.

Material	20°C	1000°C
Chromel	22.2	9.4
Fe	13.3	−7
Nicrosil	11.8	8.8
Au	2.0	4
Cu	1.9	7
Ag	1.7	
W	1.3	20.3
Pt	4.7	21.4
Nisil	−14.8	−29.8
Alumel	−18.2	−29.6
Ni	−19.5	−35.4
Constantan	−38.3	−65.6

Source: After Bentley (1998).

where S_{APt} is the Seebeck coefficient for material A relative to platinum ($\mu\text{V/K}$); and S_{BPt} the Seebeck coefficient for material B relative to platinum ($\mu\text{V/K}$).

Values for the Seebeck coefficient relative to platinum are listed in Table 14.3 for a variety of materials.

Substituting for S_A and S_B in Equation (14.11) gives

$$E = \int_{T_1}^{T_2} S_{AB} dT \quad (14.12)$$

and if the relative Seebeck coefficient can be taken as constant over the temperature range then

$$E = S_{AB}(T_2 - T_1) \quad (14.13)$$

Equation (14.13) is particularly useful as it allows the analysis of a wide range of circuits. This can be achieved using loop analysis. In loop analysis the contribution of emf due to the thermal gradient across each element is summed together to produce the total emf that would be indicated by a voltage-measuring device.

The influence of connection leads shown in Figure 14.8 on the output can be analyzed using Equation 14.13. If identical conductors are used, then by using loop analysis

$$E = S_C(T_1 - T_0) + S_A(T_2 - T_1) + S_B(T_1 - T_2) + S_C(T_0 - T_1) \quad (14.14)$$

This simplifies to

$$E = S_{AB}(T_2 - T_1) \quad (14.15)$$

The output emf is therefore dependent only on the temperature difference between the thermoelement junction and the terminus junctions. In other words, the temperature

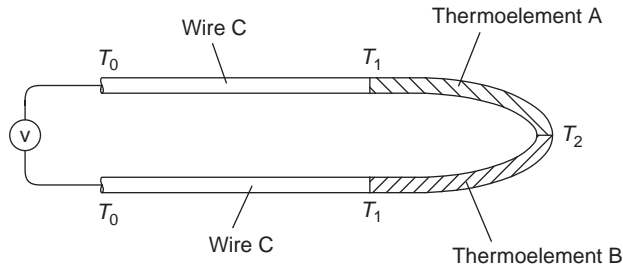


FIGURE 14.8 Use of connection leads for a thermocouple.

of the connections at the multimeter or data logger does not contribute to the emf produced by the thermocouple circuit. In effect the reference temperature T_1 has been moved from the data logger to the terminus connections.

Use of the circuit illustrated in Figure 14.8 requires the temperature T_1 to be known. This can be achieved by using an alternative method of temperature measurement at this junction such as a thermistor or PRT. Alternatively, the terminus connections can be submerged in an ice bath. The temperature at T_1 will then be known and at 0°C .

In practical use, it is uncommon to need values for the Seebeck coefficients. Instead, the emf temperature characteristics of certain selected thermocouples have been identified under controlled conditions and standardized and this data can be used. If, however, you are involved in the design of a new type of thermocouple or some unusual application, data for Seebeck coefficients of a wide range of materials is tabulated in Pollock (1991) and Mott and Jones (1958).

14.3.4.6 Thermocouple Types There are hundreds of types of thermocouple that have been developed. In principle, almost any dissimilar metals and even semiconductors can be used to form thermocouples. The wide scope of alloys indicates the range possible. Kinzie (1973) for example lists more than 300 combinations of materials for thermocouples. A series of international and national standards are however available for thermocouples listing just eight combinations of materials that are widely used. Stringent guidelines are provided for these thermocouples to ensure that devices that are manufactured to be compatible with the specific standard concerned from one company perform in a similar fashion to those manufactured by another company.

The eight standardized thermocouples fall into three general categories:

1. rare metal thermocouples (types B, R, and S),
2. nickel-based thermocouples (types K and N),
3. constantan negative thermocouples (types E, J, and T).

In the rare metal group, types B, R, and S are based on platinum and its alloys with rhodium. These are the most stable of the standard thermocouples and can be used at high temperatures (up to 1750°C) but are generally more expensive and sensitive to contamination. In nickel-based thermocouples, types N and K are commonly used for applications not requiring the elevated temperature range of the rare metal thermocouples. In constantan negative leg thermocouples, types E, J, and T have high emf outputs as constantan has a strong negative Seebeck coefficient.

TABLE 14.4 Manufacturing Tolerances for ASTM Thermocouples (ASTM E 230-98)

Type	Temperature Range (°C)	Standard Tolerance (°C)	Special Tolerance (°C)
B	870 to 1700	$\pm 0.5\%$	$\pm 0.25\%$
E	0 to 870	± 1.7 or $\pm 0.5\%$	± 1.0 or $\pm 0.4\%$
J	0 to 760	± 2.2 or $\pm 0.75\%$	± 1.1 or $\pm 0.4\%$
K	0 to 1260	± 2.2 or $\pm 0.75\%$	± 1.1 or $\pm 0.4\%$
N	0 to 1260	± 2.2 or $\pm 0.75\%$	± 1.1 or $\pm 0.4\%$
R	0 to 1480	± 1.5 or $\pm 0.25\%$	± 0.6 or $\pm 0.1\%$
S	0 to 1480	± 1.5 or $\pm 0.25\%$	± 0.6 or $\pm 0.1\%$
T	0 to 370	± 1.0 or $\pm 0.75\%$	± 0.5 or $\pm 0.4\%$
E	-200 to 0	± 1.7 or $\pm 1.0\%$	
K	-200 to 0	± 2.2 or $\pm 2.0\%$	
T	-200 to 0	± 1.0 or $\pm 1.5\%$	

The criteria for thermocouple selection include cost, maximum and minimum operating temperatures, chemical stability, material compatibility, atmospheric protection, mechanical limitations, duration of exposure, sensor lifetime, sensitivity, and output emf.

Standards organizations throughout the world have specified allowable tolerances for the deviation of the thermocouple output from standardized tables. Many of the standards are based on the results and research at national laboratories, and there is similarity between the magnitude of tolerances in the standards from one country to another. Tables 14.4 and 14.5 give the manufacturing tolerances for ASTM and British Standard thermocouples. Many countries use the term Class 1 for special tolerance and Class 2 for standard tolerance.

The data for thermoelectric emf versus temperature for two of the standardized thermocouples, types K and T, according to IEC 584.1: 1995 and BS EN 60584.1: 1996 is reproduced in Tables 14.6 and 14.7 and for all the standardized thermocouples plotted in

TABLE 14.5 Manufacturing Tolerances for Thermocouples (BS EN 60584.2)

Type	Tolerance Class 1	Tolerance Class 2
B		$\pm 0.25\%$ (600°C to 1700°C)
E	$\pm 1.5^\circ\text{C}$ (-40°C to 375°C) $\pm 0.4\%$ (375°C to 800°C)	$\pm 2.5^\circ\text{C}$ (-40°C to 333°C) $\pm 0.75\%$ (333°C to 900°C)
J	$\pm 1.5^\circ\text{C}$ (-40°C to 375°C) $\pm 0.4\%$ (375°C to 750°C)	$\pm 2.5^\circ\text{C}$ (-40°C to 333°C) $\pm 0.75\%$ (333°C to 750°C)
K	$\pm 1.5^\circ\text{C}$ (-40°C to 375°C) $\pm 0.4\%$ (375°C to 1000°C)	$\pm 2.5^\circ\text{C}$ (-40°C to 333°C) $\pm 0.75\%$ (333°C to 1200°C)
N	$\pm 1.5^\circ\text{C}$ (-40°C to 375°C) $\pm 0.4\%$ (375°C to 1000°C)	$\pm 2.5^\circ\text{C}$ (-40°C to 333°C) $\pm 0.75\%$ (333°C to 1200°C)
R	$\pm 1.0^\circ\text{C}$ (0°C to 1100°C) $\pm (1 + 0.003(t - 1100))$ (1100°C to 1600°C)	$\pm 1.5^\circ\text{C}$ (0°C to 600°C) $\pm 0.25\%$ (600°C to 1600°C)
S	$\pm 1.0^\circ\text{C}$ (0°C to 1100°C) $\pm (1 + 0.003(t - 1100))$ (1100°C to 1600°C)	$\pm 1.5^\circ\text{C}$ (0°C to 600°C) $\pm 0.25\%$ (600°C to 1600°C)
T	$\pm 0.5^\circ\text{C}$ (-40°C to 125°C) $\pm 0.4\%$ (125°C to 350°C)	$\pm 1.0^\circ\text{C}$ (-40°C to 133°C) $\pm 0.75\%$ (133°C to 350°C)

TABLE 14.6 Standard Reference Data to IEC 584.1:1995 and BS EN 60584.1 Part 4: 1996 for Type K Nickel–Chromium/Nickel–Aluminium Thermocouples Giving Thermocouple emf in Microvolts for Various Tip Temperatures Assuming a Cold Junction at 0°C

Type K											
°C	0	10	20	30	40	50	60	70	80	90	100
−200	−5891	−6035	−6158	−6262	−6344	−6404	−6441	−6458			
−100	−3554	−3852	−4138	−4411	−4669	−4913	−5141	−5354	−5550	−5730	−5891
0	0	−392	−778	−1156	−1527	−1889	−2243	−2587	−2920	−3243	−3554
0	0	397	798	1203	1612	2023	2436	2851	3267	3682	4096
100	4096	4509	4920	5328	5735	6138	6540	6941	7340	7739	8138
200	8138	8539	8940	9343	9747	10153	10561	10971	11382	11795	12209
300	12209	12624	13040	13457	13874	14293	14713	15133	15554	15975	16397
400	16397	16820	17243	17667	18091	18516	18941	19366	19792	20218	20644
500	20644	21071	21497	21924	22350	22776	23203	23629	24055	24480	24905
600	24905	25330	25755	26179	26602	27025	27447	27869	28289	28710	29129
700	29129	29548	29965	30382	30798	31213	31628	32041	32453	32865	33275
800	33275	33685	34093	34501	34908	35313	35718	36121	36524	36925	37326
900	37326	37725	38124	38522	38918	39314	39708	40101	40494	40885	41276
1000	41276	41665	42053	42440	42826	43211	43595	43978	44359	44740	45119
1100	45119	45497	45873	46249	46623	46995	47367	47737	48105	48473	48838
1200	48838	49202	49565	49926	50286	50644	51000	51355	51708	52060	52410
1300	52410	52759	53106	53451	53795	54138	54479	54819			

Figure 14.9. In the tables just part of the data available is presented in 10°C steps. Complete data sets, with listings in 1°C steps, are available in the standards and have also been reproduced in many manufacturers' catalogs.

14.3.4.7 Thermocouple Assemblies and Installation Temperature measurement can rarely be undertaken using a bare thermocouple wire. Often the thermocouple wires must be electrically isolated from the application and protected from the environment. When measuring the temperature of a moving fluid, it may also be necessary to locate the thermocouple within a specialized assembly to produce a measurement that can be related to the temperature of the flow. A thermocouple assembly therefore involves consideration of

TABLE 14.7 Standard Reference Data to IEC 584.1:1995 and BS EN 60584.1 Part 5: 1996 for Type T Copper/Copper–Nickel Thermocouples giving Thermocouple emf in Microvolts for Various Tip Temperatures Assuming a Cold Junction at 0°C

[illegible]

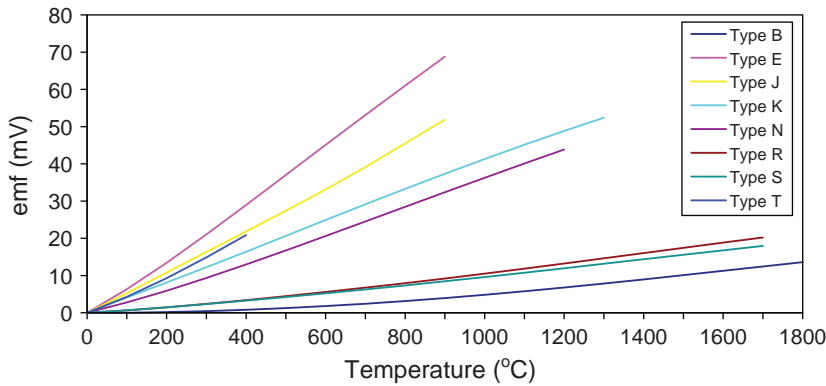


FIGURE 14.9 emf versus temperature characteristics for the standard thermocouples.

joining the wires at the tip to form the thermoelectric junction, electrical isolation of the wires, protection of the wires, installation of the assembly into the application, and connection of the thermocouple to the voltage-measuring device.

As indicated by the law of intermediate materials, the precise form of connection of the two thermoelements does not significantly affect the output provided the temperature of the bead formed is relatively uniform. Thermocouple beads can therefore be formed by twisting few millimeters of the wires' end together at the tip, which provides only limited strength; crimping; or welding as illustrated in Figure 14.10. The choice of connection depends on the requirements of the application such as the need for strength or high speed of response. For applications requiring high strength of the bead or resistance to vibration, then a welded bead may be most appropriate. The vast majority of commercial thermocouples are welded during manufacture and this need not be a concern for the bulk of users. If, however, thermocouples are being assembled from wire then the bead can be formed by using a discharge-welding machine.

Many applications require the thermocouple wires to be electrically or chemically isolated from the environment or medium of interest. Examples of insulation materials

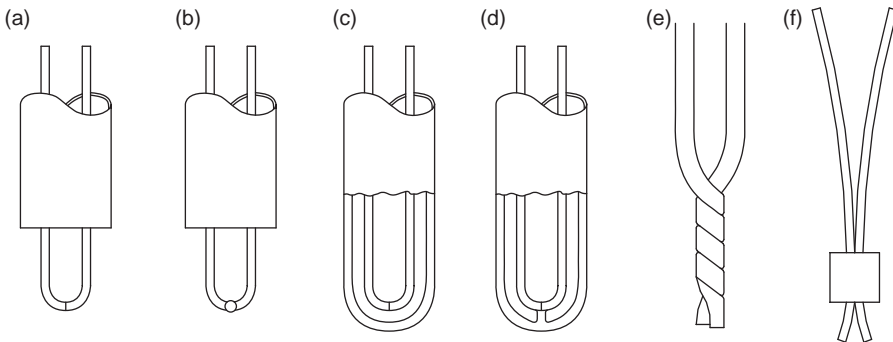


FIGURE 14.10 Bead formation options: (a) butt welded, (b) welded bead, (c) butt welded and sheathed, (d) grounded and sheathed, (e) twisted wires, and (f) crimped.

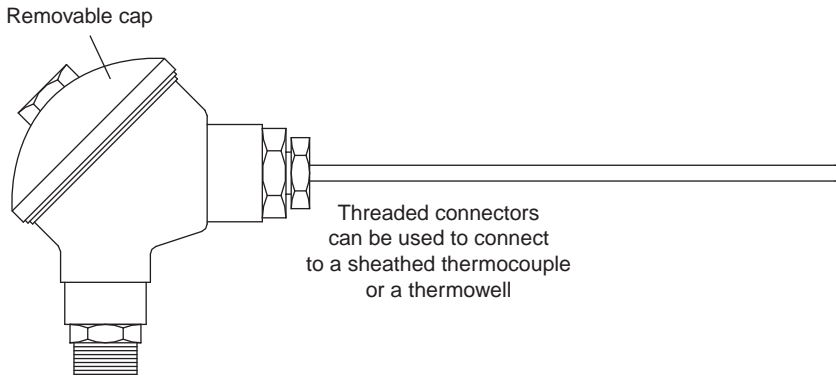


FIGURE 14.11 Some protection tube and connector designs.

include PVC (polyvinyl chloride) for temperatures from -30°C to 105°C , Teflon for -273°C to $+250^{\circ}\text{C}$, glass fiber for -50°C to 400°C , and polyimide for -269°C to 400°C . Higher temperatures can be achieved using ceramic sheaths. The requirements of electrical or chemical isolation and good thermal contact are often in conflict, giving rise to thermal disturbance errors.

In addition to the need to electrically isolate the thermocouple, it is also necessary in some applications to protect the thermocouple from exposure to the local environment that could otherwise impair the function of the measurement device. An example is the immersion of a thermocouple into a corrosive fluid. Levels of protection can be achieved by use of a protection tube or sheath around the insulation. An example of a thermocouple assembly with a protecting tube and head is illustrated in Figure 14.11. These tubes and connectors are commercially available. Choice of the insulator and sheath materials depends on the application. The ASTM manual (1993) on the use of thermocouples in temperature measurement lists a wide range of protecting tube materials for different applications. The MIMS (mineral insulated metal sheathed) thermocouple attempts to combine high temperature capability and protection from the environment in a single assembly. In these devices, a mineral such as magnesium oxide is compacted around the thermocouple wires to electrically isolate and support them and this assembly is encapsulated within a metal sheath, Figure 14.12. MIMS thermocouples can be used for the range -200°C to 1250°C .

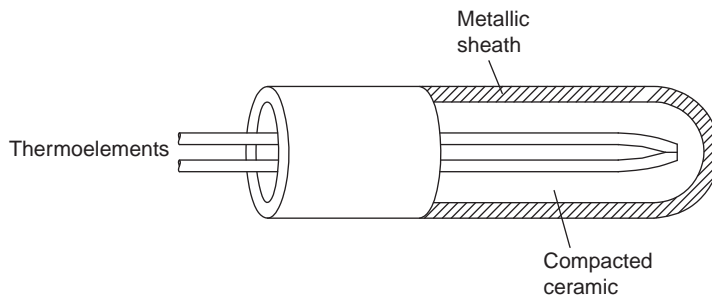


FIGURE 14.12 A MIMS (mineral insulated metal sheathed) thermocouple for high temperature capability and protection from the local environment.

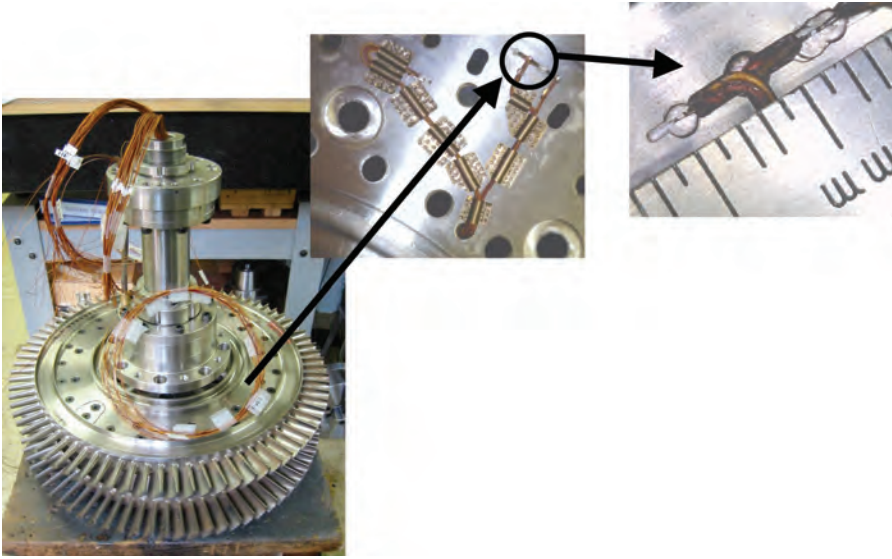


FIGURE 14.13 Thermocouple installation (Coren et al., 2010).

Thermocouples are useful for determining the temperature for a wide range of applications from surfaces and solid bodies to stationary and flowing fluids. As a thermocouple consists simply of the thermoelectric wires it is usually necessary to insulate the wires and sometimes encapsulate the wires in a protective sheath. In addition, in order to avoid errors caused by heat transfer, it may be necessary to install the thermocouple within a specialized assembly to ensure appropriate measurement of the parameter of interest.

The temperature of a solid surface can be measured using a thermocouple by a variety of means depending on the type of surface and uncertainty desired. Typically the choice is limited to surface mounting or installing the thermocouple in a hole or recess. In the case of surface mounting, an adhesive or shims may be used to secure the thermoelement wires to the surface and to improve thermal contact between the surface of interest and the thermocouple bead. An example of the use of shims for a gas turbine application is shown in Figure 14.13.

Solid temperatures and the temperature of a liquid flowing in a pipe are sometimes measured with the thermocouple immersed in a thermowell. A thermowell is a protecting tube to prevent or minimize damage from harmful atmospheres, corrosive fluids, or mechanical damage; common commercial forms are illustrated in Figure 14.14. As a general rule of thumb, a thermowell should be immersed in the liquid to a depth of 10 times the diameter. Material property requirements for thermowells are documented in ASME PTC 19.3 (1974).

A thermocouple placed in a gas environment will experience heat transfer by conduction along its wires and support, convective heat transfer with the surrounding gas both at the tip and along the length of the wires and support, and radiative heat exchange with the surroundings. The contribution due to conduction is usually relatively small in comparison to convection. Radiative heat transfer tends to be a function of the fourth power of the absolute temperature and can therefore be significant if the temperature of the gas is distinctly different to the temperature of the enclosure. Because of these exchanges of heat

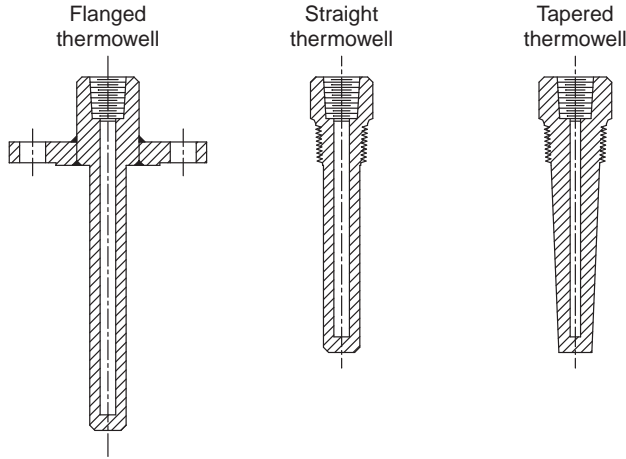


FIGURE 14.14 Thermowells (ASTM, 1993)

the measurement of gas temperature needs to be carefully thought through and undertaken (see Childs, 2001a,b).

14.3.4.8 Extension Leads and Compensation Cable In some applications the measurement location and readout instrumentation are separated by a considerable length. An example is the measurement of temperatures of various different components in an engine on test, with for example, a hundred thermocouples all connected to a data acquisition system some 30 m away. In most applications the thermal gradient and resultant emf is most significant in the first few meters of the wire. The remaining wire length serves the purpose of transmitting the emf to the data acquisition system. The thermoelectric properties of this length of wire are less critical than the part in the region of large thermal gradient. The drive for economy can make the use of a lower specification and hence cheaper wire for this region of wire attractive, and these lengths of wire are known as extension leads or extension wires. There are two types of extension wire: those with similar physical composition to the thermocouple wire itself but manufactured to a less stringent specification and those manufactured using a different material altogether. Some people use the term extension leads for wires manufactured using the same material as the thermocouple and the term compensation lead for wires manufactured using different materials. However, this terminology is not universal and the terms extension and compensation lead, wire and cable are used interchangeably. The use of extension leads in a thermocouple circuit is illustrated in Figure 14.15.

For the configuration shown in Figure 14.15 loop analysis can again be used to analyze the output. If the Seebeck coefficients of the extension leads are S_A , and S_B , respectively, then,

$$E = S_{A'}(T_1 - T_0) + S_A(T_2 - T_1) + S_B(T_1 - T_2) + S_{B'}(T_0 - T_1) \quad (14.16)$$

This simplifies to

$$E = S_{A'B'}(T_1 - T_0) + S_{AB}(T_2 - T_1) \quad (14.17)$$

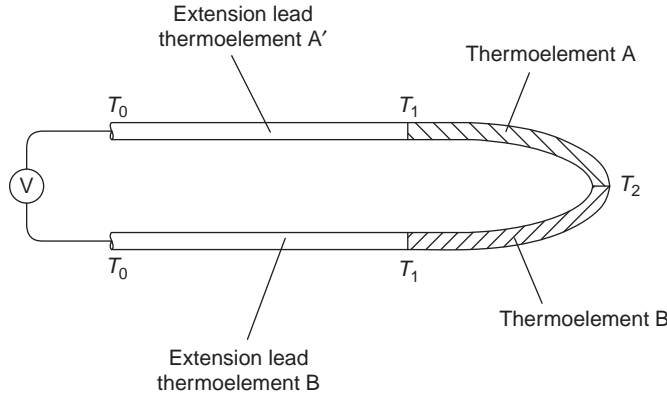


FIGURE 14.15 The use of extension leads in a thermocouple circuit (Childs, 2001a,b).

If the wire pair A'B' is selected so that it has approximately the same relative Seebeck coefficient as AB, then

$$S_{A'B'} \approx S_{AB} \quad (14.18)$$

and

$$E \approx S_{AB}(T_2 - T_0) \quad (14.19)$$

So the reference junction is effectively located at the connections between the voltmeter or the data logger and the extension leads.

14.3.5 Resistance Temperature Devices

The electrical resistance of a material is a function of temperature, and this phenomena is exploited by the class of instruments called resistance temperature detectors (RTDs). Types of RTD include platinum resistance thermometers (PRTs), thermistors, and a variety of semiconductor-based instruments. RTDs are used in a wide range of applications. Some types such as PRTs can have very low uncertainty, better than 0.002 K (Mangum and Furukawa, 1990) and are used in the definition of the ITS-90, and others such as thermistors and some semiconductor-based devices can be inexpensive in comparison to, say, thermocouples.

In an RTD, the measurement of resistance is made across the length of material. This is usually a metal although germanium and carbon are used in some cryogenic (<273 K) applications. Any metal could in principle be used but stability considerations normally limit the choice to platinum, copper, and nickel. Copper and nickel are useful in the ranges -212°C to 350°C , respectively (Claggett et al., 1995), and are cheap in comparison to platinum. Platinum is more chemically stable and tends to be the preferred material for the majority of metal-based RTDs where low uncertainty is required or a message of quality is important in branding. PRTs are used to define part of the ITS-90.

14.3.5.1 Platinum Resistance Thermometers Platinum resistance thermometers consist of a length of platinum, connection wires, and an instrument to measure the resistance. Platinum resistance thermometers fall into a number of broad categories:

- Standard platinum resistance thermometers (SPRTs) which are used for low uncertainty measurements.
- Industrial platinum resistance thermometers (IPRTs) which are used for practical laboratory and industrial use.
- Secondary SPRTs which are designed for laboratory environments and limited temperature ranges while still providing low uncertainty measurements.

The relationship between resistance and temperature can be approximated by

$$R_t = R_0(t + \alpha t) \quad (14.20)$$

where R_0 is the resistance at 0°C (Ω); R_t the resistance at temperature t (Ω); α the temperature coefficient of resistance (1°C) and t the temperature ($^\circ\text{C}$).

The temperature coefficient of resistance, α , is calculated from

$$\alpha = \frac{R_{100} - R_0}{100^\circ\text{C} \times R_0} \quad (14.21)$$

where R_{100} is the resistance at 100°C (Ω).

Much of the original work on resistance thermometry was undertaken by Siemens (1871) and Callendar (1887, 1891, 1899). Callendar found that the resistance of platinum could fairly accurately be described by a quadratic in the form

$$R_t = R_0(1 + At + Bt^2) \quad (14.22)$$

where A and B are constants.

This equation has traditionally been written in the form given in Equation (14.23), known as Callendar's equation, which simplifies the calculations necessary to determine the calibration constants α and δ .

$$R_t = R_0 \left[1 + \alpha t + \alpha \delta \left(\frac{t}{100} \right) \left(1 - \frac{t}{100} \right) \right] \quad (14.23)$$

where δ is a constant.

For temperatures below 0°C, additional terms are needed and the temperature resistance characteristic is relatively accurately defined by the Callendar–Van Dusen equation, Equation (14.24), which was the basis of the now superseded International Practical Temperature Scales of 1927, 1948, and 1968.

$$R_t = R_0(1 + At + Bt^2 + C(t - 100)t^3) \quad (14.24)$$

TABLE 14.8 Typical Values for the Constants in the Callendar–Van Dusen Equation for an SPRT and an IPRT (Nicholas and White, 1994)

Constant	SPRT	IPRT
A	$3.985 \times 10^{-3}/^{\circ}\text{C}$	$3.908 \times 10^{-3}/^{\circ}\text{C}$
B	$-5.85 \times 10^{-7}/^{\circ}\text{C}^2$	$-5.80 \times 10^{-7}/^{\circ}\text{C}^2$
C	$-4.27 \times 10^{-12}/^{\circ}\text{C}^4$	$-4.27 \times 10^{-12}/^{\circ}\text{C}^4$
α	$3.927 \times 10^{-3}/^{\circ}\text{C}$	$3.85 \times 10^{-3}/^{\circ}\text{C}$

Here C is a constant and is zero above 0°C . Typical values for the coefficients in Equation (14.24) for a standard PRT and an industrial PRT are listed in Table 14.8.

The resistance element may be a coil of fine wire or a track of platinum deposited on a surface. The design of the resistance element depends on the requirements of the thermometer (see ESDU, 2007b). Most materials tend to expand and contract with temperature and this induces strain in a material. This can affect the PRT in a number of ways. Expansion of a track of platinum can alter the cross-sectional area available for conduction and therefore alter the resistance. In addition, repetitive strain cycling can modify the microscopic structure of the platinum and therefore the resistance. A further compounding effect is any variation in the electrical properties of the insulation and connection wires.

Standards (BS 1041: Part 3, 1989; ASTM E1137, 1997; BS EN 60751, 1996) have been produced for PRTs. Values for resistance and the corresponding temperature can be readily interpolated in order to determine a given reading. Two classes of uncertainty are defined for IPRTs: Class A devices where the uncertainty is within $\pm(0.15 + 0.002|t|)$ and Class B devices, with an uncertainty $\pm(0.3 + 0.005|t|)$.

The delicate strain-free designs of an SPRT would not survive the shock and vibration encountered in some industrial environments. The IPRT typically comprises a platinum wire encapsulated within a ceramic housing or a thick film sensor coated onto a ceramic surface. The actual sensing element of an IPRT can be further protected from the environment by a metal sheath. In the design of an IPRT, the constraint for mounting the platinum wire so that strain is minimized is relaxed in favor of producing a robust device. Most designs have just two wires sealed within the sensor body as this reduces the risk of short circuits and makes the device more robust. The ceramic encapsulation used should be selected so that its thermal coefficient of expansion closely matches that of platinum and the former over the temperature range of the PRT. The ceramic material used for encapsulation can include compounds and impurities that react with the platinum and change the transducer's temperature coefficient. The uncertainty achievable for an IPRT is generally of the order of $\pm 0.01^{\circ}\text{C}$ to $\pm 0.2^{\circ}\text{C}$ over the range 0 – 300°C (see Hashemian and Petersen, 1992).

The platinum-sensing element can also be deposited as a thin or thick film on an insulating substrate such as alumina. For a thick film sensor, the surface area can be of the order of 25 mm^2 with an active area of 5 mm^2 . They are mass-produced with each sensor individually trimmed to produce the desired resistance. The thick film structure produces a device that is robust and capable of operating in applications experiencing minor shock or vibration. They are noninductive and therefore insensitive to stray electrical fields. Thick film PRTs are normally designated to Class B of IEC-751 (1993). An innovative

form of temperature measurement is the application of thin film PRTs to polyamide sheet (Jones, 1995). Massed arrays of PRTs can be formed on a polyamide sheet substrate by evaporating or sputtering platinum followed by depositing gold leads using a magnetron. The resulting sheet is flexible and can be glued to the geometry of interest.

In order to obtain a measure of temperature using an RTD, the resistance must be measured. Simplistically, this can be achieved by passing a current through the sensing resistor and measuring the potential across it. If the current is known, then the ratio of potential to current provides the measurement of the resistance from Ohm's law, and the temperature can be determined from the resistance temperature characteristic. However, even if an uncertainty of only 1°C is desired, the resistance must be measured to better than 0.4% and as a result must be undertaken with care (Nicholas and White, 1994). There are two basic methods for measuring resistance commonly used in platinum resistance thermometry, potentiometric methods, and bridge methods, and these are reviewed by Nicholas and White (1994), Connolly (1998), and Wolfendale et al. (1982).

14.3.5.2 Rhodium Iron, Doped Germanium and Carbon Resistors A number of types of RTD are suitable for cryogenic applications including platinum, rhodium iron, doped germanium, and carbon. Although used to define the ITS-90 between 1234.93 K and 13.81 K, the sensitivity of platinum falls off below about 20 K. Rhodium iron alloy has a similar resistance temperature characteristic to platinum at temperatures above 30 K. Below 30 K the sensitivity drops to a minimum between 25 K and 15 K and then rises again giving a RTD with good sensitivity at low temperatures (see Rusby, 1982; Schuster, 1992). Doped germanium resistors are available with a relatively wide temperature range for cryogenic applications, from 0.05 K to 325 K and typically comprise the semiconductor encapsulated in a 3 mm diameter, 8.5-mm long cylinder (example cited is the GR-200A from Lakeshore Cryotronics). Certain carbon resistors produce stable RTDs.

14.3.5.3 Thermistors Thermistors use the variation in resistance with temperature of ceramic semiconductor. Modern thermistors are usually mixtures of oxides of chromium, nickel, manganese, iron, copper, cobalt, titanium, other metals, and doped ceramics. Thermistors can have either a positive temperature coefficient (PTC) or a negative temperature coefficient (NTC). A typical resistance temperature characteristic is 1 Ω/0.01°C. The temperature can be determined approximately from the relationship given in Equation (14.25) (Wood et al., 1978).

$$R_T = R_0 \exp \left[1 - B \left(\frac{1}{T} - \frac{1}{T_0} \right) \right] \quad (14.25)$$

where R_0 is the resistance at T_0 and B a constant for the particular thermistor material.

The resistance characteristic of a thermistor expressed by Equation (14.25) is negative and nonlinear. This can be offset if desired using two or more matched thermistors packaged in a single device so that the nonlinearities of each device negate each other (see Beakley, 1951; Sapoff and Oppenheim, 1964; Sapoff, 1980). Thermistors are usually designated by their resistance at 25°C, with common resistances ranging from 470 Ω to 100 kΩ. The high resistivity of thermistors normally negates the need for a four-wire bridge circuit.

In order to produce a lower uncertainty fit for the resistance temperature characteristic of a thermistor, a polynomial can be used with the degree used depending on the temperature range and the type of thermistor. Equation (14.26) is adequate for most systems.

$$\ln R_T = A_0 + \frac{A_1}{T} + \frac{A_2}{T^2} + \frac{A_3}{T^3} \quad (14.26)$$

Here A_0 , A_1 , A_2 , and A_3 are constants.

Alternatively Equation (14.27) can be used where the second order term has been neglected.

$$\frac{1}{T} = b_0 + b_1 R_T + b_3 (\ln R_T)^3 \quad (14.27)$$

where b_0 , b_1 , and b_3 are constants.

14.3.6 Semiconductor Devices

A forward voltage across a transducer junction can be used to generate a temperature sensing output proportional to absolute temperature. Temperature sensors based on simple transistor circuits can be readily incorporated as a part of an integrated circuit to provide on-board diagnostic or control capability. The majority of semiconductor junction sensors use a diode connected bipolar transistor. If the base of the transistor is shorted to the collector then a constant current flowing in the remaining p-n junction (base to emitter) produces a forward voltage that is proportional to absolute temperature. This can be modeled by Equation (14.28) and in practice has a temperature coefficient of approximately $-2 \text{ mV}/^\circ\text{C}$.

$$V_F = \left(\frac{kT}{q} \right) \ln \left(\frac{I_F}{I_S} \right) \quad (14.28)$$

where k is the Boltzmann's constant ($1.38 \times 10^{-23} \text{ J/K}$); T the temperature (K); q the charge of electron ($1.6 \times 10^{-19} \text{ C}$); I_F the forward current (A); and I_S the junction's reverse saturation current (A).

14.3.7 Diode Thermometers

The forward voltage drop across a p-n junction increases with decreasing temperature. For certain semiconductors the relationship between voltage and temperature is almost linear and in silicon this occurs between 400 K and 25 K with a sensitivity of approximately 2.5 mV/K . Commonly used semiconductors for thermometry are Si and GaAs with the addition of certain elements for stability. The merits of diode thermometers include price, simple voltage temperature relationship, relatively large temperature range (1–400 K), no need for a reference junction, relatively high sensitivity, uncertainty lower than $\pm 50 \text{ mK}$ (Krause and Dodrill, 1986), and simplicity of operation with a constant current source and a digital voltmeter.

14.3.8 Noise Thermometry

The thermal motion of atoms and molecules and charge carriers within an electrical conductor generates broadband electrical noise (Johnson, 1928; Nyquist, 1928). The noise power generated is called Johnson noise and is related to temperature by

$$\text{power} = 4kT\Delta f \quad (14.29)$$

where k is the Boltzmann's constant (J/K); T the absolute temperature (K); and Δf the frequency band (Hz).

In noise thermometry the thermal noise is measured across a resistor. The thermal noise power generated in a resistor is usually very small and sensitive instrumentation is therefore necessary. In principle the temperature range of application is from a few mK to over 2500°C. The circuitry used depends on the temperature range and has been reviewed by Kamper (1972), Blalock and Shepard (1982), and White et al. (1996).

14.3.9 Pyrometric Cones

A number of deforming temperature and heat rate indicating devices have been developed, known as pyrometric cones, thermoscope bars, and Bullers rings, for use in kilns and furnaces. Pyrometric cones, also known as Harrison, Seger, or Orton cones, are slender trihedral pyramids manufactured from mixtures of various mineral oxides (Seger, 1886). On heating the effects of gravity cause the inclined cone to progressively bend as the material softens, see Figure 14.16. Pyrometric cones do not provide a precise indication of temperature as their deformation is a function of both the rate of heating and the temperature cycle. They tend to be used to provide an indication of process completion such as completion of firing.



FIGURE 14.16 Pyrometric cones.

Cones vary in shape, size, and method of use according to the manufacturer. They are an economic means of identifying firing completion and can be distributed at a number of locations in a kiln to indicate local conditions. An alternative is the use of thermocouples, RTDs, or infrared thermometers but these have a higher initial cost. One manufacture, Orton Ceramics, provides cones suitable for temperatures in the range 590–2015°C. Determination of performance equivalents between different manufacturers is standardized in ANSI/ASTM C24-79 and reported by Fairchild and Peters (1926) and Beerman (1956). The practical use of pyrometric cones is described by McGee (1988), Childs (2001a,b), and in manufacturers' guidelines. Thermoscope bars comprise a rod of material mounted on a stand. The bars soften and deform with temperature, and the deformation can be measured and compared with a look up table supplied by the manufacturer to provide an indication of temperature. Bullers rings are dust pressed rings that shrink linearly with temperature rise. Use of thermoscope bars and Bullers rings is outlined in BS 1041 Part 7 (1988).

14.4 SEMI-INVASIVE METHODS

Some temperature measurement scenarios permit the modification of a surface of interest with a thermally sensitive material that can be observed remotely. These techniques can be classed as semi-invasive and include heat-sensitive paints, thermographic phosphors, thermochromic liquid crystals, and variety of crayons and labels. Some of the materials available provide a means of continuously monitoring temperature while others merely provide a means of identifying that a certain temperature has been reached or exceeded. The latter are known as peak temperature indicators.

14.4.1 Peak Temperature-Indicating Devices

A number of products are available that provide an indication that a particular temperature has been attained or exceeded. On heating above some critical temperature the material used for the indicator melts, fuses, or changes composition providing a permanent record that a given temperature has been reached or more likely exceeded. Commercially available products include heat-sensitive crayons, pellets, labels, and paints.

Fuseable material can be applied to a surface by means of a crayon. Applications include identifying temperatures in metal treatment and welding. Self-adhesive labels consisting of a temperature-sensitive indicator sealed under a transparent window are also available. These provide a convenient means of identifying whether a component such as an electronics package has experienced an excessive temperature or in medical applications to validate that a process temperature has been attained. The indicating material permanently changes color from say a light gray to black at the rated temperature, which is printed on the label. Some labels include multiple indicators, sensitive to a range of temperatures, to enable greater resolution of the peak temperature.

14.4.2 Temperature-Sensitive Paints

The same kind of fusible material as used in the peak temperature-indicating crayons is available in paint form. Thermally sensitive paint has also been developed for high temperature applications in order to indicate the peak temperatures on turbine blades

Neumann, 1989. Early work on turbines involved the use of proprietary paints from Thermindex, Faber Castel, and Tempil. The development of a range of multichange paints is reported by Bird et al. (1998). Some paint formulations are available that provide a reversible indication of peak temperature. One principle exploited is to drive water off a colorful salt, thus changing its color. On cooling, the salt can re-absorb water vapor from the atmosphere and revert to its initial color state.

Parameter-sensitive paints (PSPs) are sensitive to a number of physical parameters including pressure and temperature. The paints comprise luminescent molecules and a polymer binder material that can be dissolved in a solvent. Temperature-sensitive paints (TSPs) can be applied by brush or spray and as the paint dries, the solvent evaporates to leave luminescent molecules embedded in a polymer matrix. The principle exploited by TSPs is photoluminescence where a probe molecule is promoted to an excited state by the absorption of a photon of appropriate energy in a fashion comparable to thermographic phosphors. The difference is that the intensity of the luminescence is related to temperature by photophysical processes known as thermal and oxygen quenching (Gallery et al., 1994) and the intensity of luminescence is inversely proportional to the temperature.

14.4.3 Thermographic Phosphors

The luminescent properties of materials, known as thermographic phosphors, such as the lanthanide doped ceramics YAG:Tb, YAG:Dy, or Y₂O₃:Eu, are temperature dependent. In order to measure the temperature of a surface temperature, a thin layer of luminescent materials can be applied, illuminated with UV light, typically from a laser, and the subsequent luminescence observed and analyzed. Thermographic phosphors can be used to indicate temperatures from cryogenic levels to about 2000°C. A phosphor-based thermometry system will generally consist of a source of excitation energy, a method of delivery of energy to the target, a fluorescing medium bonded to the target, optical, detector and data acquisition, and analysis systems. Before excitation phosphor's electronic levels are populated in their ground state. They can be excited by the absorption of energy by, for instance, electromagnetic radiation such as visible or UV light, X- or γ -rays, particle beams, or an electric field. After absorption of energy the atomic configuration of the material may not remain excited but return to its initial or some intermediate state. The intensity of emitted light is an inverse function of temperature. In the case of Europium-based phosphors experiencing continuous illumination, it is given by Equation 14.18 (Fonger and Struck, 1970).

$$I_T = \left[a_j + a_j A e^{-\Delta E/kT} \right]^{-1} \quad (14.30)$$

where I_T is the intensity; a_j the probability rate; A the factor related to a_j ; E the energy (J); T the temperature (K); k the Boltzmann's constant (J/K); and a_{CTS} the charge transfer state rate.

Sensitivities of 0.05°C and an uncertainty of 0.1–5% of the Celsius temperature reading are possible (Allison and Gillies, 1997). Most thermographic phosphors are ceramics and therefore relatively inert. Bonding of the phosphor to the application can be achieved by mixing the phosphor slurry with epoxy, paint, or glue and then applying the mixture

onto the surface by brushing or spraying. For harsh environments subject to, for example, high mechanical shock or large thermal gradients, chemical bonding by vapor deposition, RF sputtering, and laser ablation can also be considered. Thermographic phosphors are reviewed by Heyes (2004, 2009), and the development of a thermal barrier coating for turbomachinery components capable of both sustaining a high thermal gradient and indicating temperature is described by Feist et al. (2001, 2002).

14.4.4 Thermochromic Liquid Crystals

Thermochromic liquid crystals are a class of materials that display significant changes in color over discrete temperature bands. As the temperature rises, a thermochromic liquid crystal mixture will change from colorless, black against a blackened background and pass through the visible spectrum of colors; red to orange to yellow to green to blue to violet, before turning colorless again. The process of color change is reversible with cooling. The colors displayed can be related to temperature by a calibration process. The sensitivity to temperature arises from the structure within the thermochromic liquid crystal with transition between solid and liquid states. In the liquid crystal phase, molecules lose positional order becoming fluid, but retain orientation order and the optical properties of crystalline solids (Kakade et al., 2009a). When a thermochromic liquid crystal is illuminated with white light, it will selectively reflect monochromatic light with wavelength equal to the pitch of the helical molecular structure. The wavelength of the reflected light decreases as the temperature increases and the color observed will therefore change from red through the visible spectrum to blue.

Pure crystals deteriorate rapidly with age and exposure to ultraviolet light but their life can be extended by encapsulation of the crystals within a polymer coating. Thermochromic crystals are commercially available in the form of water-based slurry. This can be sprayed onto a dark or blackened surface using an airbrush (see, for example, Baughn, 1995; Roberts and East, 1996; Kakade et al., 2009a, for practical guidance). They are also available as a pre-formed layer on a blackened substrate of mylar or paper.

The temperature at which a liquid crystal formulation begins to display color is called its red-start temperature. The range of temperature over which the crystals display color is often referred to as the color play bandwidth or the temperature event range. The range of temperature over which a liquid crystal displays a single color is called the isochrome bandwidth and this can be as little as 0.1°C . For narrow band thermochromic liquid crystals the color changes over a range about 1°C and these can be used to determine surface temperature with an uncertainty of about 0.1°C . Wideband thermochromic liquid crystals are typically active across a range 5°C to 20°C and can be used to determine the distribution of surface temperature. The value of the red-start temperature and the color play bandwidth can be controlled by selecting appropriate cholesteric esters and their proportions.

A number of approaches have been adopted in determining the value of temperature from color in the use of thermochromic crystals. For instance in the use of forehead clinical thermometers, human eyesight and judgement is used comparing the displayed color to a look-up table or identifying the most significant illumination in a panel. A CCD camera can also be used to capture the images produced by liquid crystals and the resulting RGB image processed to hue, and a calibration used to convert this to a corresponding temperature (see Kakade et al., 2009a,b).

Applications of thermochromic liquid crystals have included forehead and fish-tank thermometers, novelty stickers, turbine heat transfer experiments (for example Campbell and Molezzi, 1996; Ou et al., 2000), jet engine nacelles, and vehicle interiors (Lee and Yoon, 1998).

14.5 NONINVASIVE METHODS

A number of temperature-dependent phenomena such as the emission of radiation, scattering, and luminescence can be observed remotely thereby providing the possibility of measuring temperature without disturbing the medium of interest. Remote observation can under certain circumstances be classed as noninvasive to the medium of interest or target. It should however be noted that optical techniques, particularly for methods requiring use of a laser system can involve use of bulky and complicated equipment, and while this may not require machining or installation of a measurement device on the surface of interest or location of a sensor in a fluid, the equipment may nevertheless be bulky and will likely require optical access through, say, special windows. Noninvasive methods are potentially attractive for a variety of reasons and applications including

- targets in motion,
- fragile targets,
- remote targets,
- unsteady temperature applications,
- harsh environments,
- temperature distribution required.

Noninvasive methods of temperature measurement have developed with the advances in semiconductor fabrication technology and the availability of lasers as a source of excitation. Use of infrared thermometers is now commonplace for inspection in industry and retail. Temperature measurement techniques based on infrared radiation are reviewed in Sections 14.5.1 and 14.5.2.

A series of sophisticated methods for noninvasive measurement of gas temperature are available such as index of refraction; absorption, and emission spectroscopy; line reversal; scattering; laser-induced fluorescence; light polarization; speed of sound; and speckle methods. Some of the techniques, such as laser-induced fluorescence, are highly specialized and expensive requiring extensive expertise and high-cost capital equipment. This type of method is not available off the shelf and its use should be carefully justified in terms of the value of the data it would produce. These methods are reviewed in ESDU (2002) and Childs (2001a,b).

14.5.1 Infrared Thermometry

Infrared thermometers measure the thermal radiation emitted by a body due to its temperature and have found applications from cryogenic temperatures to over 6000 K. Infrared thermometers are widely available with examples used in industry and retail where the cost is the same order of magnitude as, for example, a thermocouple indicator to highly sophisticated traceable devices for high temperature gas or surface measurement, suitable

for laboratory and research use. Any body will emit energy in the form of thermal radiation due to its temperature with the quantity of radiation rising with increasing temperature. The energy emitted over the electromagnetic spectrum due to temperature by a blackbody is modeled by Planck's law (Equation 14.31). The energy radiated reduces with temperature but the wavelength distribution shifts toward those with longer wavelengths.

$$E_{\lambda b} = \frac{C_1}{\lambda^5 [\exp(C_2/\lambda T) - 1]} \quad (14.31)$$

where $E_{\lambda b}$ is the spectral emissive power for a blackbody (W/m^3); C_1 the first radiation constant ($3.7417749 \times 10^{-16} \text{ Wm}^2$, Cohen and Taylor, 1987); λ the wavelength (m); C_2 the second radiation constant (0.01438769 m K , Cohen and Taylor, 1987); and T the absolute temperature (K).

Integration of Equation (14.31) over all wavelengths gives the total thermal radiation emitted by a blackbody, Equation (14.32).

$$E_b = \sigma T^4 \quad (14.32)$$

where σ is the Stefan–Boltzmann constant. $\sigma = 5.67051 \times 10^{-8} \text{ W/m}^2 \text{ K}^4$ (Cohen and Taylor, 1987).

An infrared system consists of three basic elements, the source, propagation medium, and measuring device. The infrared-measuring device may comprise an optical system, a detector, processing circuit, and display. The purpose of the optical system is to focus the energy emitted by the target onto the sensitive surface of the infrared detector, which usually converts the energy into an electrical signal. There are a number of types of infrared thermometer depending on whether the device is sensitive to all or a specific band of wavelengths. Those sensitive to all wavelengths are classed as total radiation or broadband thermometers. Those sensitive to radiation in a specific band of wavelengths are classed as spectral band thermometers, ratio thermometers, and thermal imagers.

The radiation emitted by real targets does not match the ideal simplification of Equations (14.31) and (14.32). Radiation emitted by real bodies is a function of both the surface temperature and surface properties. The surface property limiting the quantity of radiation is called the emissivity, ϵ , and is the ratio of the electromagnetic flux that is emitted from a surface to that emitted from a blackbody at the same temperature. Emissivity is a function of the dielectric constant and subsequently its refractive index and is generally wavelength dependent. Quantification and accounting for the effects of emissivity in undertaking a temperature measurement using an infrared thermometer is usually necessary. Target surfaces can be classified into three categories:

1. blackbodies,
2. graybodies,
3. nongraybodies.

Information about the emissivity of a wide range of materials is documented in the reference texts edited by Touloukian et al. (1970, 1972a,b) and also tends to be available from infrared thermometer suppliers.

The medium across which the radiation is transmitted can attenuate the radiation due to absorption, scattering, and turbulence with absorption and emission dependent on the energy structure of the constituent molecules. As well as nitrogen and oxygen, air typically contains a proportion of water vapor and CO₂ along with a number of trace molecules. The strong dipole moment and light hydrogen atoms in water vapor result in strong and broad absorption bands. In addition, water vapor is an asymmetric top and this gives an irregular absorption spectrum. The properties of water vapor and CO₂ along with those of other trace substances in air act to absorb a proportion of the radiation emitted from a target at specific wavelengths and even make air opaque at some frequencies. This means that using Equation (14.31) without modification to account for the radiation not transmitted would lead to an error. Information concerning the transmissive properties of air, which varies with distance and local composition, has been published by Yates and Taylor (1960) and is also modeled in a number of software packages such as LOWTRAN (Selby et al., 1972, 1975, 1976, 1978), MODTRAN, HITRAN, and FASCODE. These codes represent long-running research development and are reviewed in Smith (1993) and are available on-line (see, for example, www.cfa.harvard.edu/hitran/, HITRAN, 2010). When dealing with other transmission mediums or specific particulates such as smoke, the transmission characteristics for the system concerned must be identified in order to allow for the temperature reading to be correctly interpreted.

An approach to negate the impact of nontransmission of thermal radiation is to use spectrally sensitive detectors. These operate across a specific waveband only and a detector can be selected to match the characteristics of a given application. For example, there are a number of wavelengths, known as windows, near 0.65 μm , 0.9 μm , 1.05 μm , 1.35 μm , 1.6 μm , 2.2 μm , 4 μm , and 10 μm , where the transmittance is high and air is effectively transparent to thermal radiation. A large number of detectors have now been identified that are sensitive to thermal radiation such as PbS, PbSe, InSb, HgCdTe, and PbSnTe (for further data see for example Rogatto, 1993). As a general rule instruments used for measuring hot targets operate at shorter wavelengths (for example 0.9–1.1 μm) and those for cooler objects operate at longer wavelengths (3–5 μm or 8–14 μm). A further issue arises from the use of optical devices and windows between the target and the detector as materials used for lenses and windows also have their own spectral characteristics and only transmit a proportion of the incident radiation at specific wavelengths. An infrared sensor should only be operated when the spectral range over which a target and the media transmit and over which the detector is sensitive all overlap.

A compounding problem affecting a temperature measurement using an infrared thermometer is background radiation. An infrared thermometer will read any radiation incident on the detector whether it is emitted, transmitted, or reflected from a target. If a surface is not a perfect absorber of incident radiation then a proportion of the incident radiation can be reflected, and this can distort the indicated temperature. If, for example, a target surface is at a temperature, say 300 K, but is in close proximity to a hotter object, then it is possible for an infrared thermometer to read the radiation emitted from both the surface and a proportion of radiation from the hotter object that is reflected off the surface of the target. The temperature indicated for the target will be artificially high, without appropriate offsetting. A number of strategies are available to minimize such effects involving repositioning the detector or using screens. The practical use of infrared thermometers is described by Kaplan (1999).

When an infrared thermometer is aimed at a target it collects energy within a collecting beam, or zone, whose shape is determined by the optical system and detector and is typically conical. The cross section of the collecting beam is called the field of view and this determines the spot size; the area on the target over which a temperature measurement is made.

In principle, it is possible to side step the requirement of knowing the surface emissivity by using ratio thermometers. These measure the radiation emitted from a surface around two fixed wavelengths. The ratio of the radiation emitted is given by

$$R = \frac{\varepsilon(\lambda_1)\lambda_2^5(e^{C_2/\lambda_2 T} - 1)}{\varepsilon(\lambda_2)\lambda_1^5(e^{C_2/\lambda_1 T} - 1)} \quad (14.33)$$

If the emissivity is not spectrally sensitive and is near unity the emissivities will cancel, and provided the target temperature is low, then,

$$T = \frac{C_2(\lambda_1 - \lambda_2)/\lambda_1\lambda_2}{\ln \left[R \left(\frac{\lambda_1}{\lambda_2} \right)^5 \right]} \quad (14.34)$$

An optical fiber can be used to convey thermal radiation into a narrow wavelength band from a location of interest or a convenient viewing point to a detector. This enables the detector to be remote from the target and is particularly useful in monitoring very hot applications such as combustion processes. Optical fiber-based devices are useful for temperatures in the range 100–4000 °C. The use of this technique is reported by Dils (1983), Saaski and Hartl (1992), Sun (1992), Ewan (1998), Zhang et al. (1992), Grattan and Zhang (1995), and Krohn (2000).

14.5.2 Thermal Imaging

Infrared radiation principles can be used to measure the spatial distribution of temperature on a target surface and this is known as thermal imaging or thermography. Applications of thermal imaging are extensive ranging from plant condition monitoring, diagnostics, materials testing, process control, energy auditing for structures, and surveillance. The developments in semiconductor technology have resulted in the wider availability of thermal imaging devices, including reasonably priced items for process control. Thermal imagers typically comprise an optical system, a detector, processing electronics, and a display and are extensions of infrared thermometer technologies combined with some form of scanning optics. Thermal imagers do not require any form of additional illumination in order to operate and this makes them highly attractive in surveillance applications. It should be noted that military and surveillance thermal imagers tend to be configured to produce an image rather than quantitative information on the distribution of temperatures.

It is possible to make use of a single detector with some form of scanner to transmit the radiation signal from specific regions of the optical system to enable a two-dimensional image of the temperature distribution to be built up. The disadvantage of single detector scanners is the trade-off between the speed of response of the instrument and the signal-

to-noise ratio of the detector. The detector typically needs to be cooled and operated at performance limits in order to achieve the desired time response. Multidetector scanners comprising a linear detector array or a two-dimensional array of detectors, known as a staring array enable the temporal to spatial burden to be reduced. Cooling of the detectors can be achieved in a number of ways such as use of liquid nitrogen, Peltier coolers, or a miniature Sterling engine.

The optimum waveband for a thermal imager as for most other infrared thermometers is dictated by the wavelength distribution of the emitted radiation, the transmission characteristics of the atmospheric environment between the imager and the target and by the characteristics of the available detector technology. The following optical bands for air are defined as

- short wavelength infrared imaging band (SWIR), 1.1 to 2.5 μm ;
- medium wavelength infrared imaging band (MWIR), 2.5 to 7 μm , with a notch at 4.2 μm due to CO_2 absorption;
- long wavelength imaging band (LWIR), 5 to 15 μm ;
- further bands beyond 15 μm are classed as far infrared (FIR) and very long wave infrared (VLWIR).

The emissivity of most naturally occurring objects and organic paints is high in the long wave infrared but is lower and more variable in the medium wave infrared. Metallic surfaces tend to have lower emissivity in either band. As such, the use of a thermal imager to provide quantitative information for the temperature distribution, particularly for a surface comprising different materials has to be carefully managed. Without correction for local emissivity values the thermal imager will assume a default value and apply this to the whole image. However, basing the choice of thermal imager on emissivity alone ignores attenuation in the transmission medium. Fog particles, for example, attenuate radiation by scattering with the magnitude dependent on the particulate size. If moderately sized particulates are present, scattering affects the MWIR region more than the LWIR, and a LWIR system can generally provide better range performance.

A number of performance parameters can be used for thermal imagers, allowing different systems to be compared (see also Holst, 1995, 2000; Runciman, 1999).

- Thermal sensitivity (NEDT, noise equivalent differential temperature) refers to the smallest temperature differential that can be detected and depends on the optical system, the responsivity of the detector, and the noise of the system.
- Spatial resolution defines the smallest quantity that can be discerned and is often quantified by Airy disc size.
- Minimum resolvable temperature (MRT).
- Minimum detectable temperature (MDT).

The price of a thermal imager usually reflects the performance, ruggedness, and image processing capability. Some compact imagers can be readily handheld while other systems are designed to be mounted on a platform weighing as much as 100 kg. The availability of uncooled and Sterling engine-cooled detector arrays has made battery-powered portable devices viable. The uncertainty associated with the temperature measurement is specific to the device but typical figures are $\pm 2\text{ K}$ or $\pm 2\%$ full-scale output. Modern

thermal imagers can be particularly easy to use needing simply to be aimed at the target of interest and the image captured by pressing a button.

14.6 CONCLUSIONS

A wide variety of methods are available for measuring temperature. Despite the many types of temperature measuring systems, the choice for a given application is typically limited by a number of factors including uncertainty, temperature range, thermal disturbance, level of contact, size of the sensor, transient response, sensor protection, availability, and cost. This chapter has provided a review of a range of invasive, semi-invasive, and noninvasive instrumentation along with an introduction to the measurement process, calibration, and traceability. Some applications permit the installation of a sensor on or within the medium of interest. Alternatively, it may be desirable to observe the target remotely using an infrared pyrometer, a thermal imager, or other form of noninvasive method.

NOMENCLATURE

a_{CTS}	charge transfer state rate
a_j	probability rate
$B(T)$	second virial coefficient (cm^3/mol)
$C(T)$	third virial coefficient (cm^6/mol^2)
C_1	first radiation constant ($3.7417749 \times 10^{-16} \text{ W m}^2$, Cohen and Taylor, 1987)
C_2	second radiation constant (0.01438769 m K , Cohen and Taylor, 1987)
$D(T)$	fourth virial coefficient (cm^9/mol^3)
E	thermoelectric emf (μV); energy (J)
$E_{\lambda,b}$	spectral emissive power for a blackbody (W/m^3)
E_b	emissive power for a blackbody (W/m^2)
I_F	forward current (A)
I_S	junction's reverse saturation current (A)
I_T	intensity
k	Boltzmann's constant ($1.380658 \times 10^{-23} \text{ J/K}$, Cohen and Taylor, 1987)
n	number of moles of the gas
p	pressure (N/m^2)
q	charge of electron ($1.6 \times 10^{-19} \text{ C}$)
\mathcal{R}	universal gas constant (8.314510 J/mol K , Cohen and Taylor, 1987)
R_0	resistance at 0°C (Ω)
R_t	resistance at temperature t (Ω)
R_T	resistance at temperature T (Ω)
R_{100}	resistance at 100°C (Ω)
S	Seebeck coefficient ($\mu\text{V/K}$)
S_A	Seebeck coefficient for material A ($\mu\text{V/K}$)

S_B	Seebeck coefficient for material B ($\mu\text{V/K}$)
S_{AB}	relative Seebeck coefficient ($\mu\text{V/K}$)
S_{APt}	Seebeck coefficient for material A relative to platinum ($\mu\text{V/K}$)
S_{BPt}	Seebeck coefficient for material B relative to platinum ($\mu\text{V/K}$)
t	temperature ($^{\circ}\text{C}$)
T	temperature (K)
$T_{^{\circ}F}$	the temperature in degree Fahrenheit ($^{\circ}\text{F}$)
$T_{^{\circ}R}$	the Rankine temperature ($^{\circ}\text{R}$)
T_0	temperature at a reference condition (K)
V	volume (m^3)
V_F	forward voltage (V)
α	temperature coefficient of resistance ($\Omega\Omega^{-1}^{\circ}\text{C}^{-1}$ or $^{\circ}\text{C}^{-1}$)
δ	constant
ε	emissivity
λ	wavelength (m)
σ	Stefan–Boltzmann constant and is equal to $5.67051 \times 10^{-8} \text{W/m}^2 \text{K}^4$ (Cohen and Taylor, 1987)
ΔE	absolute Seebeck emf (μV)
Δf	frequency interval (Hz)
ΔT	temperature difference across a segment of a conductor (K),
ANSI	American National Standards Institute
ASTM	American Society for Testing of Materials
BS	British Standard
FIR	far infrared
IPRT	industrial platinum resistance thermometer
IR	infrared
ITS-90	International Temperature Scale of 1990
MDT	minimum detectable temperature
MRT	minimum resolvable temperature
MWIR	medium wavelength infrared
NEDT	noise equivalent differential temperature
NTC	negative temperature coefficient
PRT	platinum resistance thermometer
PSP	parameter-sensitive paint
PTC	positive temperature coefficient
RTD	resistance temperature detector
SPRT	standard platinum resistance thermometer
SWIR	short wavelength infrared
TSP	temperature-sensitive paints
UV	ultra violet
VLWIR	very long wave infrared

REFERENCES

- Allison SW, Gillies GT. Remote thermometry with thermographic phosphors: instrumentation and applications. *The Review of Scientific Instruments* 1997;68:2615–2650.
- ASME PTC, 19.3. Performance test codes supplement on instruments and apparatus part 3 temperature measurement. *ASME* 1974.
- ASTM E1137. Standard specification for industrial platinum resistance thermometers;1997.
- ASTM, E230-98. Standard specification and temperature-electromotive force (emf) tables for standardised thermocouples. *American Society for Testing and Materials*;1999.
- ASTM, American Society for Testing and Materials. Manual on the use of thermocouples in temperature measurement. 4th Edition. ASTM PCN 28-012093-04 1993.
- Barnard RD. *Thermoelectricity in Metals and Alloys*. Taylor and Francis; 1972.
- Baughn JW. Review—liquid crystal methods for studying turbulent heat transfer. *International Journal of Heat and Fluid Flow* 1995;16:365–375.
- Beakley WR. The design of thermistor thermometers with linear calibration. *Journal of Scientific Instruments* 1951;28:176–179.
- Beerman HP. Calibration of pyrometric cones. *Journal of the American Ceramic Society* 1956;39:47–54.
- Bentley RE. Handbook of Temperature Measurement. Vol. 3, *Theory and Practice of Thermoelectric Thermometry*. Springer; 1998.
- Bird C, Mutton JE, Shepherd R, Smith MDW, Watson HML. *Surface temperature measurement in turbines*. pp. 21-1 to 21-10 in AGARD CP 598, 1998.
- Blalock TV, Shepard RL. A decade of progress in high temperature Johnson noise thermometry. In: Schooley JF, editor. *Temperature. Its Measurement and Control in Science and Industry*. Vol. 5(2): American Institute of Physics; New York ; 1982. p. 1219–1223.
- Blatt FJ, Schroeder PA, Foiles CL, Greig D. *Thermoelectric Power in Metals*. Plenum; 1976.
- Bourassa RR, Wang SY, Lengeler B. Energy dependence of the Fermi surface and thermoelectric power of the noble metals. *Physics Review B* 1978;18:1533–1536.
- BS 1041 Part 7. 1988. Temperature measurement. Guide to the selection and use of temperature/time indicators.
- BS 1041: Part 3. 1989. Temperature measurement. Guide to the selection and use of industrial resistance thermometers.
- BS 1041: Section 2.1: 1985. British standard temperature measurement. Part 2: Expansion thermometers. Section 14.2.1 Guide to selection and use of liquid-in-glass thermometers.
- BS EN 60584: 1996. Thermocouples. Part 1: Reference tables. Part 2: Tolerances.
- BS EN 60751. 1996. Industrial platinum resistance thermometer sensors.
- Callendar HL. Notes on platinum thermometry. *Philosophical Magazine* 1899;47:191.
- Callendar HL. On construction of platinum thermometers. *Philosophical Magazine* 1891;34 (104).
- Callendar HL. On the practical measurement of temperature. Experiments made at the Cavendish laboratory, Cambridge. *Philosophical Transactions of the Royal Society of London* 1887;178:161.
- Campbell RP, Molezzi MJ. Applications of advanced liquid crystal video thermography to turbine cooling passage heat transfer measurement. *ASME* 1996; Paper 96-GT-225.
- Childs PRN. *Practical Temperature Measurement*. Butterworth: Heinemann; 2001a.
- Childs PRN. Advances in Temperature Measurement. *Advances in Heat Transfer* 2001b;36: 111–181 AIP.
- Childs PRN, Greenwood JR, Long CA. Heat flux measurement techniques. *Proceedings of I. Mechanical E., Journal of Mechanical Engineering Science*. 1999;213:655–677.

- Claggett TJ, Worrall RW, Liptak BG. Thermocouples. In: Liptak BG, editor. *Instrument Engineers' Handbook Process Measurement and Analysis*, 3rd ed. Chilton: 1995.
- Cohen ER, Taylor BN. The 1986 adjustment of the fundamental physical constants. *Reviews of Modern Physics* 1987;59:1121.
- Cohen ER, Taylor BN. The fundamental physical constants. *Physics Today*, BG5-BG9 1999.
- Connolly JJ. Resistance thermometer measurement, In: Bentley RE, edited. *Handbook of Temperature Measurement, 2, Resistance and Liquid in Glass Thermometry*. Springer; 1998.
- Coren D, Turner J, Eastwood D, Davies S, Atkins N, Childs PRN, Dixon J, Scanlon T. An advanced multi-configuration turbine stator well cooling test facility. *ASME Paper GT2010-23450* 2010.
- Dils RR. High temperature optical fiber thermometer. *Journal of Applied Physics* 1983;54: 1198–1201.
- Doebelin EO. *Measurement Systems: Application and Design*, 4th ed. McGraw Hill; 1990.
- ESDU. Temperature measurement. Resistance thermometry. ESDU 06019. ESDU Heat Transfer Series Volume 4 (Insulation and Temperature Measurement) ISBN: 978 186 246 595 4 ISSN: 0141-402X, 2007b.
- ESDU. Temperature measurement. Techniques. ESDU 02006. ESDU Heat Transfer Series Volume 4 (Insulation and Temperature Measurement) ISBN: 978 1 86246 196 3, ISSN: 0141-402X, 2002.
- ESDU. Temperature measurement. Thermocouples. ESDU 06018. ESDU Heat Transfer Series Volume 4 (Insulation and Temperature Measurement) ISBN: 978 186 246 594 7, ISSN: 0141-402X, 2007a.
- Ewan BCR. A study of two optical fiber probe designs for use in high temperature combustion gases. *Measurement Science and Technology* 1998;9:1330–1335.
- Fairchild CO, Peters MF. Characteristics of pyrometric cones. *Journal of the American Ceramic Society* 1926;9:701–743.
- Feist JP, Heyes AL, Nicholls JR. Phosphor thermometry in an EBPVD produced TBC doped with dysprosium. *Proceedings of Instrumentation and Mechanical Engineering*. 2001;215: Part G 333–341.
- Feist JP, Heyes AL, Choy KL, Nicholls JR. Thermographic phosphor thermometry: recent development for applications in gas turbines. In: Greated C, Cosgrove J, Buick JM, editors. *Optical Methods for Data Processing in Heat and Fluid Flow*; 2002.
- Figliola RS, Beasley DE. *Theory and Design for Mechanical Measurements*. 4th ed. John Wiley & Sons; 2005.
- Fonger WH, Struck CW. Eu^{+35}D resonance quenching to the charge-transfer states in $\text{Y}_2\text{O}_3\text{S}$, $\text{La}_2\text{O}_3\text{S}$ and LaOCl . *Journal of Chemical Physics* 1970;52:6364–6372.
- Gallery J, Gouterman M, Callis J, Khalil G, McLachlan B, Bell J. Luminescent thermometry for aerodynamic measurements, *Review of Scientific Instruments* 1994;65:712–720.
- Goodfellow Catalogue 2000 /2001. Goodfellow Cambridge Ltd.
- Grattan KTV, Zhang ZY. *Fibre Optic Fluorescence Thermometry*. Chapman & Hall; 1995.
- Hashemian HM, Petersen KM. Achievable accuracy and stability of industrial RTDs. In: Schooley JF, editor. *Temperature. Its Measurement and Control in Science and Industry*. American Institute of Physics; 1992. Vol. 6: p. 427–432.
- Heyes AL. Thermographic phosphor thermometry for gas turbines. In: Sieverding CH, Brouckaert JF, editors. *Advanced Measurement Techniques for Aero and Stationary Gas Turbines*. VKI LS 2004-04; 2004.
- Heyes AL. On the design of phosphors for high temperature thermometry. *Journal of Luminescence* 2009;129:2004–2009.
- HITRAN. <http://www.cfa.harvard.edu/hitran/>. (date of access 25th June 2010).

- Holst GC. *Common Sense Approach to Thermal Imaging*. JCD Publishing/SPIE Press; 2000.
- Holst GC. *Electro-Optical Imaging System Performance*. JCD Publishing; 1995.
- Huston WD. The accuracy and reliability of bimetallic temperature measuring elements. In: Herzfeld CH, editor. *Temperature. Its Measurement and Control in Science and Industry*. Vol. 3, Reinhold; 1962. p. 949–957.
- IEC, 584-1. International Standard. Thermocouples—Part 1: Reference tables. Second Edition 1995-09, International Electrotechnical Commission, 1995.
- IEC-751. 1983. Standard for industrial resistance thermometers. International Electrotechnical Commission.
- Johnson JB. Thermal agitation of electricity in conductors. *Physical Review* 1928;32:97–109.
- Jones TV. The thin film heat transfer gauge—a history and new developments. Proceedings 4th UK Conference on Heat Transfer, Paper C510/150/95, p. 1–12, 1995.
- Kakade VU, Lock GD, Wilson M, Owen JM, Mayhew JE. Accurate heat transfer measurements using thermochromic liquid crystal. Part 1: calibration and characteristics of crystals. *International Journal of Heat and Fluid Flow* 2009a;30:939–949.
- Kakade VU, Lock GD, Wilson M, Owen JM, Mayhew JE. Accurate heat transfer measurements using thermochromic liquid crystal. Part 2: application to a rotating disc. *International Journal of Heat and Fluid Flow* 2009b;30:950–959.
- Kamper RA. Survey of noise thermometry. In: Plumb HH, editor. *Temperature. Its Measurement and Control in Science and Industry*. 4(1): Instrument Society of America; 1972. p. 349–354.
- Kaplan H. *Practical Applications of Infrared Thermal Sensing and Imaging Equipment*. 2nd ed. Vol. TT34, SPIE Press; 1999.
- Kinzie PA. *Thermocouple Temperature Measurement*. Wiley-Interscience; 1973.
- Krause JK, Dodrill BC. Measurement system induced errors in diode thermometry. *Review of Social Economy* 1986;57:661–665.
- Krohn DA. *Fiber Optic Sensors*. ISA; 2000.
- Lee SJ, Yoon JH. Temperature field measurement of heated ventilation flow in a vehicle interior. *International Journal of Vehicle Design* 1998;19:228–243.
- Mangum BW, Furukawa GT. Guidelines for realizing the ITS-90. *NIST Technical note 1265*; 1990.
- McGee TD. *Principles and Methods of Temperature Measurement*. Wiley; 1988.
- Meijer GCM, van Herwaarden AW. Thermal sensors. *IOP* 1994.
- Mott NF, Jones H. *Theory and Properties of Metals and Alloys*. Dover; 1958.
- Neumann RD. Aerothermodynamic instrumentation, AGARD Report No. 761, *Special Course on Aerothermodynamics of Hypersonic Vehicles* 1989;(1–40):4.
- Nicholas JV, White DR. *Traceable Temperatures, An Introduction to Temperature Measurement and Calibration*. John Wiley & Sons; 1994.
- Nicholas JV. Liquid-in-glass thermometers. Section 32.8 In: Webster JG, editor. *The Measurement Instrumentation and Sensors Handbook*. CRC Press; 1999.
- Nyquist H. Thermal agitation of electric charge in conductors, *Physical Review* 1928;32:110–113.
- Ou S, Rivir R, Meininger M, Soechting F, Tabbita M. Transient liquid crystal measurement of leading edge film cooling effectiveness and heat transfer with high free stream turbulence. *ASME Paper* 2000-GT-0245, 2000.
- Pavese F. Manometric thermometers. Section 32.9 In: Webster JG, editor. *The Measurement Instrumentation and Sensors Handbook*. CRC Press; 1999.
- Pavese F, Molinar G. *Modern Gas-Based Temperature and Pressure Measurements*. Plenum Press; 1992.

- Pavese F, Steur PPM. ^3He constant-volume gas thermometry: calculations for a temperature scale between 0.8 K and 25 K. *Journal of Low Temperature Physics* 1987;69:91–117.
- Pollock DD. *Thermocouples: Theory and Properties*. CRC Press; 1991.
- Preston-Thomas H. The international temperature scale of 1990 (ITS-90). *Metrologia* 1990;27:3–10.
- Roberts GT, East RA. Liquid crystal thermography for heat transfer measurement in hypersonic flows: a review. *Journal of Spacecraft and Rockets* 1996;33:761–768.
- Roeser WF. Thermoelectric thermometry. *Journal of Applied Physics* 1940;11:388–407.
- Rogatto WD, editor. The infrared and electro-optical systems handbook. Vol. 3, *Electro-Optical Components*. MI and SPIE Press; 1993.
- Runciman HM. Thermal imaging. Section 35.1, In: Webster JG, editor. *The Measurement, Instrumentation and Sensors Handbook*. CRC Press; 1999.
- Rusby RL. The rhodium–iron resistance thermometer: Ten years on. In: Schooley JF, editor. *Temperature. Its Measurement and Control in Science and Industry*. 1982;5(2):829–834 American Institute of Physics, New York.
- Saaski EW, Hartl JC. Thin-film Fabry Perot temperature sensors. In: Schooley JF, editor. *Temperature. Its Measurement and Control in Science and Industry*. 1992;6(2):731–734, American Institute of Physics, New York.
- Sapoff M. Thermistors: part 4, optimum linearity techniques. *Measurements and Control*. 1980;14.
- Sapoff M, Oppenheim RM. Theory and application of self-heated thermistors. *Proceedings of the IEEE* 1964;51:1292.
- Schuster G. Temperature measurement with rhodium–iron resistors below 0.5 K. In: Schooley JF, editor. *Temperature. Its Measurement and Control in Science and Industry*. Vol. 6(1): New York: American Institute of Physics; 1992. p. 449–452.
- Seger. *Thorindustrie Zeitung*, 1886, p. 135, 229.
- Selby JEA, McClatchey RA. Atmospheric transmittance from 0.25 to 28.5 μm : computer code LOWTRAN 2. Report AFCRL-72-0745, *Air Force Cambridge Research Laboratory* 1972.
- Selby JEA, McClatchey RA. Atmospheric transmittance from 0.25 to 28.5 μm : computer code LOWTRAN 3. Report AFCRL-75-0255, *Air Force Cambridge Research Laboratory* 1975.
- Selby JEA, McClatchey RA. Atmospheric transmittance from 0.25 to 28.5 μm : computer code LOWTRAN 3B. Report AFCRL-TR-76-0258, *Air Force Cambridge Research Laboratory*. 1976.
- Selby JEA, Kneizys FX, Chetwynd JH, McClatchey RA. Atmospheric transmittance from 0.25 to 28.5 μm : computer code LOWTRAN 4. Report AFGL-TR-78-0053, *Air Force Cambridge Research Laboratory* 1978.
- Siemens WH. On the increase of electrical resistance in conductors with rise of temperature and its application to the measure of ordinary and furnace temperatures; also on a simple method of measuring electrical resistances. *Proceedings of Royal Society, London* 1871;19:443.
- Smith FG, editor. The infrared and electro-optical systems handbook. Vol. 2, *Atmospheric Propagation of Radiation*. MI and SPIE Press; 1993.
- Stephenson RJ, Moulin AM, Welland ME. Bimaterials thermometers. In: Webster JR, editor. *The Measurement Instrumentation and Sensors Handbook*. CRC Press; 1999.
- Sun M. Fibreoptic thermometry based on photoluminescent decay times. In: Schooley JF, editor. *Temperature: Its Measurement and Control in Science and Industry*. Vol. 6: American Institute of Physics; 1992. p. 715–719.
- Timoshenko SP. *The Collected Papers*. McGraw Hill; 1953.
- Touloukian YS, DeWitt DP. Thermophysical properties of matter, Vol. 7, *Thermal Radiative Properties. Metallic Elements and Alloys*. IFI/Plenum; 1970.

- Touloukian YS, DeWitt DP. Thermophysical properties of matter, Vol. 8, *Thermal Radiative Properties. Nonmetallic Solids*. IFI/Plenum; 1972a.
- Touloukian YS, DeWitt DP, Hecnitz RS. Thermophysical properties of matter, Vol. 9, *Thermal Radiative Properties. Coatings*. IFI/Plenum; 1972b.
- Ween S. Care and use of liquid-in-glass thermometers. *ISA Transactions* 1968;7:94.
- White DR, Galleano R, Actis A, Bixy H, DeGroot M, Dumbledam J, Reesink AL, Elder F, Sakurai H, Shepard RL, Gallop JC. The status of Johnson noise thermometry. *Metrologia* 1996;33:325–335.
- Wise JA. Liquid in glass thermometry. *NBS Monograph* 1976; 150.
- Wolfendale PCF, Yewen JD, Daykin CI. A new range of high precision resistance bridges for resistance thermometry. In: Schooley JF, editor. *Temperature. Its Measurement and Control in Science and Industry*. Vol. 5(2): American Institute of Physics; New York: 1982. p. 729–732.
- Wood SD, Mangum BW, Filliben JJ, Tillett SB. An investigation of the stability of thermistors. *Journal Research of the NBS* 1978;247–263.
- Yates HW, Taylor JH. NRL Report 5453 US Naval Research Laboratory, Washington DC: 1960.
- Zhang Z, Grattan KTV, Palmer AW. Fibre optic high temperature sensor based on the fluorescence lifetime of alexandrite. *Review of Scientific Instruments* 1992;63:3869–3873.

15

PRESSURE AND VELOCITY MEASUREMENTS

RICHARD S. FIGLIOLA AND DONALD E. BEASLEY

- 15.1 Pressure concepts
- 15.2 Pressure reference instruments
 - 15.2.1 McLeod gauge
 - 15.2.2 Barometer
 - 15.2.3 Manometer
 - 15.2.4 Deadweight testers
- 15.3 Pressure transducers
 - 15.3.1 Bourdon tube
 - 15.3.2 Bellows and capsule elements
 - 15.3.3 Diaphragms
 - 15.3.4 Strain gauge elements
 - 15.3.5 Capacitance elements
 - 15.3.6 Piezoelectric crystal elements
- 15.4 Pressure transducer calibration
 - 15.4.1 Static calibration
 - 15.4.2 Dynamic calibration
- 15.5 Pressure measurements in moving fluids
 - 15.5.1 Total pressure measurement
 - 15.5.2 Static pressure measurement
- 15.6 Modeling pressure and fluid systems
- 15.7 Design and installation: transmission effects
 - 15.7.1 Liquids
 - 15.7.2 Gases
 - 15.7.3 Heavily damped systems
- 15.8 Fluid velocity measuring systems
 - 15.8.1 Pitot-static pressure probe
 - 15.8.2 Thermal anemometry
 - 15.8.3 Doppler anemometry
 - 15.8.4 Particle image velocimetry
 - 15.8.5 Selection of velocity measuring methods
 - 15.8.6 Pitot-static pressure methods

- 15.8.7 Thermal anemometer
- 15.8.8 Laser doppler anemometer
- 15.8.9 Particle image velocimetry

Nomenclature

References

15.1 PRESSURE CONCEPTS

Pressure represents a contact force per unit area. It acts inwardly, and normally to a surface. To better understand the origin and nature of pressure, consider the measurement of pressure at the wall of a vessel containing an ideal gas. As a gas molecule with some amount of momentum collides with this solid boundary, it rebounds off in a different direction. From Newton's second law, we know that this change in linear momentum of the molecule produces an equal but opposite (normal, inward) force on the boundary. It is the net effect of these collisions averaged over brief instants in time that yields the pressure sensed at the boundary surface. Because there are so many molecules per unit volume involved (e.g., in a gas there are roughly 10^{16} molecules per mm^3), pressure is usually considered to be continuous. Factors that affect the frequency or the number of the collisions, such as temperature and fluid density, affect the pressure. In fact, this reasoning is the basis of the kinetic theory from which the ideal gas equation of state is derived.

A pressure scale must be related to molecular activity, since a lack of any molecular activity must form the limit of absolute zero pressure. A pure vacuum, which contains no molecules, provides the limit for a primary standard for absolute zero pressure. As shown in Figure 15.1, the absolute pressure scale is quantified relative to this absolute

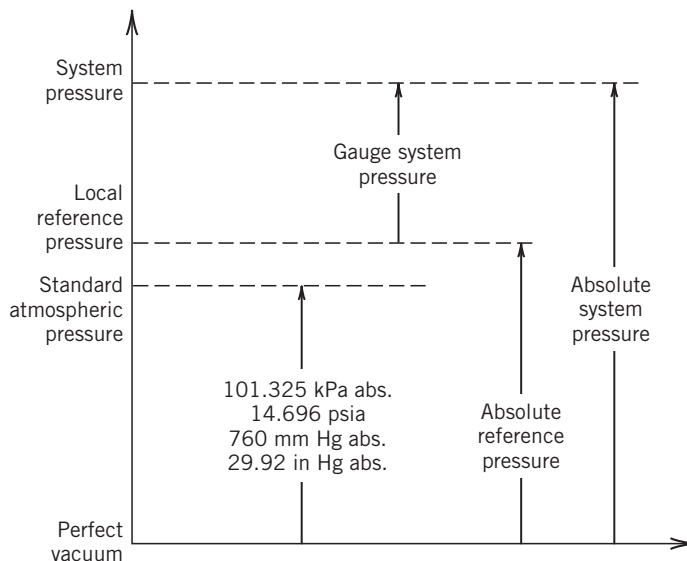


FIGURE 15.1 Relative pressure scales.

zero pressure. The pressure under standard atmospheric conditions is defined as 1.01320×10^5 Pa absolute (where $1 \text{ Pa} = 1 \text{ N/m}^2$) (Brombacher et al.,). This is equivalent to

- 101.32 kPa absolute
- 1 atm absolute
- 14.696 lb/in.² absolute (written as psia)
- 1.013 bar absolute (where 1 bar = 100 kPa)

The term “absolute” can be abbreviated as “a” or “abs.”

Also indicated in Figure 15.1 is a gauge pressure scale. The gauge pressure scale is measured relative to some absolute reference pressure, which is defined in a manner convenient to the measurement. The relation between an absolute pressure, p_{abs} , and its corresponding gauge pressure, p_{gauge} , is given by

$$p_{\text{gauge}} = p_{\text{abs}} - p_0 \quad (15.1)$$

where p_0 is a reference pressure. A commonly used reference pressure is the local absolute atmospheric pressure. Absolute pressure is a positive number. Gauge pressure can be positive or negative depending on the value of measured pressure relative to the reference pressure. A differential pressure, such as $p_1 - p_2$, is a relative measure of pressure.

Pressure can also be described in terms of the pressure exerted on a surface that is submerged in a column of fluid at depth h , as depicted in Figure 15.2. From hydrostatics, the pressure at any depth within a fluid of specific weight γ can be written as

$$p_{\text{abs}}(h) = p(h_0) + \gamma h = p_0 + \gamma h \quad (15.2)$$

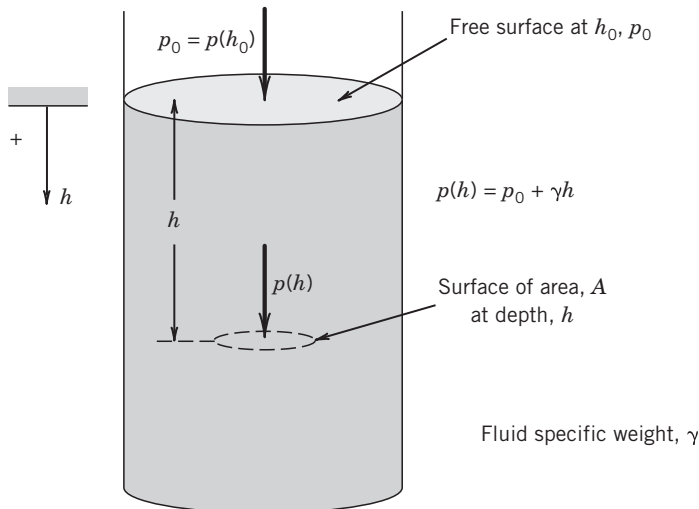


FIGURE 15.2 Hydrostatic-equivalent pressure head and pressure.

where p_0 is the pressure at an arbitrary datum line at h_0 , and h is measured relative to h_0 . The fluid specific weight is given by $\gamma = \rho g$ where ρ is the density. When Equation (15.2) is rearranged, the equivalent head of fluid at depth h becomes

$$h = [p_{\text{abs}}(h) - p(h)]/\gamma = (p_{\text{abs}} - p_0)/\gamma \quad (15.3)$$

The equivalent pressure head at one standard atmosphere ($p_0 = 0$ absolute) is

$$\begin{aligned} 760 \text{ mm Hg abs} &= 760 \text{ torr abs} = 1 \text{ atm abs} \\ &= 10,350.8 \text{ mm H}_2\text{O abs} = 29.92 \text{ in Hg abs} \\ &= 407.513 \text{ in H}_2\text{O abs} \end{aligned}$$

The standard is based on mercury (Hg) with a density of $0.0135951 \text{ kg/cm}^3$ at 0°C and water at $0.000998207 \text{ kg/cm}^3$ at 20°C (Brombacher et al.,).

15.2 PRESSURE REFERENCE INSTRUMENTS

The units of pressure can be defined through the standards of the fundamental dimensions of mass, length, and time. In practice, pressure transducers are calibrated by comparison against certain reference instruments, which also serve as pressure measuring instruments. This section discusses several basic pressure reference instruments that can serve either as working standards or as laboratory instruments.

15.2.1 McLeod Gauge

The McLeod gauge, originally devised by Herbert McLeod in 1874 (McLeod, 1874), is a pressure-measuring instrument and laboratory reference standard used to establish gas pressures in the subatmospheric range of 1 mmHg abs down to 0.1 mmHg abs. A pressure that is below atmospheric pressure is also called a vacuum pressure. One variation of this instrument is sketched in Figure 15.3a, in which the gauge is connected directly to the low-pressure source. The glass tubing is arranged so that a sample of the gas at an unknown low pressure can be trapped by inverting the gauge from the sensing position, depicted as Figure 15.3a, to that of the measuring position, depicted as Figure 15.3b. In this way, the gas trapped within the capillary is isothermally compressed by a rising column of mercury. Boyle's law is then used to relate the two pressures on either side of the mercury to the distance of travel of the mercury within the capillary. Mercury is the preferred working fluid because of its high density and very low vapor pressure.

At the equilibrium and measuring position, the capillary pressure, p_2 , is related to the unknown gas pressure to be determined, p_1 , by $p_2 = p_1(\forall_1/\forall_2)$ where \forall_1 is the gas volume of the gauge in Figure 15.3a (a constant for a gauge at any pressure), and \forall_2 is the capillary volume in Figure 15.3b. But $\forall_2 = Ay$, where A is the known cross-sectional area of the capillary and y is the vertical length of the capillary occupied by the gas. With γ as the specific weight of the mercury, the difference in pressures is related by $p_2 - p_1 = \gamma y$ such that the unknown gas pressure is just a function of y :

$$p_1 = \gamma Ay^2/(\forall_1 - Ay) \quad (15.4)$$

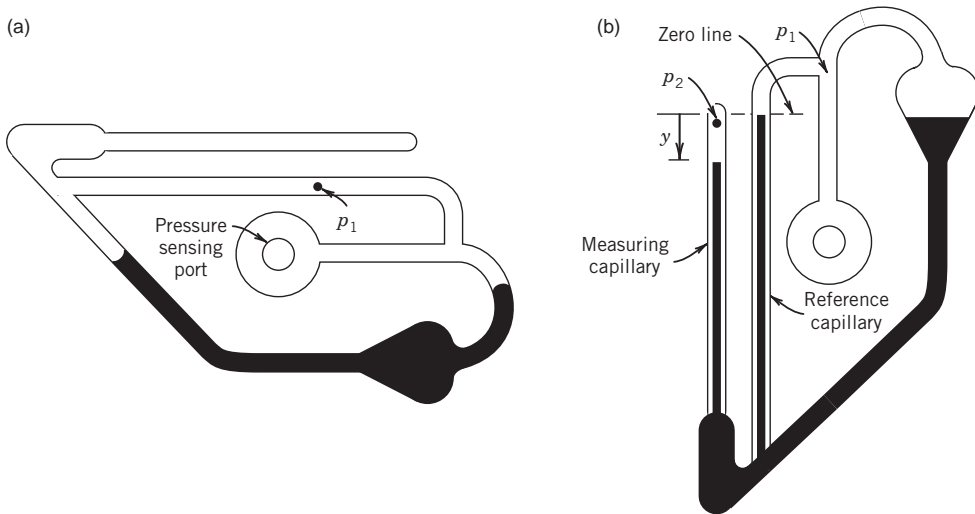


FIGURE 15.3 McLeod gauge: (a) sensing position and (b) Indicating position.

In practice, a commercial McLeod gauge has the capillary etched and calibrated to indicate either pressure, p_1 , or its equivalent head, p_1/γ , directly. The McLeod gauge generally does not require correction. The reference stem offsets capillary forces acting in the measuring capillary. Instrument systematic uncertainty is on the order of 0.5% (95%) at 1 mm Hg abs and increases to 3% (95%) at 0.1 mm Hg abs.

15.2.2 Barometer

A barometer consists of an inverted tube containing a fluid and is used to measure atmospheric pressure. To create the barometer, the tube, which is sealed at only one end, is evacuated to zero absolute pressure. The tube is immersed with the open end down within a liquid-filled reservoir as shown in the illustration of the Fortin barometer in Figure 15.4. The reservoir is open to atmospheric pressure, which forces the liquid to rise up the tube.

From Equations (15.2) and (15.3), the resulting height of the liquid column above the reservoir free surface is a measure of the absolute atmospheric pressure in the equivalent head (Eq. 15.3). Evangelista Torricelli (1608–1647), a colleague of Galileo, can be credited with developing and interpreting the working principles of the barometer in 1644.

As Figure 15.4 shows, the closed end of the tube is at the vapor pressure of the barometric liquid at room temperature. So the indicated pressure is the atmospheric pressure minus the liquid vapor pressure. Mercury is the most common liquid used because it has a very low vapor pressure, and so, for practical use, the indicated pressure can be taken as the local absolute barometric pressure. However, for very accurate work the barometer needs to be corrected for temperature effects, which change the vapor pressure, for temperature and altitude effects on the weight of mercury, and for deviations from standard

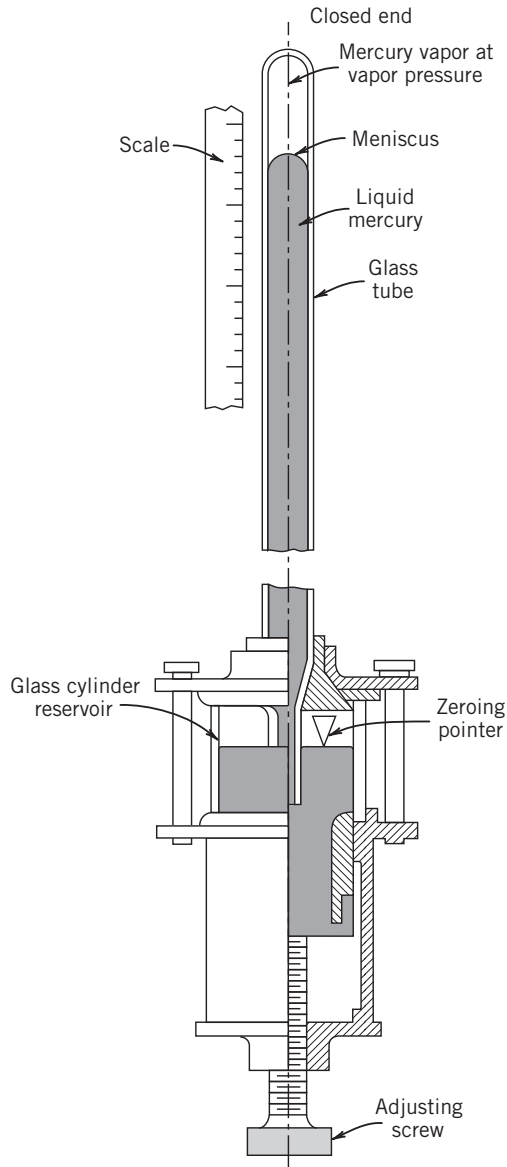


FIGURE 15.4 Fortin barometer.

gravity (9.80665 m/s^2 or 32.17405 ft/s^2). Correction curves are provided by instrument manufacturers.

Barometers are used as local standards for the measurement of absolute atmospheric pressure. Under standard conditions for pressure temperature and gravity, the mercury rises 760 mm (29.92 in.) above the reservoir surface. The U.S. National Weather Service always reports a barometric pressure that has been corrected to sea-level elevation.

15.2.3 Manometer

A manometer is an instrument used to measure differential pressure based on the relationship between pressure and the hydrostatic equivalent head of fluid. Several design variations are available, allowing measurements ranging from the order of 0.001 mm of manometer fluid to several meters.

The U-tube manometer in Figure 15.5 consists of a transparent tube filled with an indicating liquid of specific weight γ_m . This forms two free surfaces of the manometer liquid. The difference in pressures p_1 and p_2 applied across the two free surfaces brings about a deflection, H , in the level of the manometer liquid. For a measured fluid of specific weight γ , the hydrostatic equation can be applied to the manometer of Figure 15.5 as

$$p_1 = p_2 + \gamma x + \gamma_m H - \gamma(H + x)$$

which yields the relation between the manometer deflection and applied differential pressure,

$$p_1 - p_2 = (\gamma_m - \gamma)H \quad (15.5)$$

From Equation (15.5), the static sensitivity of the U-tube manometer is given by $K = 1/(\gamma_m - \gamma)$. To maximize manometer sensitivity, we want to choose manometer liquids that minimize the value of $(\gamma_m - \gamma)$. From a practical standpoint, the manometer fluid must not be soluble with the working fluid. The manometer fluid should be selected to provide a deflection that is measurable yet not so great that it becomes awkward to observe.

A variation in the U-tube manometer is the micromanometer shown in Figure 15.6. These special-purpose instruments are used to measure very small differential pressures, down to 0.005 mm H₂O (0.0002 in. H₂O). In the micromanometer, the manometer

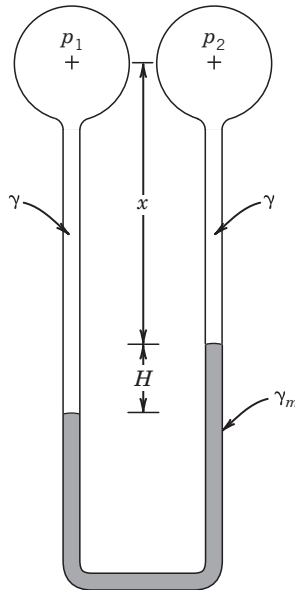


FIGURE 15.5 U-tube manometer.

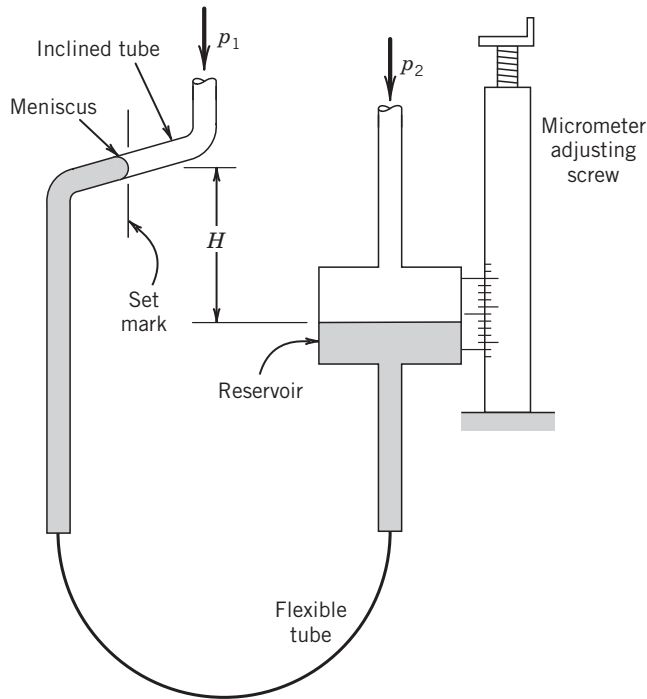


FIGURE 15.6 Micromanometer.

reservoir is moved up or down until the level of the manometer fluid within the reservoir is at the same level as a set mark within a magnifying sight glass. At that point the manometer meniscus is at the set mark, and this serves as a reference position. Changes in pressure bring about fluid displacement so that the reservoir must be moved up or down to bring the meniscus back to the set mark. The amount of this repositioning is equal to the change in the equivalent pressure head. The position of the reservoir is controlled by a micrometer or other calibrated displacement measuring device so that relative changes in pressure can be measured with high resolution.

The inclined tube manometer is also used to measure small changes in pressure. It is essentially a U-tube manometer with one leg inclined at an angle θ , typically from 10° to 30° relative to the horizontal. As indicated in Figure 15.7, a change in pressure equivalent to a deflection of height H in a U-tube manometer would bring about a change in position of the meniscus in the inclined leg of $L = H/\sin \theta$. This provides an increased sensitivity over the conventional U-tube by the factor of $1/\sin \theta$.

A number of elemental errors affect the instrument uncertainty of all types of manometers. These include scale and alignment errors, zero error, temperature error, gravity error, and capillary, and meniscus errors. The specific weight of the manometer fluid varies with temperature but can be corrected. For example, the common manometer fluid of mercury has a temperature dependence approximated by

$$\gamma_{\text{Hg}} \frac{133.084}{1 + 0.000067T} (\text{N/m}^3) = \frac{848.707}{1 + 0.000101(T - 32)} (\text{lb/ft}^3)$$

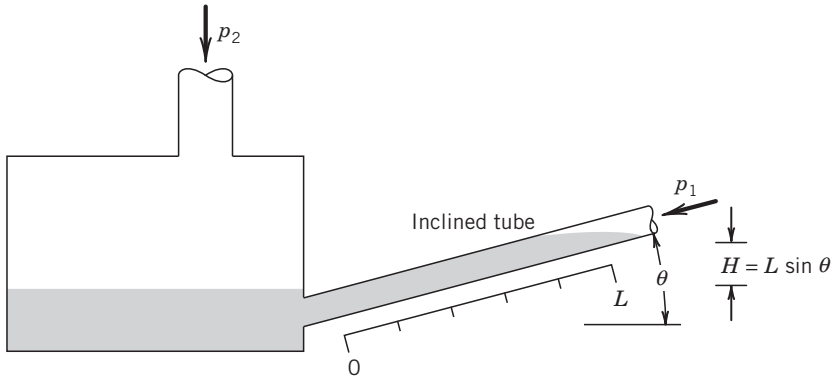


FIGURE 15.7 Inclined tube manometer.

with T in $^{\circ}\text{C}$ or $^{\circ}\text{F}$, respectively. A gravity correction for elevation z and latitude ϕ corrects for gravity error effects using the dimensionless correction,

$$e_1 = -(2.637 \times 10^{-3} \cos 2\phi + 9.6 \times 10^{-8} z + 5 \times 10^{-5})_{\text{US}} \quad (15.6a)$$

$$= -(2.637 \times 10^{-3} \cos 2\phi + 2.9 \times 10^{-8} z + 5 \times 10^{-5})_{\text{metric}} \quad (15.6b)$$

where ϕ is in degrees and z is in feet for Equation (15.6a) and meters in Equation (15.6b). Tube-to-liquid capillary forces lead to the development of a meniscus. Although the actual effect varies with purity of the manometer liquid, these effects can be minimized by using manometer tube bores of greater than about 6 mm (0.25 in.). In general, the instrument uncertainty in measuring pressure can be as low as 0.02–0.2% of the reading.

15.2.4 Deadweight Testers

The deadweight tester makes direct use of the fundamental definition of pressure as a force per unit area to create and to determine the pressure within a sealed chamber. These devices are a common laboratory standard for the calibration of pressure-measuring devices over the pressure range from 70 to $7 \times 10^7 \text{ N/m}^2$ (0.01–10,000 psi). A deadweight tester, such as that shown in Figure 15.8, consists of an internal chamber filled with an oil and a close-fitting piston and cylinder. Chamber pressure acts on the end of the carefully machined piston. A static equilibrium exists when the external pressure exerted by the piston on the fluid balances the chamber pressure. This external piston pressure is created by a downward force acting over the equivalent area A_e of the piston. The weight of the piston plus the additional weight of calibrated masses are used to provide this external force F . At static equilibrium the piston floats and the chamber pressure can be deduced as

$$p = \frac{F}{A_e} + \sum \text{error corrections} \quad (15.7)$$

A pressure transducer can be connected to a reference port and calibrated by comparison to the chamber pressure. For most calibrations, the error corrections can be ignored.

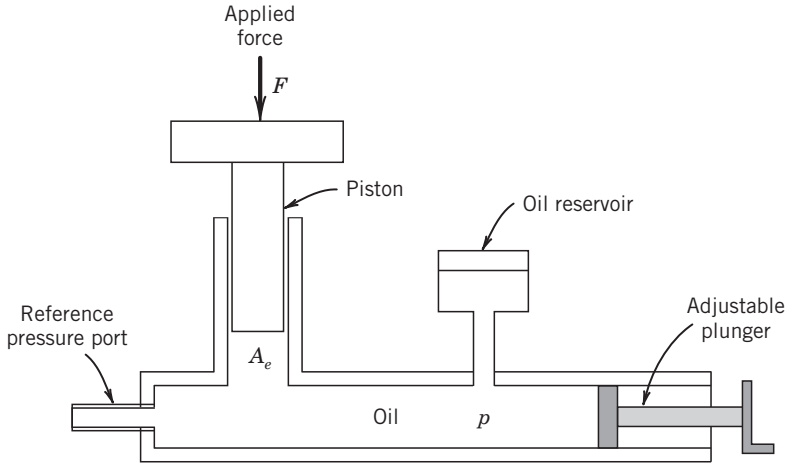


FIGURE 15.8 Deadweight tester.

When error corrections are applied, the instrument uncertainty in the chamber pressure using a deadweight tester can be as low as 0.005–0.01% of the reading. A number of elemental errors contribute to Equation (15.7), including air buoyancy effects, variations in local gravity, uncertainty in the known mass of the piston and added masses, shear effects, thermal expansion of the piston area, and elastic deformation of the piston (American Society of Mechanical Engineers (ASME), 1987).

An indicated pressure, p_i , can be corrected for gravity effects, e_1 , from Equation (15.6a) or (15.6b), and for air buoyancy effects, e_2 , by

$$p = p_i(1 + e_1 + e_2) \quad (15.8)$$

where

$$e_2 = -\gamma_{\text{air}}/\gamma_{\text{masses}} \quad (15.9)$$

The tester fluid lubricates the piston so that the piston is partially supported by the shear forces in the oil in the gap separating the piston and the cylinder. This error varies inversely with the tester fluid viscosity, so high-viscosity fluids are preferred. In a typical tester, the uncertainty due to this error is less than 0.01% of the reading. At high pressures, elastic deformation of the piston affects the actual piston area. For this reason, the effective area is based on the average of the piston and cylinder diameters.

15.3 PRESSURE TRANSDUCERS

A pressure transducer is a device that converts a measured pressure into a mechanical or electrical signal. The transducer is actually a hybrid sensor-transducer. The primary sensor is usually an elastic element that deforms or deflects under the measured pressure relative to a reference pressure. Several common elastic elements used, as shown in Figure 15.9, include the Bourdon tube, bellows, capsule, and diaphragm. A secondary transducer element converts the elastic element deflection into a readily measurable signal

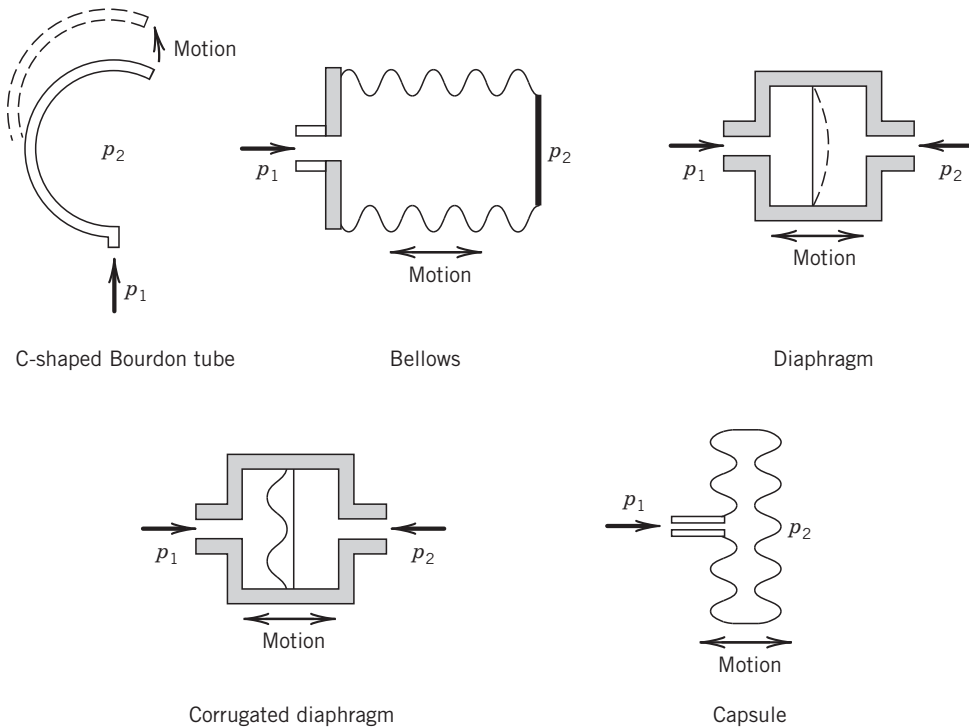


FIGURE 15.9 Elastic elements used as pressure sensors.

such as an electrical voltage or mechanical rotation of a pointer. There are many methods available to perform this transducer function, and we examine a few common ones.

General categories for pressure transducers are absolute, gauge, vacuum, and differential. These categories reflect the application and reference pressure used. Absolute transducers have a sealed reference cavity held at a pressure of absolute zero, enabling absolute pressure measurements. Gauge transducers have the reference cavity open to atmospheric pressure and are intended to measure above or below atmospheric pressure or both. Differential transducers measure the difference between two applied pressures. Vacuum transducers are a special form of absolute transducer for low-pressure measurements.

Pressure transducers are subject to some or all of the following elemental errors: resolution, zero shift error, linearity error, sensitivity error, hysteresis, noise, and drift due to environmental temperature changes. Electrical transducers are also subject to loading error between the transducer output and its indicating device (see Chapter 6). Loading errors increase the transducer nonlinearity over its operating range. When this is a consideration, a voltage follower (see Chapter 6) can be inserted at the output of the transducer to isolate transducer load.

15.3.1 Bourdon Tube

The Bourdon tube is a curved metal tube having an elliptical cross section that mechanically deforms under pressure. In practice, one end of the tube is held fixed and the input pressure is applied internally. A pressure difference between the outside and the inside of

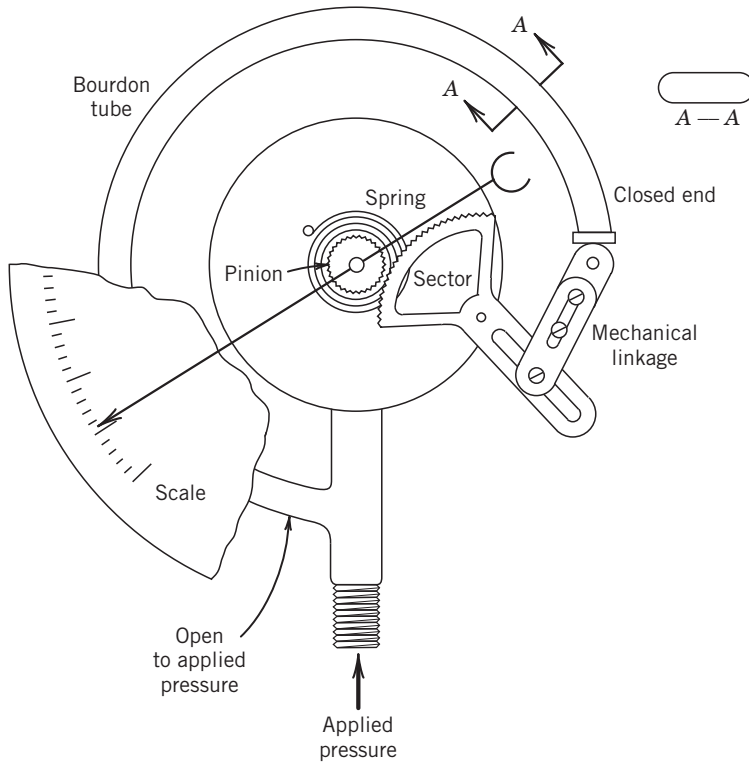


FIGURE 15.10 Bourdon tube pressure gauge.

the tube brings about tube deformation and a deflection of the tube free end. This action of the tube under pressure can be likened to the action of a deflated balloon that is subsequently inflated. The magnitude of the deflection of the tube end is proportional to the magnitude of the pressure difference. Several variations exist, such as the C shape (Figure 15.9), the spiral, and the twisted tube. The exterior of the tube is usually open to atmosphere (hence, the origin of the term “gauge” pressure referring to pressure referenced to atmospheric pressure), but in some variations the tube may be placed within a sealed housing and the tube exterior exposed to some other reference pressure, allowing for absolute and for differential designs.

The Bourdon tube mechanical dial gauge is a commonly used pressure transducer. A typical design is shown in Figure 15.10, in which the secondary element is a mechanical linkage that converts the tube displacement into a rotation of a pointer. Designs exist for low or high pressures, including vacuum pressures, and selections span a wide choice in range. The best Bourdon tube gauges have instrument uncertainties as low as 0.1% of the full-scale deflection of the gauge, with values of 0.5–2% more common. But the attractiveness of this device is that it is simple, portable, and robust, lasting for years of use.

15.3.2 Bellows and Capsule Elements

A bellows sensing element is a thin-walled, flexible metal tube formed into deep convolutions and sealed at one end (Figure 15.9). One end is held fixed and pressure is applied

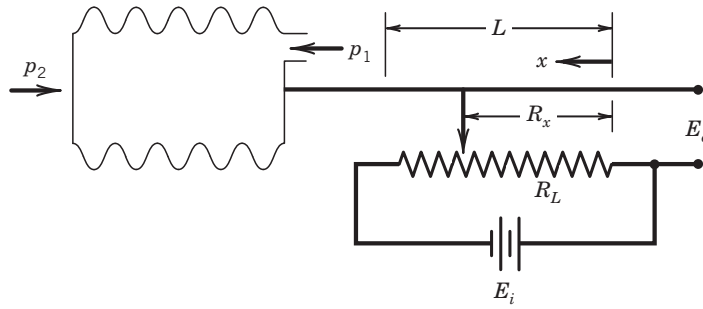


FIGURE 15.11 Potentiometer pressure transducer.

internally. A difference between the internal and external pressures causes the bellows to change in length. The bellows is housed within a chamber that can be sealed and evacuated for absolute measurements, vented through a reference pressure port for differential measurements, or opened to atmosphere for gauge pressure measurements. A similar design, the capsule sensing element, is also a thin-walled, flexible metal tube whose length changes with pressure, but its shape tends to be wider in diameter and shorter in length (Figure 15.9).

A mechanical linkage is used to convert the translational displacement of the bellows or capsule sensors into a measurable form. A common transducer is the sliding arm potentiometer (voltage-divider, Chapter 6) found in the potentiometric pressure transducer shown in Figure 15.11. Another type uses a linear variable displacement transducer (LVDT; see Chapter 12) to measure the bellows or capsule displacement. The LVDT design has a high sensitivity and is commonly found in pressure transducers rated for low pressures and for small pressure ranges, such as zero to several hundred mm Hg absolute, gauge, or differential.

15.3.3 Diaphragms

An effective primary pressure element is a diaphragm (Figure 15.9), which is a thin elastic circular plate supported about its circumference. The action of a diaphragm within a pressure transducer is similar to the action of a trampoline; a pressure differential on the top and bottom diaphragm faces acts to deform it. The magnitude of the deformation is proportional to the pressure difference. Both membrane and corrugated designs are used. Membranes are made of metal or nonmetallic material, such as plastic or neoprene. The material chosen depends on the pressure range anticipated and the fluid in contact with it. Corrugated diaphragms contain a number of corrugations that serve to increase diaphragm stiffness and to increase the diaphragm effective surface area.

Pressure transducers that use a diaphragm sensor are well suited for either static or dynamic pressure measurements. They have good linearity and resolution over their useful range. An advantage of the diaphragm sensor is that the very low mass and relative stiffness of the thin diaphragm give the sensor a very high natural frequency with a small damping ratio. Hence, these transducers can have a very wide frequency response and very short 90% rise and settling times. The natural frequency (rad/s) of a circular

diaphragm can be estimated by (Hetenyi, 1950)

$$\omega_n = 10.21 \sqrt{\frac{E_m t^2}{12(1 - \nu_p^2) r^4}} \quad (15.10)$$

where E_m is the material bulk modulus (psi or N/m²), t the thickness (in. or m), r the radius (in. or m), ρ the material density (lb_m/in.³ or kg/m³), and ν_p the Poisson's ratio for the diaphragm material. The maximum elastic deflection of a uniformly loaded, circular diaphragm supported about its circumference occurs at its center and can be estimated by

$$y_{\max} = \frac{3(p_1 - p_2)(1 - \nu_p^2)r^4}{16E_m t^3} \quad (15.11)$$

provided that the deflection does not exceed one-third the diaphragm thickness. Diaphragms should be selected so as to not exceed this maximum deflection over the anticipated operating range.

Various secondary elements are available to translate this displacement of the diaphragm into a measurable signal. Several methods are discussed below.

15.3.4 Strain Gauge Elements

A common method for converting diaphragm displacement into a measurable signal is to sense the strain induced on the diaphragm surface as it is displaced. Strain gauges, devices whose measurable resistance is proportional to their sensed strain (see Chapter 11), can be bonded directly onto the diaphragm, integrated within the diaphragm material or onto a deforming element (such as a thin beam) attached to the diaphragm so as to deform with the diaphragm and to sense strain. Metal strain gauges can be used with liquids. Strain gauge resistance is reasonably linear over a wide range of strain and can be directly related to the sensed pressure (Way, 1934). A diaphragm transducer using strain gauge detection is depicted in Figure 15.12.

By using semiconductor technology in pressure transducer construction, we now have a variety of very fast, very small, highly sensitive strain gauge diaphragm transducers. Silicone piezoresistive strain gauges can be diffused into a single crystal of silicone wafer, which forms the diaphragm. Semiconductor strain gauges have a static sensitivity that is 50times greater than conventional metallic strain gauges. Because the piezoresistive gauges are integral to the diaphragm, they are relatively immune to the thermoelastic strains prevalent in conventional metallic strain gauge–diaphragm constructions. Furthermore, a silicone diaphragm does not creep with age (as does a metallic gauge), thus minimizing calibration drift over time. However, uncoated silicone does not tolerate liquids.

15.3.5 Capacitance Elements

Another common method to convert diaphragm displacement to a measurable signal is a capacitance sensor. One version uses a thin metallic diaphragm as one plate of a capacitor paired with a fixed plate to complete the capacitor. The diaphragm is exposed to the process pressure on one side and to a reference pressure on the other or to a differential

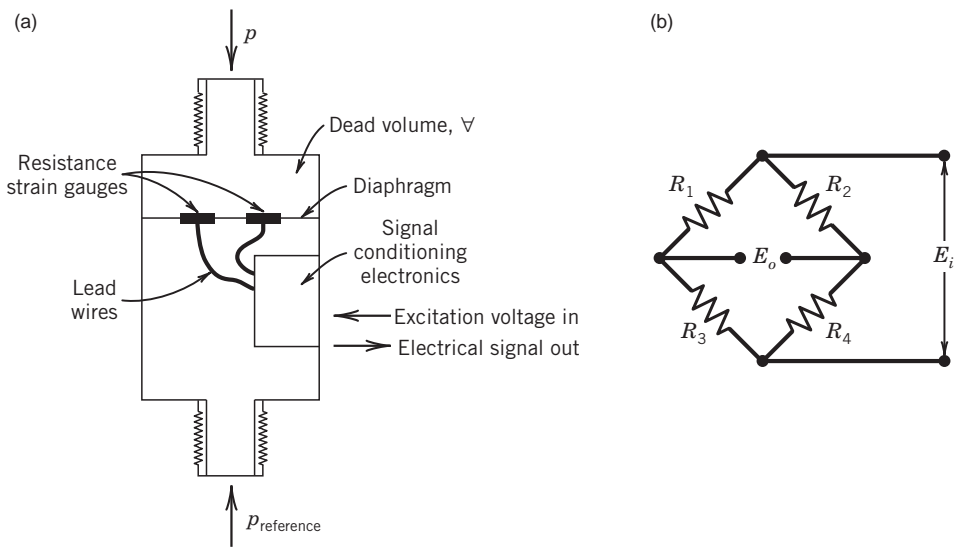


FIGURE 15.12 Diaphragm pressure transducer: (a) sensing scheme and (b) bridge–strain gauge circuit for pressure diaphragms.

pressure. When pressure changes, so as to deflect the diaphragm, the gap between the plates changes, which causes a change in capacitance.

To illustrate this, a transducer using this method is depicted in Figure 15.13. The capacitance C developed between two parallel plates separated by average gap t is determined by

$$C = c\epsilon A/t \quad (15.12)$$

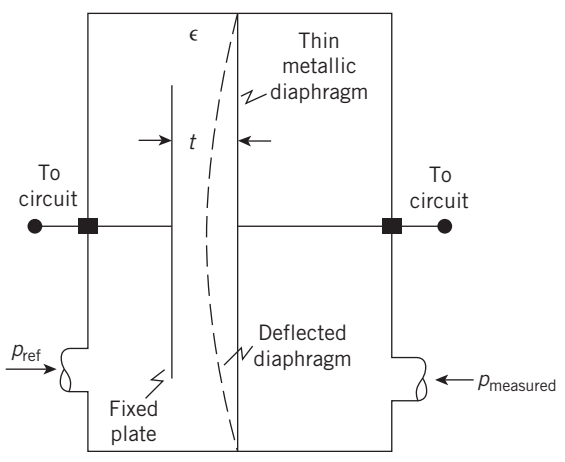


FIGURE 15.13 Capacitance pressure transducer. In this schematic, the diaphragm is conductive and its deflection exaggerated.

where the product $c\epsilon$ is the permittivity of the material between the plates relative to a vacuum ($\epsilon = 8.85 \times 10^{-12}$ F/m; c = dielectric constant), and A is the overlapping area of the two plates. The dielectric constant depends on the material in the gap, which for air is $c = 1$ but for water is $c = 80$. The capacitance responds to an instantaneous change in the area-averaged plate gap separation from which the time-dependent pressure is determined. However, the capacitance change is small relative to the absolute capacitance. Oscillator and bridge circuits are commonly used to operate the circuit and to measure the small capacitance change providing an output voltage E_0 .

The capacitance pressure transducer has the attractive features of other diaphragm transducers, including small size and a wide operating range. Many common and inexpensive pressure transducers use this measuring principle and are suitable for many engineering measurement demands, including a niche as the on-board car tire pressure sensor of choice. However, the principle is sensitive to temperature changes and has a relatively high impedance output. If attention is paid to these shortcomings, an accurate and stable transducer can be made.

15.3.6 Piezoelectric Crystal Elements

Piezoelectric crystals form effective secondary elements for dynamic (transient) pressure measurements. Under the action of compression, tension, or shear, a piezoelectric crystal deforms and develops a surface charge q , which is proportional to the force acting to bring about the deformation. In a piezoelectric pressure transducer, a preloaded crystal is mounted to the diaphragm sensor, as indicated in Figure 15.14. Pressure acts normal to the crystal axis and changes the crystal thickness t by a small amount Δt . This sets up a charge, $q = K_q p A$, where p is the pressure acting over the electrode area A and K_q is the crystal charge sensitivity, a material property. A charge amplifier (see Chapter 6) is used to convert charge to voltage so that the voltage developed across the electrodes is

$$E_o = q/C \quad (15.13)$$

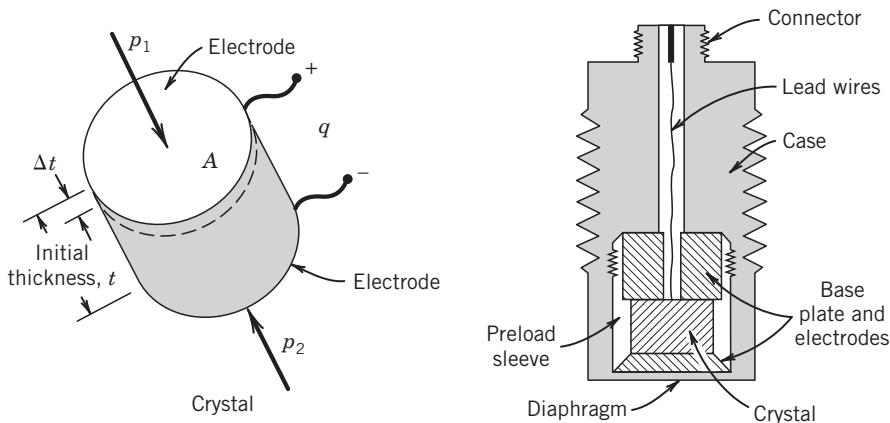


FIGURE 15.14 Piezoelectric pressure transducer.

where C is the capacitance of the crystal–electrode combination, again given by Equation (15.12). The operating equation becomes

$$E_o = K_q tp / c\epsilon = Kp \quad (15.14)$$

where K is the overall transducer gain. The crystal sensitivity for quartz, the most common material used, is $K_q = 2.2 \times 10^{-9}$ coulombs/N.

15.4 PRESSURE TRANSDUCER CALIBRATION

15.4.1 Static Calibration

A static calibration of a pressure transducer is usually accomplished either by direct comparison against any of the pressure reference instruments discussed (Section 15.3) or against the output from a certified laboratory standard transducer. For the low-pressure range, the McLeod gauge or the appropriate manometric instruments along with the laboratory barometer serve as convenient working standards. The approach is to pressurize (or evacuate) a chamber and expose both the reference instrument, which serves as the standard, and the candidate transducer to the same pressure for a side-by-side measurement. For the high-pressure range, the deadweight tester is a commonly used pressure reference standard.

15.4.2 Dynamic Calibration

The rise time and frequency response of a pressure transducer are found by dynamic calibration, and there are a number of clever ways to accomplish this (Instrument Society of America (ISA), 2002). As discussed in Chapter 3, the rise time of an instrument is found through a step change in input. For under-damped systems, natural frequency and damping ratio can be measured based on the ringing behavior (see Chapter 3), giving the frequency response. The frequency response can be found directly by applying a constant amplitude periodic input signal and varying its frequency. A flush-mounted transducer places the sensor in direct contact with the fluid at the measurement site. Because pressure transducers could also be attached by means of a pressure tap or by connecting tubing to the tap, this length (called the transmission line) affects the overall response and should be included as part of the dynamic calibration.

An electrical switching valve or flow control valve can create a step change in pressure. But the mechanical lag of the valve limits its use to transducers having an expected rise time of 50 ms or more. Faster applications might use a shock tube calibration or some equivalent diaphragm burst test.

As shown in Figure 15.15, the shock tube consists of a long pipe separated into two chambers by a thin diaphragm. The transducer is mounted into the pipe wall of one chamber, the expansion section, at pressure p_1 . The pressure in the other chamber, the driver section, is raised from p_1 to p_2 . Some mechanism, such as a mechanically controlled needle, is used to burst the diaphragm on command. Upon bursting, the pressure differential causes a pressure shock wave to move down the low-pressure chamber. A shock wave has a thickness on the order of $1 \mu\text{m}$ and moves at the speed of sound, a . So as the shock passes the transducer, the transducer experiences a change in pressure from p_1 to p_3 over a time $t = d/a$, where d is the diameter

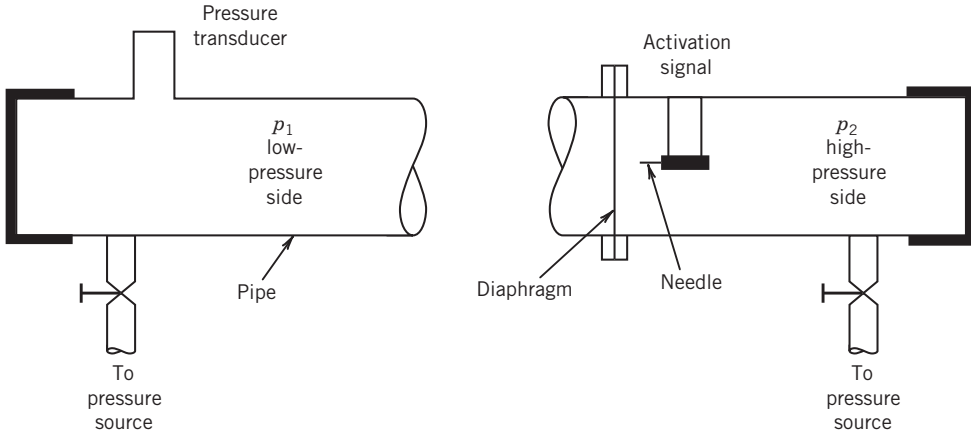


FIGURE 15.15 Schematic of a shock tube facility.

of the transducer pressure port, and pressure p_3 is

$$p_3 = p_1[1 + (2k/k + 1)(M_1^2 - 1)] \quad (15.15)$$

where k is the gas-specific heat ratio and M_1 is the Mach number calculated using normal shock wave tables and absolute pressure p_1 . The velocity of the shock wave can also be deduced from the output of fast-acting standard pressure sensors mounted in the shock tube wall. Typical values of t are on the order of 1–10 μs , so this method is at least four orders of magnitude faster than a switching valve. The transducer rise time is calculated from the output record.

A common verification check for system response is the “pop test,” which is well suited for liquids or gases and systems needing just moderate response times. In this situation, the transducer and connecting tubing are pressurized to a steady value, perhaps by using a small syringe or hand pump. The system is suddenly vented to atmosphere. The recorded transducer response gives an indication of the system rise time and ringing behavior. One variation of this approach attaches a balloon or similar flexible material to one end of the connecting tubing/transducer system. After pressurizing, the balloon is popped to suddenly vent the system.

A reciprocating piston within a cylinder can generate a sinusoidal variation in pressure for frequency-response calibration. The piston can be driven by a variable speed motor and its displacement measured by a fast-responding transducer, such as an LVDT (see Chapter 12). Under properly controlled conditions (Ex. 1.2), the actual pressure variation can be estimated from the piston displacement. Other techniques include an encased loud-speaker or an acoustically resonant enclosure, which serves as a frequency driver instead of a piston, or using an oscillating flow control valve to vary system pressure with time (Instrument Society of America (ISA), 2002).

15.5 PRESSURE MEASUREMENTS IN MOVING FLUIDS

Pressure measurements in moving fluids warrant special consideration. Consider the flow over the bluff body shown in Figure 15.16. Assume that the upstream flow is uniform and

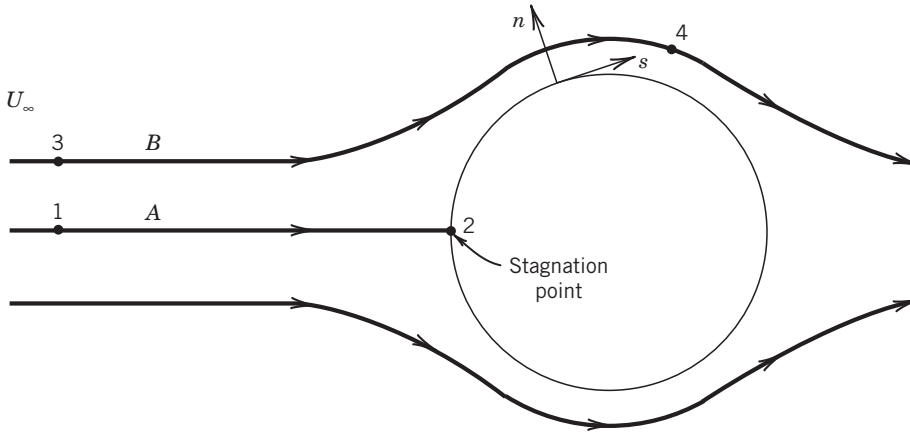


FIGURE 15.16 Streamline flow over a bluff body.

steady with negligible losses. Along streamline A, the upstream flow moves with a velocity U_1 , such as at point 1. As the flow approaches point 2, it must slow down and finally stop at the front end of the body. Above streamline A, flow moves over the top of the bluff body, and below streamline A, flow moves under the body. Point 2 is known as the stagnation point and streamline A the stagnation streamline for this flow. Along streamline B, the velocity at point 3 is U_3 and because the upstream flow is considered to be uniform it follows that $U_1 = U_3$. As the flow along B approaches the body, it is deflected around the body. From the conservation of mass principles, $U_4 > U_3$. Application of conservation of energy between points 1 and 2 and between 3 and 4 yields

$$\begin{aligned} p_1 + \rho U_1^2/2 &= p_2 + \rho U_2^2/2 \\ p_3 + \rho U_3^2/2 &= p_4 + \rho U_4^2/2 \end{aligned} \quad (15.16)$$

However, because point 2 is the stagnation point, $U_2 = 0$, and

$$p_2 = p_t = p_1 + \rho U_1^2/2 \quad (15.17)$$

Hence, it follows that $p_2 > p_1$ by an amount equal to $p_v = \rho U_1^2/2$, called the *dynamic pressure*, an amount equivalent to the kinetic energy per unit mass of the flow as it moves along the streamline. If no energy is lost through irreversible processes, such as through a transfer of heat,¹ this translational kinetic energy is transferred completely into p_2 . The value of p_2 is known as the *stagnation* or the *total pressure* and is noted as p_t . The total pressure can be determined by bringing the flow to rest at a point in an isentropic manner.

The pressures at 1, 3, and 4 are known as static pressures² of the flow. The *static pressure* is that pressure sensed by a fluid particle as it moves with the same velocity as the

¹This is a realistic assumption for subsonic flows. In supersonic flows, the assumption is not valid across a shock wave.

²The term “static pressure” is a misnomer in moving fluids, but its use here conforms to common expression. “Stream pressure” is more appropriate and is sometimes used.

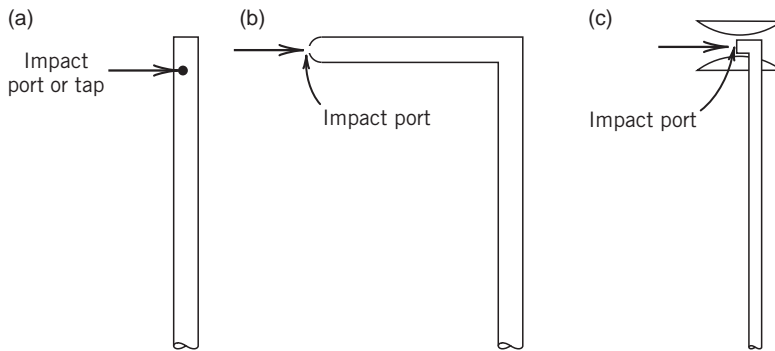


FIGURE 15.17 Total pressure measurement devices. (a) Impact cylinder. (b) Pitot tube. (c) Kiel probe.

local flow. The static pressure and velocity at points 1 and 3 are given the special names of the “freestream static pressure” and “freestream velocity.” Since $U_4 > U_3$, Equation (15.16) shows that $p_4 < p_3$. It follows from Equation (15.17) that the total pressure is the sum of the static and dynamic pressures anywhere in the flow.

15.5.1 Total Pressure Measurement

In practice, the total pressure is measured using an impact probe, such as those depicted in Figure 15.17. A small hole in the impact probe is aligned with the flow so as to cause the flow to come to rest at the hole. The sensed pressure is transferred through the impact probe to a pressure transducer or other pressure-sensing device such as a manometer. Alignment with the flow is somewhat critical, although the probes in Figure 15.17a and b are relatively insensitive (within $\sim 1\%$ error in indicated reading) to misalignment within a $\pm 7^\circ$ angle (American Society of Mechanical Engineers (ASME), 1987). A special type of impact probe shown in Figure 15.17c, known as a Kiel probe, uses a shroud around the impact port. The effect of the shroud is to force the local flow to align itself with the shroud axis so as to impact directly onto the impact port. This effectively eliminates total pressure sensitivity to misalignment up to $\pm 40^\circ$ (American Society of Mechanical Engineers (ASME), 1987).

15.5.2 Static Pressure Measurement

The local value of static pressure in a moving stream is measured by sensing the pressure in the direction that is normal to the flow streamline. Within ducted flows, static pressure is sensed by wall taps, which are small, burr-free, holes drilled into the duct wall perpendicular to the flow direction at the measurement point. The tap is fitted with a hose or tube, which is connected to a pressure gauge or transducer. A recommended design for a wall tap is shown in Figure 15.18. The tap hole diameter d is typically between 1% and 10% of the pipe diameter, with the smaller size preferred (Franklin and Wallace, 1970). The tap must be perpendicular to the local tangent to the wall with no drilling burrs (Rayle, 1959).

Alternatively, a static pressure probe can be inserted into the flow to measure local stream pressure. It should be a streamlined design to minimize the disturbance of the

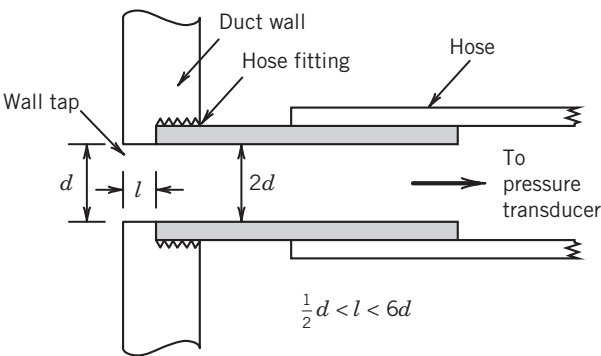


FIGURE 15.18 Anatomy of a static pressure wall tap.

flow. It should be physically small so as not to cause more than a negligible increase in velocity in the vicinity of measurement. As a rule, the frontal area of the probe should not exceed 5% of the pipe flow area. The static pressure sensing port should be located well downstream of the leading edge of the probe so as to allow the streamlines to realign themselves parallel with the probe. Such a concept is built into the improved Prandtl tube design shown in Figure 15.19a.

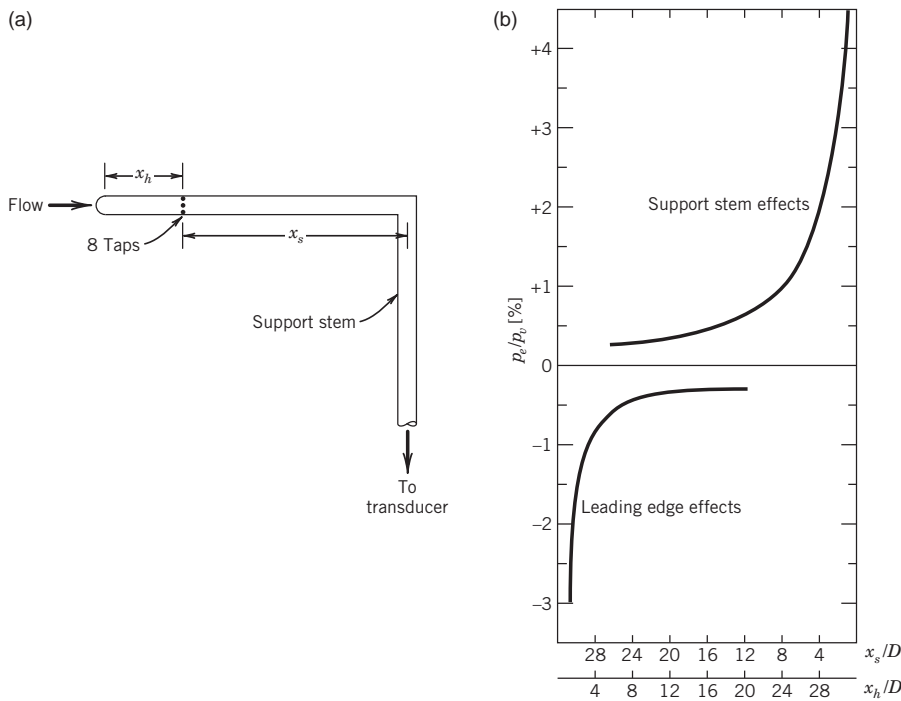


FIGURE 15.19 Improved Prandtl tube for static pressure. (a) Design. (b) Relative static error along tube length.

A Prandtl tube probe consists of eight holes arranged about the probe circumference and positioned 8–16 probe diameters downstream of the probe leading edge and 16 probe diameters upstream of its support stem. A pressure transducer or manometer is connected to the probe stem to measure the sensed pressure. The hole positions are chosen to minimize static pressure error caused by the disturbance to the flow streamlines due to the probe's leading edge and stem. This is illustrated in Figure 15.19b, where the relative static error, $p_e/p_v = (p_i - p)/(\frac{1}{2}\rho U^2)$, is plotted as a function of tap location along the probe body. Real viscous effects around the static probe cause a slight discrepancy between the actual static pressure and the indicated static pressure. To account for this, a correction factor, C_0 , is used with $p = C_0 p_i$ where $0.99 < C_0 < 0.995$ and p_i is the indicated (measured) pressure.

15.6 MODELING PRESSURE AND FLUID SYSTEMS

Fluid systems can be modeled using lumped parameter ideal elements just as common resistor–inductor–capacitor electrical loops and mass–damper–spring mechanical systems are used. The common elements are inertance, resistance, and compliance.

Inertance describes the inertial properties of a mass in motion, such as that of a mass of fluid moving within a vessel. For a fluid of density ρ and a vessel of cross-sectional area A and length ℓ the inertance is written as

$$L_f = \rho\ell/A \quad (15.18)$$

When modeling inertial forces in laminar flows, this value should be increased by a factor of $\frac{4}{3}$. Inertance is the direct analog to electrical inductance.

Fluid *resistance* describes the opposition to motion. This is the pressure change required to move a volume of fluid per unit time, Q . It is written

$$R = \Delta p^n / Q = \Delta E / I \quad (15.19)$$

where $n = 1$ for laminar and $n = 0.5$ for turbulent flows. Hence, $Q = \frac{1}{R} \Delta p^n$.

The resistance of the laminar flow of a Newtonian fluid through a circular pipe is $R = \frac{128\mu\ell}{\pi d^4}$, where μ is the fluid viscosity. In electrical systems, it has the analogous meaning as the opposition to current flow for an imposed voltage potential, such as an electrical resistor.

Compliance describes a measure of the volume change associated with a corresponding pressure change, such as

$$C_{vp} = \Delta v / \Delta p \quad (15.20)$$

It is a measure of the flexibility in a structure, component, or substance, and so it is the inverse of the system *stiffness*. Compliance is the direct analog to electrical capacitance.

15.7 DESIGN AND INSTALLATION: TRANSMISSION EFFECTS

Consider the configuration depicted in Figure 15.20 in which a tube of volume \forall_t with length ℓ and diameter d is used to connect a pressure tap to a pressure transducer of internal dead volume \forall (e.g., Figure 15.12). Under static conditions, the pressure transducer

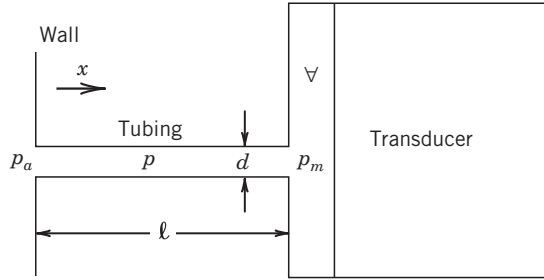


FIGURE 15.20 Wall tap to pressure transducer connection: the transmission line.

indicates the static pressure at the tap. But if the pressure at the tap is a time-dependent pressure, $p_a(t)$, the response behavior of the tubing influences the time-indicated output from the transducer, $p(t)$.

By considering the one-dimensional pressure forces acting on a lumped mass of fluid within the connecting tube, balancing inertance, compliance, and resistance against forcing function, we can develop a model for the pressure system response. A network model is shown in Figure 15.21a in which the circuit is driven between two pressures, the applied pressure $p_a(t)$ at the tap and the measured pressure $p_m(t)$ at the transducer sensor. Using the electrical analog, inertance is modeled by the inductor, fluid resistance by a resistor, and compliance by a capacitor (Figure 15.21b). The circuit analysis of the two loops gives

$$L\ddot{I} + RI + \frac{1}{C} \int Idt = E_a \quad \text{and} \quad \frac{1}{C} \int Idt = E_m \quad (15.21)$$

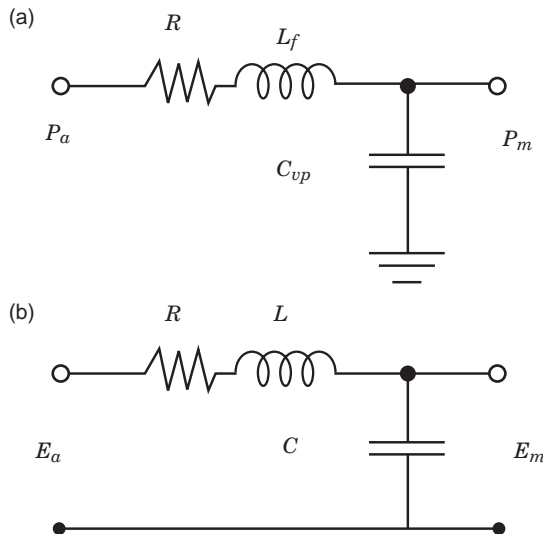


FIGURE 15.21 An equivalent lumped parameter network of the pressure transmission line model of Figure 15.20 using an electrical analogy.

Taking the derivative of the second loop to get \dot{E}_m and \ddot{E}_m in terms of \dot{I} and \ddot{I} , then substituting these back into Equation (15.21) with $E_a = p_a$, $E_m = p_m$, $L = L_f$, and $C = C_{vp}$ gives

$$L_f C_{vp} \ddot{p}_m + R C_{vp} \dot{p}_m + p_m = p_a(t) \quad (15.22)$$

Substituting Equations (15.18) to (15.20) into Equation (15.22) gives the working system response equation for the applied and measured pressures,

$$\frac{16\ell\rho C_{vp}}{3\pi d^2} \ddot{p}_m + \frac{128\mu\ell C_{vp}}{\pi d^4} \dot{p}_m + p_m = p_a(t) \quad (15.23)$$

in which we have augmented the fluid inertial force by $\frac{4}{3}$ (Munson et al., 2009).

In this simple model, the system compliance lumps the compliance of the fluid, tube walls, and transducer into a single value, C_{vp} . As these individual compliances could be modeled separately by using capacitors in parallel, the total capacitance is simply the sum of each. If one compliance dominates, the others can be neglected. Further, the inertance and resistance of the connecting tube and the transducer cavity are lumped into single values. Improved models use distributed, lumped parameters, such as are commonly used to model physiological vascular systems (Migliavacca, 2006). A long-standing approach to modeling the transmission line (Delio et al., 1949) examines the forces acting on a fluid element. In this model, small pressure changes act on the element, moving it back and forth by a distance x within the tube. Summing the forces and substituting for p_m again results in Equation (15.23) (Doebelin, 2003). Be aware that models are based on simplifying assumptions and should be used only as a guide in the design of a system, not as a replacement for *in situ* calibration.

We can study the transient and frequency response of the system represented by Equation (15.23) by extracting values for ω_n and ζ ,

$$\omega_n = \frac{d}{4} \sqrt{3\pi/\rho\ell C_{vp}} \quad (15.24)$$

$$\zeta = \frac{16\mu}{d^3} \sqrt{3\ell C_{vp}/\pi\rho} \quad (15.25)$$

The total system compliance could be measured using Equation (15.20) by closing the pressure tap, increasing the fluid volume in the tubing by a small known amount, such as by syringe, and measuring the corresponding pressure change.

15.7.1 Liquids

Liquids are relatively incompressible, so that the compression–restoring force in the transmission line is due primarily to the compliance in the transducer, which is a transducer specification, for use in Equations (15.24) and (15.25). Connecting tubing can usually be considered rigid for the underlying assumptions of the above lumped parameter analysis. Thick-walled, flexible tubing is often used, but this is fairly rigid and its compliance can be ignored. If need be, the compliance can be measured.

15.7.2 Gases

For gases, we simplify by assuming that the system is rigid relative to the compressibility of the gas. Compliance is then modeled in terms of the fluid's adiabatic bulk modulus of elasticity, $E_m = \forall / C_{vp}$. This gives

$$\omega_n = \frac{d}{4} \sqrt{3\pi E_m / \rho \ell \forall} \quad (15.26)$$

$$\zeta = \frac{16\mu}{d^3} \sqrt{3\ell \forall / \pi \rho E_m} \quad (15.27)$$

Equations (15.26) and (15.27) can also be written in terms of the speed of sound for the gas, a , which is related to its compressibility by $a = \sqrt{E_m / \rho}$ and for a perfect gas by $a = \sqrt{kRT}$, where T is the gas absolute temperature, giving

$$\omega_n = \frac{ad}{4} \sqrt{3\pi / \ell \forall} \quad (15.28)$$

$$\zeta = \frac{16\mu}{ad^3} \sqrt{3\ell \forall / \pi} \quad (15.29)$$

When the tube volume, $\forall_t \gg \forall$, then a series of standing pressure waves develop and we can expect $\omega \sim O(a/\ell)$. Hougén et al. (1963) discuss an improved prediction as

$$\omega_n = \frac{a}{\ell \sqrt{0.5 + \forall / \forall_t}} \quad (15.30)$$

$$\zeta = \frac{16\mu \ell}{\rho a d^2} \sqrt{0.5 + \forall / \forall_t} \quad (15.31)$$

Note that in all cases, larger diameter and shorter length tubes improve pressure system response.

15.7.3 Heavily Damped Systems

In systems in which we estimate a damping ratio greater than 1.5, the frequency response model can be simplified further. The behavior of the pressure-measuring system closely follows that of a first-order system. Again, the typical pressure transducer-tubing system has a compliance C_{vp} , which is a measure of the transducer and tubing volume change relative to an applied pressure change. The response of the first-order system is indicated through its time constant, which is estimated by neglecting the second-order term in Equation 15.23 so that (Doebelin, 2003)

$$\tau = \frac{128\mu \ell C_{vp}}{\pi d^4} \quad (15.32)$$

Equation 15.32 shows that the time constant is proportional to ℓ/d^4 . Long- and small-diameter connecting tubes promote a more sluggish system response to changes in pressure.

15.8 FLUID VELOCITY MEASURING SYSTEMS

Velocity measuring systems are used to measure the local velocity in a moving fluid. Desirable information can consist of the mean velocity, as well as any of the dynamic components of the velocity. Dynamic components are found in pulsating, phasic, or oscillating flows, or in turbulent flows. For most general engineering applications, information about the mean flow velocity is usually sufficient. The dynamic velocity information is often sought during applied and basic fluid mechanics research and development, such as in attempting to study airplane wing response to air turbulence, a complex periodic waveform as the wing sees it. In general, the instantaneous velocity can be written as

$$U(t) = \bar{U} + u \quad (15.33)$$

where \bar{U} is the mean velocity and u is the time-dependent dynamic (fluctuating) component of the velocity. The instantaneous velocity can also be expressed in terms of a Fourier series:

$$U(t) = \bar{U} + \sum_i C_i \sin(\omega_i t + \phi'_i) \quad (15.34)$$

so that the mean velocity and the amplitude and frequency information concerning the dynamic velocity component can be resolved with a Fourier analysis of the time-dependent velocity signal.

15.8.1 Pitot-Static Pressure Probe

For a steady, incompressible, isentropic flow, Equation 15.16 can be written at any arbitrary point x in the flow field as

$$p_t = p_x + \frac{1}{2} \rho U_x^2 \quad (15.35)$$

or, rearranging,

$$p_v = p_t - p_x + \frac{1}{2} \rho U_x^2 \quad (15.36)$$

Again p_v , the difference between the total and static pressures at any point in the flow, is the *dynamic pressure*. Hence, measuring the dynamic pressure of a moving fluid at a point provides a method for estimating the local velocity,

$$U_x = \sqrt{\frac{2p_v}{\rho}} = \sqrt{\frac{2(p_t - p_x)}{\rho}} \quad (15.37)$$

In practice, Equation 15.37 is utilized through a device known as a *pitot-static pressure probe*. Such an instrument has an outward appearance similar to that of an improved Prandtl static pressure probe (Figure 15.19a), except that the pitot-static probe contains an interior pressure tube attached to an impact port at the leading edge of the probe, as shown in Figure 15.22. This creates two coaxial internal cavities within

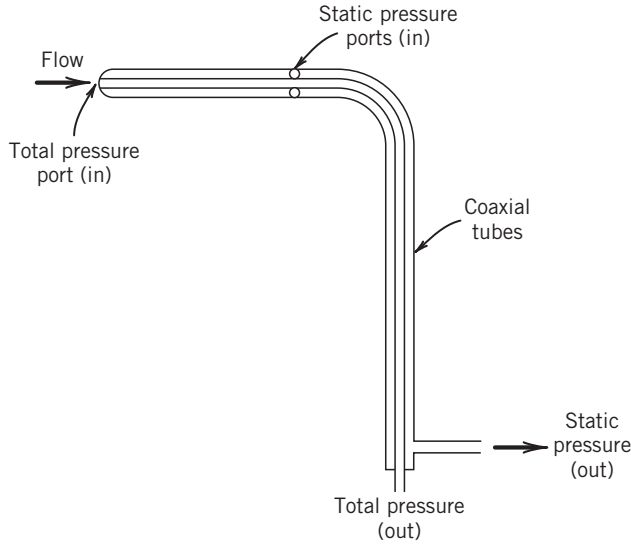


FIGURE 15.22 Pitot-static pressure probe.

the probe, one exposed to the total pressure and the second exposed to the static pressure. The two pressures are typically measured using a differential pressure transducer so as to indicate p_v directly.

The pitot-static pressure probe is relatively insensitive to misalignment over the yaw angle range of ± 15 degrees (American Society of Mechanical Engineers (ASME), 1987). When possible, the probe can be rotated until a maximum signal is measured, a condition that indicates that it is aligned with the mean flow direction. However, the probes have a lower velocity limit of use that is brought about by strong viscous effects in the entry regions of the pressure ports. In general, viscous effects should not be a concern, provided that the Reynolds number based on the probe radius, $Re_r = \bar{U}r/\nu$, is greater than 500, where ν is the kinematic viscosity of the fluid. For $10 < Re_r < 500$, a correction to the dynamic pressure should be applied, $p_v = C_v p_i$, where

$$C_v = 1 + (4/Re_r) \quad (15.38)$$

and p_i is the indicated dynamic pressure from the probe. However, even with this correction, the measured dynamic pressure has a systematic uncertainty on the order of 40% at $Re_r \approx 10$ but decreases to 1% for $Re_r \geq 500$.

In high-speed gas flows, compressibility effects near the probe leading edge require a closer inspection of the governing equation for a pitot-static pressure probe. An energy balance along a streamline for a perfect gas between any point x and the stagnation point can be written as

$$\frac{U^2}{2} = c_p(T_t - T_\infty) \quad (8.34)$$

For an isentropic process, the relationship between temperature and pressure can be stated as

$$\frac{T_x}{T_t} = \left(\frac{p_x}{p_t} \right)^{(k-1)/k} \quad (15.39)$$

where k is the ratio of specific heats for the gas, $k = c_p/c_v$. The Mach number of a moving fluid relates its local velocity to the local speed of sound,

$$M = U/a \quad (15.40)$$

where the speed of sound, also called the acoustic wave speed, for a perfect gas is $a = \sqrt{kRT_x}$ where T_x is the absolute temperature of the gas at the point of interest. Combining these equations and using a binomial expansion gives a relationship between total pressure and static pressure at any point x in a moving compressible flow,

$$p_v = p_t - p_x = \frac{1}{2} \rho U_x^2 \left[1 + M^2/4 + (2-k)M^4/24 + \dots \right] \quad (15.41)$$

Equation 15.41 reduces to Equation 15.36 when $M \ll 1$. The error in the estimate of p_v , based on the use of Equation 15.36 relative to the true dynamic pressure, becomes significant for $M > 0.3$, as shown in Figure 15.23. Thus, $M \sim 0.3$ is used as the incompressible limit for perfect gas flows.

For $M > 1$, the local velocity is found through iteration using the Rayleigh relation:

$$p_t/p = \left(\frac{(k+1)^2 M^2}{4kM^2 - 2(k-1)} \right)^{k/k-1} \left(\frac{1-k+2kM^2}{k+1} \right) \quad (15.42)$$

and Equation 15.40 where both p and p_t are the measured values.

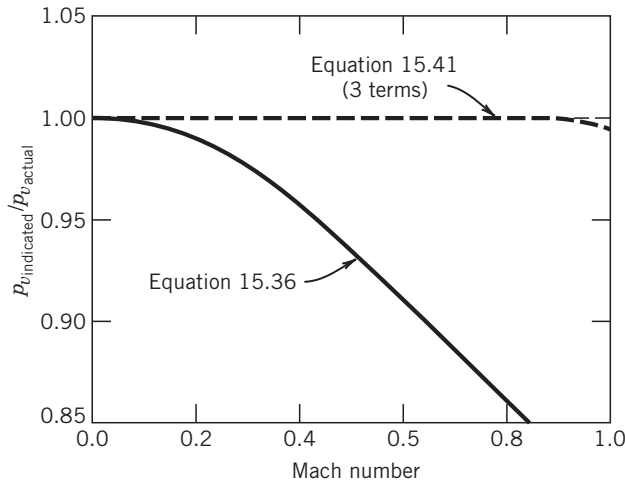


FIGURE 15.23 Relative error in the dynamic pressure between using Equations 15.36 and 15.41 at increasing flow speeds.

15.8.2 Thermal Anemometry

The rate at which energy, \dot{Q} , is transferred between a warm body at T_s and a cooler moving fluid at T_f is proportional both to the temperature difference between them and to the thermal conductance of the heat transfer path, hA . This thermal conductance increases with fluid velocity, thereby increasing the rate of heat transfer at any given temperature difference. Hence, a relationship between the rate of heat transfer and velocity exists forming the working basis of a thermal anemometer.

A thermal anemometer utilizes a sensor, a metallic resistance temperature detector (RTD) element, that makes up one active leg of a Wheatstone bridge circuit, as indicated in Figure 15.24. The resistance–temperature relation for such a sensor was shown in Chapter 8 to be well represented by

$$R_s = R_0[1 + \alpha(T_s - T_0)] \quad (15.43)$$

so that sensor temperature T_s can be inferred through a resistance measurement. A current is passed through the sensor to heat it to some desired temperature above that of the host fluid. The relationship between the rate of heat transfer from the sensor and the cooling fluid velocity is given by King's law (King, 1914) as

$$\dot{Q} = I^2 R = A + BU^n \quad (15.44)$$

where A and B are constants that depend on the fluid and sensor physical properties and operating temperatures, and n is a constant that depends on sensor dimensions (Hinze, 1959). Typically, $0.45 \leq n \leq 0.52$ (Collis and Williams,). A , B , and n are found through calibration.

Two types of sensors are common: the hot-wire and the hot-film. As shown in Figure 15.25, the hot-wire sensor consists of a tungsten or platinum wire ranging from 1 to 4 mm in length and from 1.5 to 15 μm in diameter. The wire is supported between two rigid needles that protrude from a ceramic tube that houses the lead wires. A hot-film sensor usually consists of a thin (2 μm) platinum or gold film deposited onto a glass substrate and covered with a high thermal conductivity coating. The coating acts to electrically insulate the film and offers some mechanical protection. Hot wires are generally used in electrically nonconducting fluids, while hot films can be used in conducting fluids or in nonconducting fluids when a more rugged sensor is needed.

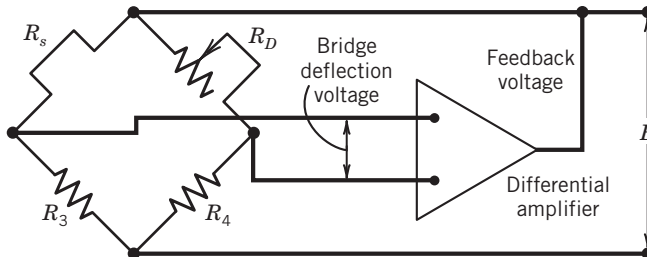


FIGURE 15.24 Thermal anemometer circuit, shown in constant resistance mode.

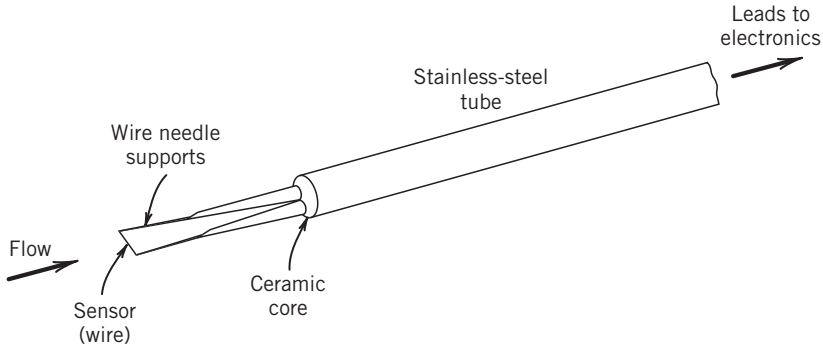


FIGURE 15.25 Schematic of a hot-wire probe.

Two anemometer bridge operating modes are possible: (1) constant current and (2) constant resistance. In constant-current operation, a fixed current is passed through the sensor to heat it. The sensor resistance and therefore its temperature, based on Equation 15.43, are permitted to vary with the rate of heat transfer between the sensor and its environment. Bridge-deflection voltage provides a measure of the cooling velocity. The more common mode of operation for velocity measurements is constant resistance. In constant-resistance operation, the sensor resistance of Equation 15.43 is originally set by adjusting the bridge balance. The sensor resistance is then maintained constant by using a differential feedback amplifier to sense small changes in bridge balance, which would be equivalent to sensing changes in the sensor set-point resistance; that is, the circuit acts as a closed loop controller using the bridge balance as the error signal. The feedback amplifier rapidly readjusts the bridge applied voltage, thereby adjusting the sensor current to bring the sensor back to its set-point resistance and corresponding temperature. Because the current through the sensor varies with changes in the velocity, the instantaneous power ($I^2 R_s$) required to maintain this constant temperature is equivalent to the instantaneous rate of heat transfer from the sensor (\dot{Q}). In terms of the instantaneous applied bridge voltage, E , required to maintain a constant sensor resistance, the velocity is found by the general correlation

$$E^2 = C + DU^n \quad (15.45)$$

where the values of C , D , and n are found by calibration under a fixed sensor and fluid temperature condition. An electronic or digital linearizing scheme is usually employed to condition the signal by performing the transformation

$$E_1 = K \left(\frac{E^2 - C}{D} \right)^{1/n} \quad (15.46)$$

such that the measured output from the linearizer, E_1 , is

$$E_1 = KU \quad (15.47)$$

where K is found through a static calibration.

For mean velocity measurements, the thermal anemometer is a straightforward device to use. It has a better usable sensitivity than the pitot-static tube at lower velocities. Multiple velocity components can be measured by using multiple sensors, each sensor aligned differently to the mean flow direction and operated by independent anemometer circuits (Hinze, 1959; Rodi, 1975). Because it has a high-frequency response, fluctuating (dynamic) velocities can be measured. In highly turbulent flows with root-mean-square (rms) fluctuations of $\sqrt{u^2} \geq 0.1\bar{U}$, signal interpretation can become complicated, but it has been well investigated (Rodi, 1975). Low-frequency fluid temperature fluctuations can be compensated for by placing resistor R_3 directly adjacent to the sensor and exposed to the flow. An extensive bibliography of thermal anemometry theory and signal interpretation can be found elsewhere (Freymuth, 1978).

In constant temperature mode using a fast-responding differential feedback amplifier, a hotwire system can attain a frequency response that is flat up to 100,000 Hz, which makes it particularly useful in fluid mechanics turbulence research. However, less expensive and more rugged systems are commonly used for industrial flow monitoring, where a fast dynamic response is desirable. There is a natural upper frequency limit on a cylindrical sensor of diameter d that is brought about by the natural oscillation in the flow immediately downstream of a body. This vibrates the sensor. The frequency of this oscillation, known as the Strouhal frequency (and explained further in Section 10.6), occurs at approximately

$$f \approx 0.22[\bar{U}/d](10^2 < \text{Re}_d < 10^7) \quad (15.48)$$

The heated sensor warms the fluid within its proximity. Under flowing conditions this does not cause any measurable problems so long as the condition

$$\text{Re}_d \geq Gr^{1/3} \quad (15.49)$$

is met where $\text{Re}_d = \bar{U}d/\nu$, $Gr = d^3 g \beta (T_s - T_{\text{fluid}})/\nu^2$, and β is the coefficient of thermal expansion of the fluid. Equation 15.49 ensures that the inertial forces of the moving fluid dominate over the buoyant forces brought on by the heated sensor. This forms a lower velocity limit on the order of 0.6 m/s for using hot-wire sensors in air.

15.8.3 Doppler Anemometry

The Doppler effect describes the phenomenon experienced by an observer whereby the frequency of light or sound waves emitted from a source that is traveling away from or toward the observer is shifted from its original value and by an amount proportional to its speed. Most readers are familiar with the change in pitch of a train as heard by an observer as the train changes from approaching to receding. Any radiant energy wave, such as a sound or light wave, experiences a Doppler effect. The effect was recognized and modeled by Christian Johann Doppler (1803–1853). The observed shift in frequency, called the Doppler shift, is directly related to the speed of the emitter relative to the observer. To an independent observer, the frequency of emission is perceived to be higher than actual if the emitter is moving toward the observer and lower if moving away, because the arrival of the emission at the observer location is affected by the relative velocity of the emission source. The Doppler effect is used in astrophysics to measure the velocity of distant objects by monitoring the frequency of light emitted from a particular

gas, usually hydrogen. Since, in the visible light spectrum, frequency is related to color, the common terms of red shift or blue shift refer to frequency shifts toward the red side of the spectrum or toward the blue side.

Doppler anemometry refers to a class of techniques that utilize the Doppler effect to measure the local velocity in a moving fluid. In these techniques, the emission source and the observer remain stationary. However, small scattering particles suspended in and moving with the fluid can be used to generate the Doppler effect. The emission source is a coherent narrow incident wave. Either acoustic waves or light waves are used.

When a laser beam is used as the incident wave source, the velocity measuring device is called a laser Doppler anemometer (LDA). LDA measures the time-dependent velocity at a point in the flow. Yeh and Cummins (1964) in 1964 discussed the first practical laser Doppler anemometer system. A laser beam provides a ready emission source that is monochromatic and remains coherent over long distances. As a moving particle suspended in the fluid passes through the laser beam, it scatters light in all directions. An observer viewing this encounter between the particle and the beam perceives the scattered light at a frequency, f_s :

$$f_s = f_i \pm f_D \quad (15.50)$$

where f_i is the frequency of the incident laser beam and f_D is the Doppler shift frequency. Using visible light, an incident laser beam frequency is on the order of 10^{14} Hz. For most engineering applications, the velocities are such that the Doppler shift frequency, f_D , is on the order of 10^3 – 10^7 Hz. Such a small shift in the incident frequency can be difficult to detect in a practical instrument. An operating mode that overcomes this difficulty is the dual-beam mode shown in Figure 15.26. In this mode, a single laser beam is divided into two coherent beams of equal intensity using an optical beam splitter. These incident beams are passed through a focusing lens that focuses the beams to a point in the flow. The focal point forms the effective measuring volume (sensor) of the instrument. Particles suspended in and moving with the fluid scatter light as they pass through the beams. The frequency of the scattered light is that given by Equation 15.50 everywhere but at the measuring volume. There, the two beams cross and the incident information from the two beams mix, a process known as optical heterodyne. The outcome of this mixing is a separation of the incident frequency from the Doppler frequency. A stationary observer, such as an optical photodiode, focused on the measuring volume, sees two distinct frequencies, the Doppler shift frequency and the unshifted incident frequency, instead of seeing their sum. It is a simple matter to separate the much smaller Doppler frequency from the incident frequency by filtering.

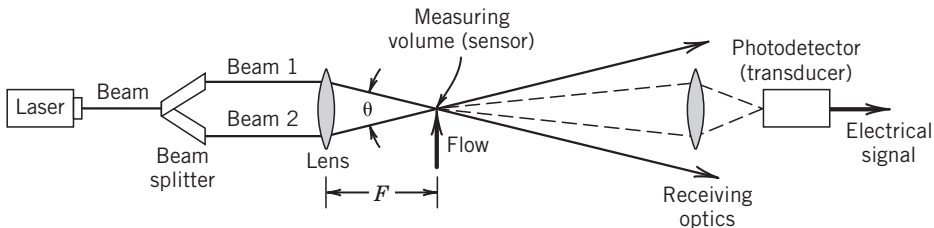


FIGURE 15.26 Laser Doppler anemometer, shown here in the dual-beam mode of operation.

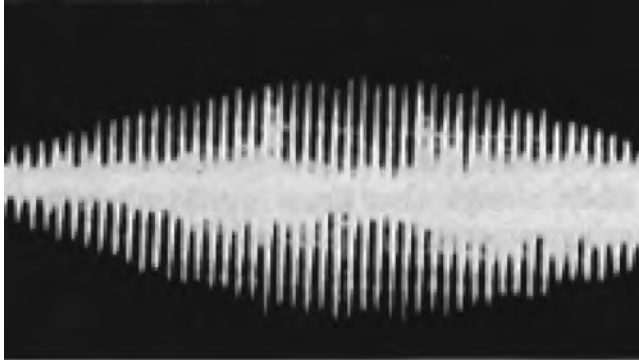


FIGURE 15.27 Oscilloscope trace of a photodiode output showing the Doppler frequency from a single particle moving through the measuring volume.

For the setup shown in Figure 15.26, the velocity is related directly to the Doppler shift by

$$U = \frac{\lambda}{2\sin\theta/2} f_D = d_f f_D \quad (15.51)$$

where the component of the velocity measured is that which is in the plane of and bisector to the crossing beams. In theory, by using beams of different color or polarization, different velocity components can be measured simultaneously. However, the dependence of the lens focal length on color causes a small displacement between the different measuring volumes formed by the different colors. For most applications this can be corrected. The LDA technique requires no direct calibration beyond explicit determination of the parameters in d_f and the ability to measure f_D .

In dual-beam mode, the output from the photodiode transducer is a current of a magnitude proportional to the square of the amplitude of the scattered light seen and of a frequency equal to f_D . This effect is seen as a Doppler “burst” shown in the typical oscilloscope trace of Figure 15.27. The Doppler burst is the frequency signal created by a particle moving through the measuring volume. If the instantaneous velocity of a dynamic flow varies with time, the Doppler shift from successive scatters will vary with time. This time-dependent frequency information can be extracted by any variety of processing equipment that can interpret the signal current. The most common is the burst analyzer.

Burst analyzers extract Doppler frequency information by performing a Fourier analysis (see Chapter 2) on the input signal. This is done by first discretizing the photodetector analog signal at a high sample rate and then analyzing the signal.

The analysis can work in one of two ways. In the first approach, the analyzer performs the fast Fourier transform (FFT) continuously on small blocks of data from which the Doppler frequency is directly determined. In the second mode, the correlation mode, the sampled signal is correlated with itself through a mathematical transformation of the form

$$R_j = \sum_{i=1}^n x(i)x(i+j) \quad (15.52)$$

where i refers to the sample value at time t and j to the sample value at time delay Δt . This operation improves the signal-to-noise ratio (SNR) of the signal; the frequency is

determined from the correlation function and the sample rate. From Equation 15.51, the Doppler frequency can be converted to a velocity and provided as an output signal. The acquisition and analysis occur rapidly so that the signal appears nearly continuous in time and with only a short time lag.

All methods output a voltage that is proportional to the instantaneous velocity, which makes their signal easy to process, or a digital output path to a digital file for signal analysis. At very low light levels and very few scattering particles, the signal level to noise level can be very low. In such cases, photon correlation techniques are successful (Cummins and Pike, 1973; Durst et al., 1976).

The LDA technique measures velocity at a point. Thus, its probe volume needs to be moved around to map out the flow field. If the SNR and seeding is good, the LDA can measure time-dependent velocities well. The LDA technique is particularly useful where probe blockage effects render other methods unsuitable, where fluid density and temperature fluctuations occur, or where environments hostile to physical sensors exist. An extended discussion of LDA techniques can be found elsewhere (Durst et al., 1976; Goldstein, 1996).

15.8.4 Particle Image Velocimetry

Particle image velocimetry (PIV) measures the full-field instantaneous velocities in a planar cross section of a flow. The technique tracks the time displacement of particles, which are assumed to follow the flow. Principle components for the technique are a coherent light source (laser beam), optics, a CCD-camera, and dedicated signal interrogation software.

In a simple overview, the image of particles suspended in the flow are illuminated and recorded during very-short-duration repetitive flashes of a laser beam. These images are recorded and compared. The distance traveled by any particle during the period between flashes is a measure of its velocity. By repeatedly flashing the laser, in the manner of a strobe light, the particle positions can be tracked and velocity as a function of time obtained.

In a typical layout, such as shown in Figure 15.28, a pulsed flash laser beam is passed through a cylindrical lens, which converts the beam into a two-dimensional (2-D) sheet of light. This laser sheet is mechanically situated to illuminate an appropriate cross section of the flow field. The camera is positioned and focused to record the view of the illuminated field. The laser flash and camera shutter are synchronized to capture the flow image. The acquired digital image is stored and processed by interrogation software, resulting in a full-field instantaneous velocity mapping of the flow.

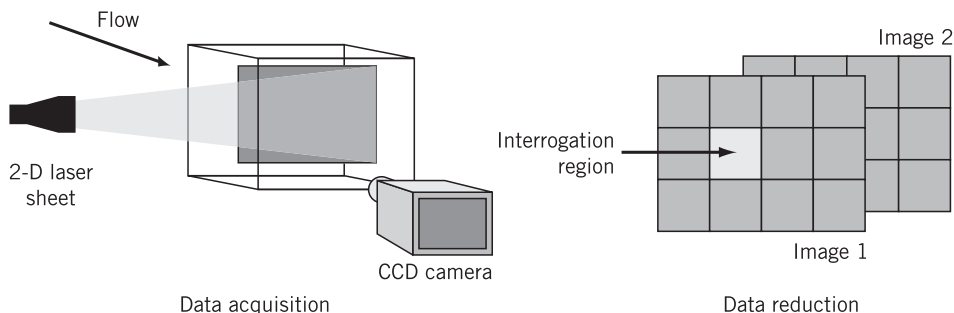


FIGURE 15.28 Basic layout of a digital particle image velocimeter.

The operating principle is based on particle displacement with time

$$\vec{U} = \Delta \vec{x} / \Delta t \quad (15.53)$$

where \vec{U} where \vec{U} is the instantaneous particle velocity vector based on its spatial position $\vec{x}(x,y,z,t)$. The camera records particle position at each flash into separate image frames. To obtain velocity data in a rapid manner, each image is divided into small areas, called interrogation areas. The corresponding interrogation areas between two images, I_1 and I_2 , are cross-correlated with each other, on a pixel-by-pixel basis. A particular particle movement from position \vec{x}^1 to \vec{x}^2 shows up as a signal peak in the correlation $R_{12}(\Delta \vec{x})$, where

$$R_{12}(\Delta \vec{x}) = \int_A \int I_1(\Delta \vec{x}) I_2(\vec{x} + \Delta \vec{x}) d\vec{x} \quad (15.54)$$

so identifying the common particle and allowing the estimate of the particle displacement $\Delta \vec{x}$. By repeating the cross-correlation between images for each interrogation area, a velocity vector map of the full area results.

A number of variations of this process have been developed but the concept remains the same. The technique works in gases or liquids. Three-dimensional information can be obtained by using two cameras. As with LDA, particle size and properties should be chosen relative to the fluid and flow velocities expected so that the particles move with the fluid, but large enough relative to camera pixel size to avoid peak-locking errors (Raffel et al., 2007). The maximum flow speed measurable is limited by the interrogation area size. The resolution depends on laser flash width and separation time, flow velocity, camera recording time, and image magnification. Raffel et al. (2007) discuss the technique, its variations, and error estimates.

15.8.5 Selection of Velocity Measuring Methods

Selecting the best velocity measuring system for a particular application involves a number of factors that an engineer needs to weight accordingly:

1. Required spatial resolution
2. Required velocity range
3. Sensitivity to velocity changes only
4. Required need to quantify dynamic velocity
5. Acceptable probe blockage of flow
6. Ability to be used in hostile environments
7. Calibration requirements
8. Low cost and ease of use.

When used under appropriate conditions, the uncertainty in velocity determined by any of the discussed methods can be as low as 1% of the measured velocity, although under special conditions LDA methods can have an uncertainty one order of magnitude lower (Goldstein and Kried, 1967).

15.8.6 Pitot-Static Pressure Methods

The pressure probe methods are best suited for finding the mean velocity in fluids of constant density. Relative to other methods, they are the simplest and cheapest method available to measure velocity at a point. Probe blockage of the flow is not a problem in large ducts and away from walls. Fluid particulate blocks the impact ports, but aspirating models are available for such situations. They are subject to mean flow misalignment errors. They require no calibration and are frequently used in the field and laboratory alike.

15.8.7 Thermal Anemometer

Thermal anemometers are best suited for use in clean fluids of constant temperature and density. They are well suited for measuring dynamic velocities with very high resolution. However, signal interpretation in strongly dynamic flows can be complicated (Rodi, 1975; Yavuzkurt, 1984). Hot-film sensors are less fragile and less susceptible to contamination than hot-wire sensors. Probe blockage is not significant in large ducts and away from walls. Thermal anemometers are 180° directionally ambiguous (i.e., flows from the left or right directions give the same output signal), an important factor in flows that may contain flow reversal regions. An industrial-grade system can be built rather inexpensively. The thermal anemometer is usually calibrated against either pressure probes or an LDA.

15.8.8 Laser Doppler Anemometer

The laser Doppler anemometer (LDA) is a relatively expensive and technically advanced point velocity measuring technique that can be used for most types of flows but is also well suited to hostile, combusting, or highly dynamic (unsteady, pulsatile, or highly turbulent) flow environments. It offers good frequency response, small spatial resolution, no probe blockage, and simple signal interpretation, but requires optical access and the presence of scattering particles. This method provides very good temporal resolution for time-accurate measurements in turbulent flows. The method measures the velocity of particles suspended in the moving fluid, not the fluid velocity, so careful planning is required in particle selection to ensure that the particle velocities represent the fluid velocity exactly. The size and concentration of the particles govern the system frequency response (Maxwell and Seaholtz, 1974; Dring and Suo, 1982).

15.8.9 Particle Image Velocimetry

Particle image velocimetry (PIV) is a relatively expensive and technically advanced full-field velocity measuring technique that can be used for most types of flows, including hostile and combusting flows. There is no probe blockage of the flow, but it requires optical access and the presence of scattering particles. The method provides an instantaneous snapshot of the flow, providing excellent views of flow structures. Time-dependent quantification of such dynamic flows is possible, but frequency bandwidth is limited to camera frame rate and spatial resolution. As with LDA, this method measures the velocity of particles suspended in the moving fluid, not the fluid velocity, so careful planning is required in particle selection to ensure that the particle velocities represent the fluid velocity exactly.

NOMENCLATURE

d	diameter (l)
e_i	elemental errors
d_f	fringe spacing (l)
f_D	Doppler frequency (Hz)
h	depth (l)
h_0	reference depth (l)
k	ratio of specific heats
p	pressure ($m^{-1}lt^{-2}$)
p_a	applied pressure ($m^{-1}lt^{-2}$)
p_{abs}	absolute pressure ($m^{-1}lt^{-2}$)
p_e	relative static pressure error ($m^{-1}lt^{-2}$)
p_i	indicated pressure ($m^{-1}lt^{-2}$)
p_m	measured pressure ($m^{-1}lt^{-2}$)
p_t	total or stagnation pressure ($m^{-1}lt^{-2}$)
p_v	dynamic pressure ($m^{-1}lt^{-2}$)
q	charge (C)
r	radius (l)
t	thickness (l)
y	displacement (l)
ρ	density (ml^{-3})
ι	time constant (t)
ϕ	latitude
ω_n	natural frequency (t^{-1})
z	altitude (l)
C	capacitance (F)
E	voltage (V)
E_m	bulk modulus of elasticity ($m^{-1}lt^{-2}$)
Gr	Grash of number
H	manometer deflection height (l)
K	static sensitivity
K_q	charge sensitivity ($m^{-1}lt^{-2}$)
K_E	voltage sensitivity ($Vm^{-1}t^{-2}$)
M	Mach number
Re_d	Reynolds number, $Re = Vd/\nu$
S	specific gravity
U	velocity (lt^{-1})
\forall	volume (l^3)
γ	specific weight ($MI^{-2}T^{-2}$)
ϵ	dielectric constant

λ	wavelength (l)
μ	absolute viscosity ($\text{MT}^{-1}\text{L}^{-1}$)
ν	kinematic viscosity (l^2/t)
ν_p	Poisson ratio

REFERENCES

- American Society of Mechanical Engineers (ASME), *PTC 19.2-1987: Pressure Measurement*. New York: ASME International; 1987.
- Brombacher WG, Johnson DP, Cross JL. Mercury Barometers and Manometers. *National Bureau Standards Monograph* 1960;8.
- Collis DC, Williams MJ. Two-dimensional convection from heated wires at low Reynolds numbers. *Journal of Fluid Mechanics* 1959;6.
- Cummins HZ, Pike ER, *Photon Correlation And Light Beating Spectroscopy*. Proceedings of the NATO ASI. New York: Plenum; 1973.
- Delio G, Schwent G, Cesaro R. Transient behavior of lumped-constant systems for sensing gas pressures, National Advisory Council on Aeronautics (NACA) TN-1988, 1949.
- Doebelin EO. *Measurement Systems: Application and Design*. 5th ed. New York: McGraw-Hill Science/Engineering/Math; 2003.
- Dring RP, Suo M. Particle trajectories in swirling flows. *Transactions of the ASME, Journal of Fluids Engineering* 104;1982.
- Durst F, Melling A, Whitelaw JH. *Principles and Practice of Laser Doppler Anemometry*. New York: Academic Press; 1976.
- Franklin RE, Wallace JM. Absolute measurements of static-hole error using flush transducers. *Journal of Fluid Mechanics* 42;1970.
- Freythuth P. A bibliography of thermal anemometry. *TSI Quarterly* 4;1978.
- Goldstein RJ, editor. *Fluid Mechanics Measurements*. 2nd ed. New York: CRC Press; 1996.
- Goldstein RJ, Kried DK. Measurement of laminar flow development in a square duct using a laser Doppler flowmeter. *Journal of Applied Mechanics* 1967; 34.
- Hetenyi M, editor. *Handbook of Experimental Stress Analysis*. New York: Wiley; 1950.
- Hinze JO. *Turbulence*. New York: McGraw-Hill; 1959.
- Hougen, J, Martin O, Walsh R. Dynamics of pneumatic transmission lines. *Control Engineering* 1963.
- Instrument Society of America (ISA), *A Guide for the Dynamic Calibration of Pressure Transducers*. ISA-37.16.01-2002, Instrument Society of America; 2002.
- King LV. On the convection from small cylinders in a stream of fluid: determination of the convection constants of small platinum wires with application to hot-wire anemometry. *Proceedings of the Royal Society, London* 1914; 90.
- Maxwell BR, Seaholtz RG. Velocity lag of solid particles in oscillating gases and in gases passing through normal shock waves, National Aeronautics and Space Administration TN-D-7490.
- McLeod HG. Vacuum gauge. *Philosophical Magazine* 1874;48:110, Reprinted in *History of Vacuum Science and Technology*, Madey T. E., Brown W. C., editors. American Vacuum Society, New York, 1984, p. 102–105.
- Migliavacca F. Multiscale modelling in an biofluidynamics: application to reconstructive paediatric cardiac surgery. *Journal of Biomechanics* 2006; 39(6).
- Munson B, Young D, Okishi T. *Fundamentals of Fluid Mechanics*. 6th ed. New York: Wiley; 2009.

- Raffel M, Willert CE, Wereiey ST, Kompenhans J. *Particle Image Velocimetry*. 2nd ed. Heidelberg: Springer; 2007.
- Rayle RE. Influence of orifice geometry on static pressure measurements, ASME Paper No. 59-A-234, 1959.
- Rodi W. A new method for analyzing hot-wire signals in a highly turbulent flow and its evaluation in a round jet, *DISA Information*. Denmark: Dantek Electronics; 1975. See also, Bruun, H. H., Interpretation of X-wire signals, *DISA Information*, Dantek Electronics, Denmark; 1975.
- Way S. Bending of circular plates with large deflection. *Transactions of the ASME* 1934; 56.
- Yavuzkurt S. A guide to uncertainty analysis of hot-wire data. *Transactions of the ASME, Journal of Fluids Engineering* 1984; 106.
- Yeh Y, Cummins H. Localized fluid flow measurement with a He–Ne laser spectrometer. *Applied Physics Letters* 1964; 4.

16

LUMINESCENT METHOD FOR PRESSURE MEASUREMENT

GAMAL E. KHALIL, JIM W. CRAFTON, SERGEY D. FONOV, MARVIN SELLERS, and DANA DABIRI

- 16.1 Introduction
- 16.2 Principles of pressure-sensitive paint
- 16.3 Pressure-sensitive luminescent dyes
- 16.4 PSP polymer and binder
- 16.5 Measurement methods
 - 16.5.1 Intensity-based measurements
 - 16.5.2 Reference method tool and binary pressure-sensitive paint
 - 16.5.3 Lifetime-based measurements
 - 16.5.4 PSP by singlet oxygen emission
 - 16.5.5 Error analysis
- 16.6 Pressure-sensitive paint measurements
 - 16.6.1 Research facilities
 - 16.6.2 Production wind tunnels
 - 16.6.3 Low-speed applications of pressure-sensitive paint
 - 16.6.4 Pressure-sensitive paint on rotating machinery
 - 16.6.5 Fast-responding pressure-sensitive paint
- Acknowledgments
- References

16.1 INTRODUCTION

Conjugated organic and organometallic molecules such as pyrene, ruthenium complexes, and platinum porphyrins exhibit luminescence quenching by oxygen (Birks, 1970; Khalil et al., 1989; Wolfbeis, 1991). These oxygen-sensitive luminophore can be

used to monitor oxygen concentrations in various environments through luminescence intensity or luminescence decay (lifetime) rate measurements. A number of oxygen sensors based on these luminophore are now commercially available, and are commonly used in a wide variety of applications (Hach Company, *Hach 9182 Dissolved Oxygen Analyzer*, Loveland, CO; Ocean Optics, Inc., *FOXY Fiber Optic Oxygen Sensors*, Dunedin, FL; Oxford Optronix, Ltd, *OxyLiteTM/OxyLabTMpO2 Sensors*, Oxford, UK). The University of Washington research group pioneered the combination of these luminophore with oxygen-permeable polymer matrices, commonly referred to as binders, to form pressure-sensitive paint. PSP made it possible to obtain high resolution, two-dimensional pressure distributions over aerodynamic surfaces in wind tunnel environments (Kavandi et al., 1990; Demas et al., 1999; Bell et al., 2001). PSP has a number of advantages over discrete pressure taps traditionally used on conventional wind tunnel models: (i) PSP measures pressure distributions over a surface area thereby allowing visualization of dynamic flow processes with high spatial resolution limited only by the imaging system, (ii) PSP allows measurement in regions where conventional pressure taps cannot be used, such as leading and trailing edges, (iii) PSP allows for experimentally guided Computational Fluid Dynamics, resulting in accurate simulations, and (iv) PSP allows for studies of surface flow features, such as shock location and boundary layer separation and reattachment. The experimental setup requires that the paint be either sprayed or airbrushed onto the surface of interest, and a light source emitting blue or UV light to excite the dye. The surface is imaged through a filter that isolates the excitation light from the pressure-sensitive luminescence of the paint. Each pixel on the camera then acts as a pressure tap, and therefore, continuous distributions of the pressure on the painted surface are acquired. Standard pressure paints can be used for mean measurements with response times of about 1 Hz, and fast response (Gregory et al., 2008) systems can be used for high frequency measurements, with a bandwidth of over 100 kHz.

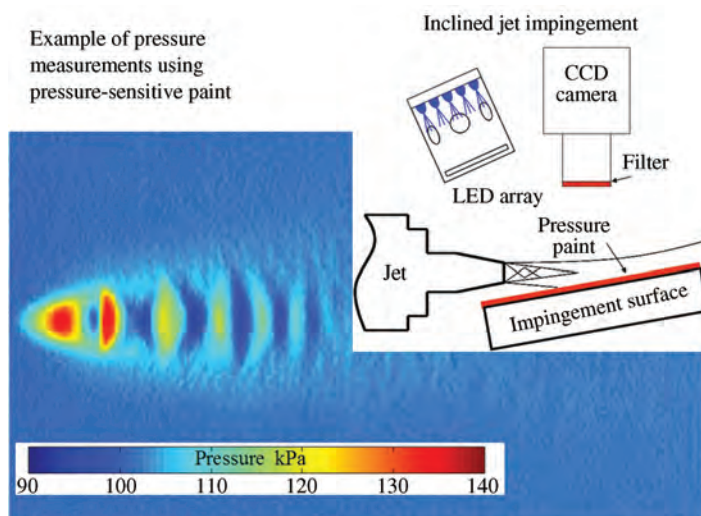


FIGURE 16.1 Example of PSP setup. (See the color version of this figure in Color Plates Section.)

An example of the utility of PSP is shown in Figure 16.1. Here, a jet is impinging onto a flat surface at an inclined angle. While this is a simple experimental setup, the resulting flow-field can produce a complex series of shock and expansion waves on the impingement surface if the jet is operated in the sonic under-expanded regime. In this under-expanded regime, slight changes in the jet pressure ratio can have a dramatic effect on the magnitude and location of these waves. It would be impractical and intrusive to pressure tap the impingement surface. PSP, however, can easily be applied to this problem and produce data with the spatial resolution necessary to resolve the pressure field.

The use of PSP is becoming more common in large transonic tunnels, with production systems in use in several facilities such as AEDC (Ruyten and Sellers, 2004), ARA (Vardaki et al., 2010), TsAGI (Mosharov et al., 1998), and DLR (Engler et al., 2005). Fast-responding PSP has been used in unsteady aerodynamic applications, such as airflow over rotor blades (Bencic, 1998). An excellent review of unsteady aerodynamic applications of PSP has been recently presented by Gregory et al. (2008). The PSP for insect flight has also been evaluated in the Callis laboratory (McGraw, 2004.) with the ultimate goal of mapping the pressure distribution over an insect's wings in flight. Paint formulations, test equipment, and software are now available commercially by Innovative Scientific Solutions, Inc. (Dayton, OH).

16.2 PRINCIPLES OF PRESSURE-SENSITIVE PAINT

PSP is made of an oxygen-sensitive luminescent molecule that is incorporated into an oxygen-permeable polymer binder and dissolved in a volatile solvent to form a paint that can be easily applied to surfaces (Figure 16.2). Exposing the luminescent molecule, or luminophor, to light of an appropriate wavelength places it in an excited state with a finite lifetime. The excited-state luminophor will eventually release its energy primarily by emitting photons or by transferring its energy to a diatomic oxygen molecule. The energetics of this process is presented in a Jablonski diagram in Figure 16.3. The luminophor begins in a singlet ground state, 1S_0 , and is excited into a singlet excited state, 1S_x , upon exposure to UV radiation. It then rapidly decays nonradiatively to its lowest singlet excited state, 1S_1 . From here, the luminophor can transfer its energy in one of three ways. First, it can lose the energy nonradiatively in the form of heat through a first order rate process represented by the rate constant, k_1 . It can also

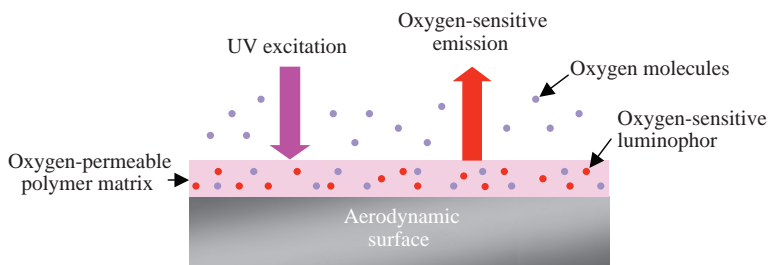


FIGURE 16.2 A schematic depicting the application of pressure-sensitive paint. (See the color version of this figure in Color Plates Section.)

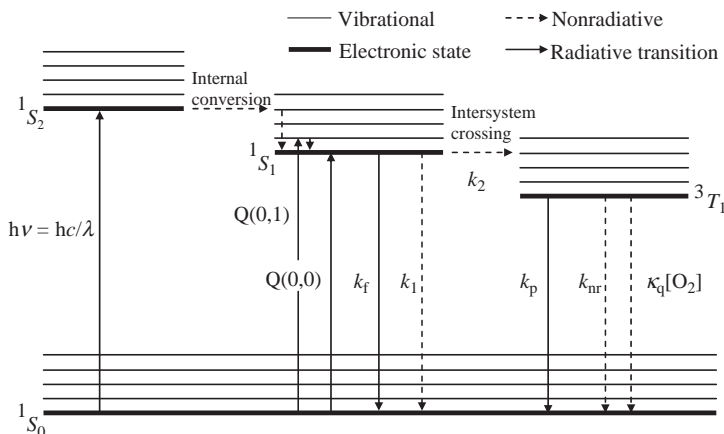


FIGURE 16.3 A general Jablonski diagram for oxygen-sensitive luminophors.

lose its energy by emitting photons in a process known as fluorescence (k_f). Alternatively, the energy can be transferred from the singlet excited state to a triplet excited state ($3T_1$), through intersystem crossing (k_2).

At this point, the luminophor will relax back to its singlet ground state through non-radiative heat transfer (k_{nr}), through the emission of photons in the form of luminescent (k_p), or through energy transfer to molecular oxygen. Many oxygen-sensitive luminophors are organometallic compounds that feature a heavy, metallic central atom surrounded by an organic ligand. The direct result of this arrangement is strong spin-orbital coupling which helps the molecule to overcome the forbidden transition. Because of this, the triplet state is a relatively long-lived species, and the luminescent decay can occur over a period approaching milliseconds, depending on the exact luminophor (Lakowicz, 1983). This long lifetime allows energy transfer to molecular oxygen to take place through molecular collisions. The rate of energy transfer is directly proportional to the concentration of oxygen present in the surrounding environment, and can be represented by $\kappa_q(O_2)$, where κ_q is the pseudo-first order bimolecular quenching constant and (O_2) is the concentration of oxygen. When the concentration of oxygen is high, collision quenching of the triplet state molecule with oxygen becomes significant. As a result, more energy is transferred to oxygen than is released as photons. Therefore, the triplet state lifetime, τ , can be defined as the inverse of the sum of the three energy decay processes:

$$\tau = (k_{nr} + k_p + \kappa_q[O_2])^{-1} \quad (16.1)$$

The concentration of oxygen, in turn, can be directly related to pressure through Henry's law:

$$[O_2] = \sigma P_{O_2} \quad (16.2)$$

where σ is the Henry's Law constant for oxygen, and P_{O_2} is the partial pressure of oxygen. There are two common ways of quantifying the amount of luminescent intensity being emitted by the luminophor, and both will be described in Section 16.5.

16.3 PRESSURE-SENSITIVE LUMINESCENT DYES

Different types of luminescent dyes are available for measuring oxygen concentration and, hence, pressure. By selecting the appropriate dye to be incorporated in the particular matrix, one can produce PSP for the specific application. This method is based on using the dyes with different excited state lifetime. The rate of the luminescent quenching is directly proportional to the oxygen concentration and the luminescent dye in the excited state. Therefore, the longer the lifetime of the luminescent dye, the higher the probability of being quenched by an oxygen molecule. Similarly, a higher concentration of oxygen dissolved in the polymer corresponds to a higher probability of collisional quenching. Thus, a luminescent dye with a long excited state lifetime is more optimal for lower oxygen concentration, while a short-lived luminescent dye is more optimal for higher oxygen concentration. In principal, it is thereby possible to “tune” the sensitivity for detecting a certain oxygen level by the proper choice of the particular polymer or matrix that has certain oxygen concentration and matching it to a particular dye. The lifetime of the luminophor and the diffusion rate of oxygen in the polymer are matched so that there is a measurable change in the luminescence level for the pressure range to be monitored.

University of Washington has done much research on the particular class of luminescent dyes known as porphyrins. Porphyrins and their derivatives have gained much momentum in chemical sensors application (McDonagh et al., 2008; Di Natale et al., 2010; Amao and Okura, 2009; Khalil et al., 2010; Gewehr, and Delpy, 1993). The electronic structure of the inner 16-membered ring with 18 electrons is responsible for the main features of their optical properties (Khalil et al., 1989). Porphyrin is highly symmetrical molecule with two pairs of low energy excited states, the quasi-forbidden Q states in the visible (650–500 nm) and the strongly allowed B states in the near UV (440–380 nm). Any chemical substitution on the porphyrin ring will reduce the symmetry and has the possibility of affecting the Q-transition intensity as well as its spectral position, displacing the transition to a longer wavelength. Symmetry breaking and symmetry reduction have immediate effects on the spectral properties of porphyrin ring. An important discovery was the high quantum yield and short lifetime of the triplet state of platinum (II) porphyrins (PtP), which has promoted many practical applications. It is believed that the strong electronic interaction between the platinum filled *d* orbital and the porphyrin empty $\pi\pi$ molecular orbitals, and the strong spin of Pt coupling produced approximately 100% triplet quantum yield and 90% phosphorescent emission yield. The concept of using PtP phosphorescence emission changes to study oxygen concentration was originally developed to monitor the oxygen concentration in blood (Khalil et al., 1989). It was originally conceived as a tiny volume of oxygen permeable polymer with luminescent PtP at the end of an optical fiber.

Porphyrins, such as other aromatic molecules, undergo dark and photo oxidation that causes degradation and changes to the spectral properties thus reducing its use life. Porphyrin reactivity with oxygen is influenced by the inductive effects of the functional group attached. To address this problem the photo oxidation process is reduced by the use of an electron withdrawing group such as fluorine. Perfluoro substitutions on porphyrins have been shown to reduce the electron density and increase the photostability (Khalil et al., 1989).

Table 16.1 lists pressure-sensitive dye examples from the porphyrins family as well as other known dyes. For each dye listed, the lifetime, relative quantum yield and wavelengths of operation are reported.

TABLE 16.1 List of Pressure-Sensitive Dyes

Pressure-Sensitive Dye	Lifetime ms (77 K)	Relative Yield (0% O ₂)	Wavelength (nm) λ_{ex} = Excitation λ_{em} = Emission
Platinum octaethylporphine (Kavandi et al., 1990; Gouterman, 1978; Kadish et al., 2000)	0.12	90	λ_{ex} 380, 500, 535 λ_{em} 650
Platinum tetra(pentafluorophenyl)porphine (Khalil et al., 1989; Khalil, 2002)	0.12	90	λ_{ex} 390, 506, 540 λ_{em} 650
Platinum tetra(pentafluorophenyl) porpholactone (Khalil, 2002)	0.072	60	λ_{ex} 396, 536, 574 λ_{em} 733
Platinum tetrabenzotetraphenylporphine (Borek et al., 2007)	0.073	70	λ_{ex} 430, 611 λ_{em} 765
Palladium octaethylporphine (Gouterman, 1978; Kadish et al., 2000)	1.93	50	λ_{ex} 390, 510, 545 λ_{em} 660
Palladium tetra(pentafluorophenyl)por- phine (Khalil et al., 1989; Khalil, 2002)	1.58	20	λ_{ex} 406, 519, 552 λ_{em} 660
Palladium tetra(pentafluorophenyl) porpholactone (Khalil et al., 1993; Gouterman et al., 1989)	1.1	20	λ_{ex} 412, 544, 584 λ_{em} 758
Galadonium octaethylporphine (Kavandi et al., 1990; Gamal and Khalil, 2007)	0.085	55	λ_{ex} 410, 540 λ_{em} 700, 786, 836
Galadonium tetra(pentafluorophenyl) porphine (Borek et al., 2007)	0.14	20	λ_{ex} 420, 550 λ_{em} 726, 810, 922
Tetra(pentafluorophenyl)porphine (Spellane et al., 1980; Wan, 1993; Baron et al., 1993)	0.00001	12	λ_{ex} 410, 505, 535, 580 λ_{em} 700
Thallium tetraphenylporphine (Khalil, 1973)	2.5	45	λ_{ex} 400, 527, 565 λ_{em} 712
Pyrene (Xu et al., 1995)	0.0005	80	λ_{ex} 350 λ_{em} 400
Ruthenium (Xu et al., 1994) tris(4,7- diphenyl-1,1-phenanthroline)Cl ₂	0.005	50	λ_{ex} 460 λ_{em} 610
Osmium (Carlson et al., 2009) tris (bathophenanthroline)dichloride	0.002	37	λ_{ex} 395 λ_{em} 620
Iridium (Fischer et al., 2009) tris(2-(beilzo [b]thiopliene-2-yl)pyridine)	0.0086	—	λ_{ex} 922, 366, 408 λ_{em} 596, 645

16.4 PSP POLYMER AND BINDER

PSP coatings consist of a luminophor dispersed in an oxygen-permeable polymer matrix or binder (see Figure 16.4). The choice of the polymer and matching it to the proper pressure-sensitive dye has been one of most challenging development for the PSP research community. PSP carriers can include polydimethylsiloxane, poly(vinyl chloride), poly(methyl methacrylate), polystyrene, poly(vinyl acetate), silica particles and anodized aluminum, as well as others (Gregory et al., 2008).

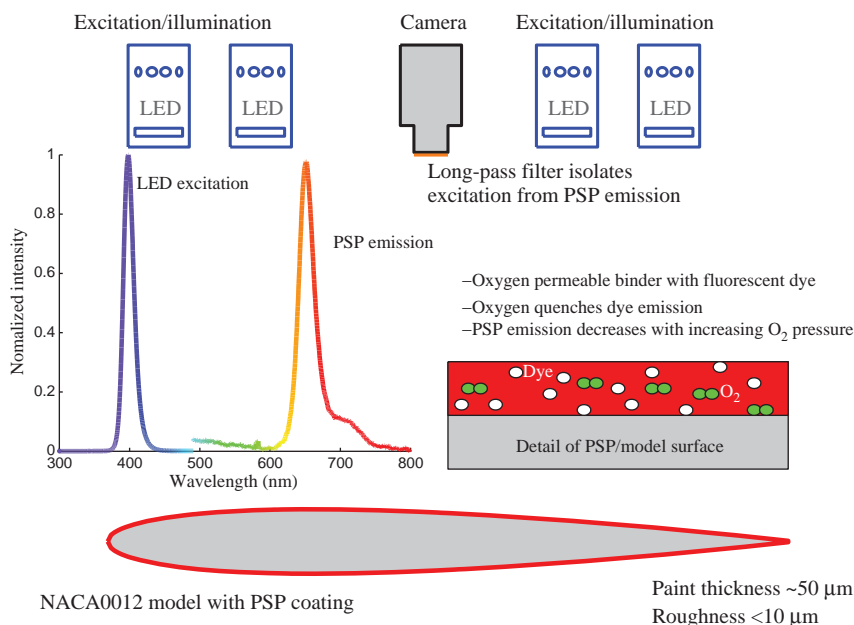


FIGURE 16.4 Basic PSP system showing dye/binder details. (See the color version of this figure in Color Plates Section.)

One of the most persistent problems of a luminescent oxygen sensing method, and hence for pressure-sensitive paint measurement, is the inherent temperature dependence. From the initial studies of PSP, it was clear that PSP accuracy requires temperature correction.

Temperature dependence is caused by several physical processes, such as temperature dependent oxygen permeability in the paint and thermal quenching of the luminescent probe. For most pressure-sensitive paints, the temperature sensitivity of the paint is a function of pressure, and the pressure sensitivity of the paint is a function of temperature. To solve this problem, a new polymer, FIB, was introduced (Carlson and Gouterman, 1997; Puklin et al., 2000), which is a random copolymer of heptafluoro-*n*-butyl methacrylate and hexafluoroisopropyl methacrylate. The FIB polymer has two important attributes: it consistently produces smaller temperature dependence values than other commonly used polymers; and most importantly, it has the same temperature dependence at vacuum and at atmospheric pressure. It was also shown that the small temperature dependence of FIB is due to its low activation energy of quenching and is relatively constant over the temperature range of interest (10–50°C) (Gouin and Gouterman, 2000). This allows for a simple temperature dependence correction.

The luminescence intensity $I(P, T)$ of Pt porphyrin in FIB polymer has been shown to have particularly simple temperature dependence:

$$I(P, T) = I_a(P)I_b(T) \quad (16.3)$$

This is shown by the superposition of the temperature dependence $I(T)$ taken at vacuum and at one atmosphere.

$$I(P_o, T)/I(P, T) = A + B(P/P_o) \quad (16.4)$$

TABLE 16.2 University of Washington PSP History of Pressure-Sensitive Paint

	1990	2000	2005
Sensor (Lakowicz, 1983; Jockusch et al., 2008)	Pt porphyrin	Fluorinated Pt porphyrin	Dual fluorinated Pt/Mg porphyrins
Matrix (Baron, 1996)	Silicone polymer	Fluorinated polymer (FIB)	Fluorinated polymer (FIB)
Pressure	1.0–0.3 Bar	1.0–0.05 Bar	1.0–0.001 Bar
Dynamic range			
90% Response time (ms)	2500	10	0.07
Temp. coeff. (Khan, 1985) (% int./deg.)	–2.5	–0.6	–0.05
Temp error (Baron, 1996) (mbar/deg.)	34	10	1
Applications (Mosharov et al., 1998)	High speed (airfoil) $M > 0.3$, $Re = 10^6$	Low speed (automobile) $M > 0.1$, $Re = 10^5$	Kinetic low speed (insect) $M \sim 0.02$, $Re = 10^3$

Over the past 15 years, the versatility and performance of pressure-sensitive paint has greatly increased. These improvements are described in Table 16.2.

16.5 MEASUREMENT METHODS

The relationship between surface illumination and paint luminescence is linear; therefore, any change in surface illumination will result in an equal change in paint luminescence. Errors in pressure measurements caused by variations in surface illumination can stem from several sources. Consider utilizing a point source for the illumination of a surface, as shown in Figure 16.5. The relationship between illumination intensity at a point on the surface and the distance between the source and the point of interest are an inverse function of the distance squared. Any movement of the painted surface or illumination source will result in a change in the distance between these two points, and thus a change in the illumination intensity at the surface. This movement can result from deformation of the model surface (bending or twisting of the wing due to aerodynamic loading), or physical displacement of the model (movement of the model on the sting due to aerodynamic loading). Even small deformations or displacements of the model can produce large errors in pressure measurements, as demonstrated in Figure 16.5. Another source of illumination errors is physical movement or temporal stability of the illumination source. Any variation of the intensity of the illumination source between the wind-off and the wind-on images will register as an error in illumination. Regardless of the source, illumination errors must be eliminated for the acquisition of quantitative PSP data, especially at low speeds.

For radiometric PSP, errors in pressure measurements due to temperature are largely the result of changes in the temperature of the model between the acquisition of the wind-off and the wind-on image. Any temperature gradient on the model surface, however, will result in a temperature-induced error in the pressure measurements. These temperature gradients can be the result of model construction, tunnel operation, or fluid dynamics. A rapid prototype model, for example, is constructed using an internal metal structure and a

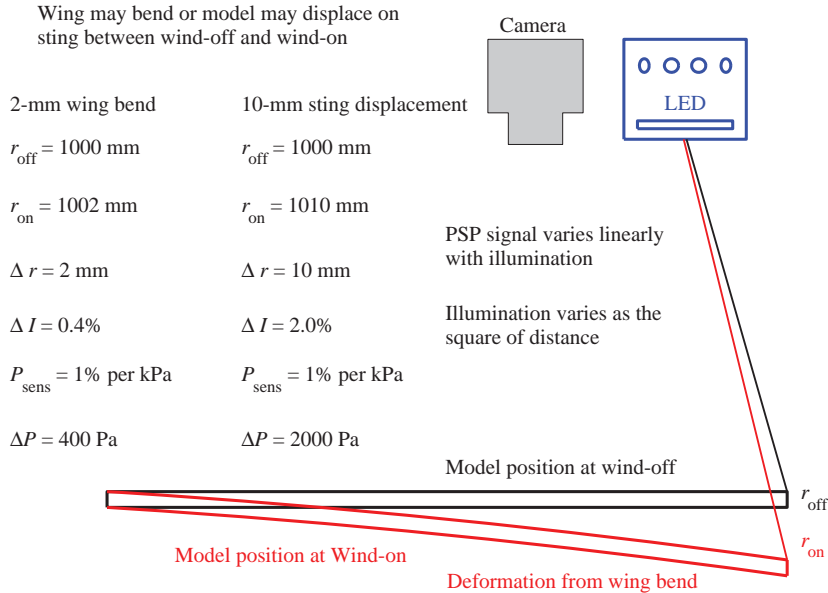


FIGURE 16.5 Illumination error. (See the color version of this figure in Color Plates Section.)

polymer resin. The thermal signature of the internal structure is apparent when the surface of the model is subjected to a heat flux. The model is exposed to a heat flux due to changes in tunnel Mach number, for example, this condition is most apparent during tunnel startup. Temperature errors can be minimized by using temperature-controlled tunnels and constructing the model from materials with high thermal conductivity. In fact, model construction and tunnel operation are key considerations for effective PSP measurements.

16.5.1 Intensity-Based Measurements

The intensity-based method commonly uses a CCD camera to directly measure the luminescent emission intensity, I , of the PSP while as it is being continuously excited by a light source of appropriate wavelength dictated by the choice of luminophor (Gouterman, 1997). For example, platinum octaethylporphyrin (PtOEP) features a strong absorption band at 390 nm, representing the $^1S_0 \rightarrow ^1S_x$ energy transition. This band is commonly known as the Soret band. Additionally, two weaker absorbing bands, known as the Q-bands, appear at 500 nm and 535 nm, and these represent transitions from the singlet ground state to different vibrational groups of the first singlet excited state, 1S_1 (Gouterman, 1997). Therefore, excitation at any of these wavelengths will be sufficient to result in an oxygen-sensitive emission centered at 650 nm.

The amount of phosphorescent emission will be directly related to the number of triplets produced in the intersystem crossing process. The triplet yield, Φ_T , is defined as the number of triplets produced divided by the total number of photons absorbed:

$$\Phi_T = \frac{k_2}{k_2 + k_1 + k_f} \quad (16.5)$$

For PtOEP, it has been shown that Φ_T is nearly 100%.

The triplet yield, in turn, dictates the phosphorescent quantum yield, Φ_P , which is the product of the triplet yield multiplied by the ratio of photons emitted as phosphorescence over the total number of photons absorbed:

$$\Phi_P = \Phi_T \frac{k_p}{k_p + k_{nr} + \kappa_q [O_2]} \quad (16.6)$$

Since (O_2) is directly proportional to pressure, Equation (16.4) becomes

$$\Phi_P = \Phi_T \frac{k_p}{k_p + k_{nr} + \kappa_q \sigma P_{O_2}} \quad (16.7)$$

Ultimately, the phosphorescence quantum yield is directly proportional to the emission intensity of PtOEP. Assuming that the temperature in the environment surrounding the PtOEP remains constant, the rates of phosphorescence (k_p) and collision quenching ($\kappa_q [O_2]$) will be the main parameters that determine the amount of intensity emitted. These processes are inversely proportional—that is, the higher the rate of collisional quenching by oxygen, the less energy remains for phosphorescent emission. In other words, higher oxygen concentrations results in less phosphorescence. Again, through Henry's Law, oxygen concentration can be directly related to pressure. Therefore, quantitatively measuring the phosphorescent emission of PtOEP in a PSP can be used to provide pressure measurements on aerodynamic surfaces.

A ratiometric calibration process is employed to obtain quantitative pressure measurements from emission intensities. First, an initial intensity measurement, I_o , is obtained at a reference pressure, P_o (usually 1 atm). The pressure surrounding the PSP is then changed, and its intensity response is measured. This response can then be quantitatively described by the Kavandi equation (Equation 16.8):

$$\frac{I_o}{I} = \frac{\Phi_o}{\Phi_p} = \frac{k_p + k_{nr} + \kappa_q \sigma P}{k_p + k_{nr} + \kappa_q \sigma P_o} = A + B \frac{P}{P_o} \quad (16.8)$$

$$A = \frac{k_p + k_{nr}}{k_p + k_{nr} + \kappa_q \sigma P_o} \quad (16.9a)$$

$$B = \frac{\kappa_q \sigma P}{k_p + k_{nr} + \kappa_q \sigma P_o} \quad (16.9b)$$

where A and B are known as the Kavandi parameters and are experimentally determined. These parameters are then used to directly equate measured intensities with pressure.

In practice, since PSP is a continuous layer that can cover large two-dimensional test surfaces, pressure distributions with high spatial resolutions over those surfaces can be obtained. This is accomplished by illuminating the PSP with a light source emitting the appropriate wavelength, such as a xenon lamp fitted with a 390 nm bandpass filter. PSP intensity measurements (I_o and I) can be obtained with a CCD fitted with a 650 nm bandpass filter to isolate only the photons resulting from the phosphorescent emission of the PSP. Postprocessing of the images yields I_o/I ratios over the entire imaged region. The Kavandi parameters obtained during calibration of the PSP can then be used to calculate the pressure distribution for the surface.

The intensity ratio serves a dual purpose. In addition to providing a linear relationship between intensity and pressure, intensity images that are compared with a reference image can correct for variations in excitation illumination, paint thickness, or luminophor concentration within the polymer matrix of the paint. However, the reference image requirement can be a significant limitation toward PSP applications, as it is necessary that the test surface remain static between the times that the reference image was obtained and the test images were captured. This approach cannot be used for moving objects, where it may be impossible to replicate the exact illumination conditions for both the reference and test images.

16.5.2 Reference Method Tool and Binary Pressure-Sensitive Paint

One means of dealing with sources of error such as illumination and temperature is to employ a reference probe, and thus, produce a binary pressure-sensitive paint. Several research teams (Bykov et al., 1997; Klein et al., 1999; Khalil et al., 2000) have successfully demonstrated this approach.

Several spectroscopic methods for measuring oxygen concentration or pressure are available that provide an internal referencing. These methodologies are designed to correct for many problems that exist with intensity measurement. Internal reference methods avoid several problems that arise with intensity measurement such as light source changes, index of refraction, optical geometry variations and fluctuations in film thickness, and dye concentration.

Advances in PSP technology have provided a solution to this limitation: a second, pressure-insensitive luminophor could be paired with a pressure-sensitive luminophor to result in self-referencing PSP (Khalil et al., 2004). Sometimes known as binary, or dual-luminophor PSP, the measurement relies on using the pressure-insensitive luminophor to provide reference emission intensity. Ideally, this reference luminophor can be excited at the same wavelength as the pressure-sensitive luminophor, and will have an emission that can be separately resolved.

An example of this is would be a binary-luminophor PSP using platinum tetra(pentafluorophenyl)porpholactone (PtTFPL) as a pressure sensor, and magnesium tetra(pentafluorophenyl)porphyrin (MgTFPP) as a reference. Both luminophors can be excited at 400 nm; PtTFPL provides the pressure-sensitive emission at 733 nm and MgTFPP provides the reference emission at 615 nm. Dual-luminophor PSP eliminates the need for a separate wind-off intensity measurement, as the reference intensity can be simultaneously imaged with the sensor intensity, usually through the use of a second CCD fitted with the appropriate bandpass filter. The binary luminophor PtTFPL and MgTFPP in the FIB polymer produced “ideal” PSP measurements with pressure sensitivity of 4.5% per psi and a temperature dependency of less than $-0.1\%/^{\circ}\text{C}$. The temperature dependency of the dual luminophor intensity ratio ($I_{\text{ref}}/I_{\text{sen}}$) is substantially reduced compared with the use of $I(\text{PtTFPP})$ alone which has a temperature dependence in FIB of $0.60\%/^{\circ}\text{C}$. The use and benefits of binary-luminophor PSP have been demonstrated numerous times, including a study performed by Youssef Mébarki of the NRC of Canada on low speed (20–40 m/s) automotive testing (Gouterman et al., 2004). In general, using a ratiometric method of the two intensities produces a more robust pressure measurement. Binary-luminophor PSP has since been commercialized by Innovative Scientific Solutions, Inc.

16.5.3 Lifetime-Based Measurements

Once illumination errors caused by model movement were identified as a significant source of error for Radiometric PSP systems, several teams began investigating lifetime techniques for wind tunnel applications. Approaches to measuring luminescent lifetimes for PSP have included phase-sensitive detection (Holmes, 1998) and multigate integration (Goss et al., 1999). Among the advantages of the lifetime approach is the elimination or minimization of illumination errors. Theoretically, all data are acquired at the wind-on condition and this minimizes or eliminates illumination as a source of error in the measurement. Errors due to temperature are still a source of uncertainty in lifetime-based PSP measurements as is shown by a calibration of UF-400 using the two-gate approach in Figure 16.6.

The practical implementation of the lifetime approach involves multigate integration using pulsed LEDs and frame transfer cameras with on-chip accumulation. A schematic of the two-gate lifetime approach is shown in Figure 16.7. The paint is excited to fluoresce using a short illumination pulse. Data are acquired in two distinct windows, in this case the first of these windows occurs during the illumination pulse and the second is centered on the decay of the fluorescence. The ratio of the integrated signal from the two gates is computed and the resulting signal is plotted versus pressure. The position and width of the gates is selected to maximize the systems sensitivity to pressure while maintaining a favorable signal to noise ratio.

A second technique measures the phosphorescent lifetime of the pressure-sensitive luminophor (Bell et al., 2001; Coyle, 1999). First, the PSP is excited with a quick burst of light at the appropriate wavelength from a pulsed light source, forcing the pressure-sensitive luminophor into its excited state. The luminophor will then lose its energy by emitting photons or transferring the energy to molecular oxygen. The loss of energy in the system ideally follows an exponential decay of the form:

$$I = I_0 e^{-t/\tau} \quad (16.10)$$

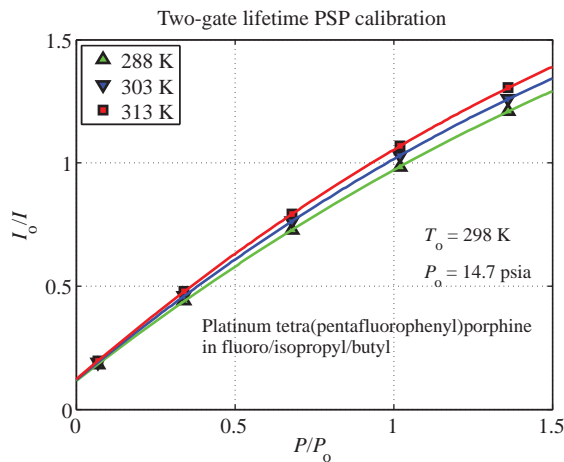


FIGURE 16.6 Lifetime calibration of UF-400 PSP using two-gate integration. (See the color version of this figure in Color Plates Section.)

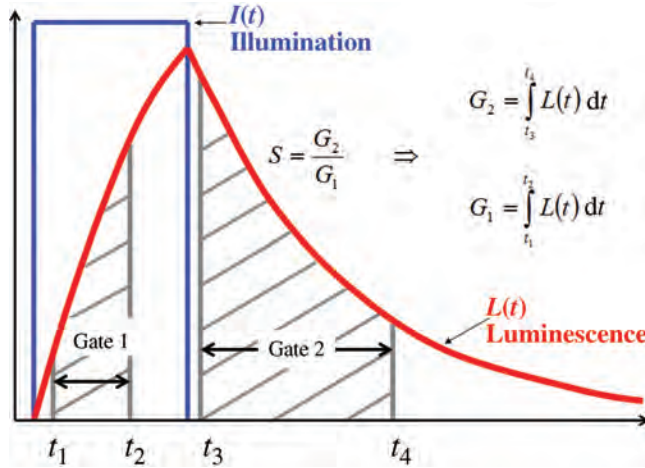


FIGURE 16.7 Two-gate lifetime measurement approach. (See the color version of this figure in Color Plates Section.)

where I_0 represents the initial intensity of luminophor, t is time, and τ is the lifetime of the luminophor. As the lifetime is inversely proportional to the oxygen concentration of the environment, it can provide a measure of pressure. Traditionally, lifetimes of luminescent systems are measured using photodiodes or photomultiplier tubes designed to collect emission intensity in a time dependent manner. However, these detectors cannot provide spatially resolved lifetime data over two-dimensional surfaces.

Fortunately, it is possible to use a CCD to obtain such lifetime distributions. It can be shown that by dividing the phosphorescent decay curve into two regions (I and II) split at time t_d (Figure 16.8) and integrating each region, a lifetime value can be calculated using (Baron, 1996)

$$\tau = \frac{t_d}{\ln\left(\frac{I + II}{II}\right)} \quad (16.11)$$

where I and II are the integrated areas of regions I and II, respectively. Specially designed CCDs have the ability to obtain two images in rapid succession, separated by times as small as 200 ns. Known as the double image feature (DIF) (Terminology of Princeton Instruments MicroMAX CCD), it allows the user to set the exposure time of both windows, which can be as short as 1 μ s. As each image effectively integrates the emission intensity of their respective portions of the luminescence decay curve, it is possible to calculate the spatially resolved lifetime distributions, and thus the pressure distribution, over the test surface.

One of the attractive features of lifetime PSP is the potential for near real-time data presentation. Since there is no significant movement or deformation of the model between gate 1 and gate 2, conversion of the images to pressure requires only simple math. The background is subtracted from each gate, the ratio of gate 2 over gate 1 is computed, and this ratio is converted to pressure using a calibration. For the final data, the PSP data must be corrected using the pressure taps; however, the preliminary data are sufficiently accurate for real-time investigation and decision making.

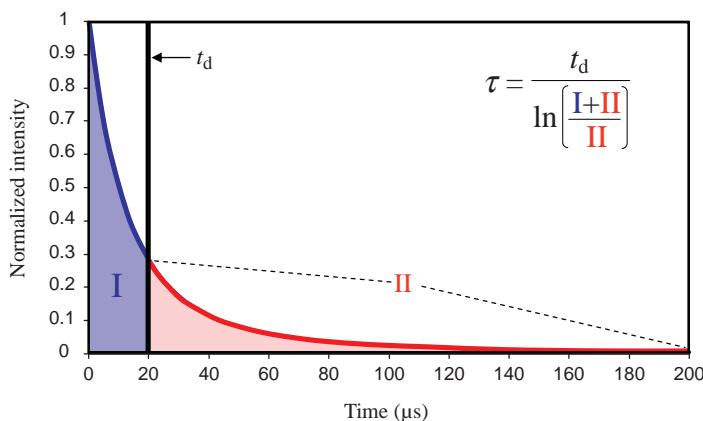


FIGURE 16.8 A plot displaying the integrated regions of the phosphorescent decay curve utilized in modified rapid lifetime determination. (See the color version of this figure in Color Plates Section.)

Modern lifetime systems use pulsed LEDs for illumination and CCD cameras with signal accumulation for detection. The LED can be pulsed repetitively at several kHz and the camera can gate and integrate the fluorescent signal from the PSP. This combination of LEDs and signal accumulation is necessary as the number of photons from the PSP is limited. Generally, it is necessary to accumulate signal for several thousand pulses on each gate.

As lifetime determination of PSP is independent of lighting conditions, coating uniformity or concentration variations, it is often seen as being a more robust method of measuring pressure, as opposed to intensity-based measurements. However, lifetime-based measurements involve more complicated hardware setups and can be more difficult to obtain experimentally.

16.5.4 PSP by Singlet Oxygen Emission

PSP performance may be validated if it is compared to a direct oxygen concentration measurement method. We examined the possibility to measure oxygen concentration directly by looking at oxygen alone.

The ground state of molecular oxygen is a triplet, $^3\Sigma_g^-$. Its two lowest excited states are $^1\Delta_g$ and $^1\Sigma_g^+$, known as singlet molecular oxygen, which lie $7,882\text{ cm}^{-1}$ ($\sim 1,270\text{ nm}$) and $13,121\text{ cm}^{-1}$ ($\sim 760\text{ nm}$) above the ground state, respectively. In free molecular oxygen, $^3\Sigma_g^- \rightarrow ^1\Delta_g$ and $^3\Sigma_g^- \rightarrow ^1\Sigma_g^+$ transitions are very weak since these transitions are both electric dipole and spin forbidden. The production of singlet oxygen by direct stimulation is very inefficient, but the process of triplet quenching of PtP produces singlet oxygen very effectively (Jockusch et al., 2008). The radiative lifetimes of singlet oxygen are 45 min for $^1\Delta_g$ and 11 s for $^1\Sigma_g^+$. The observed lifetimes for $^1\Delta_g$ range from milliseconds in gaseous phase to microseconds in aqueous media (Khan, 1985; Wasserman and Murray, 1979; Schaap, 1976). The lifetime of $^1\Sigma_g^+$, on the other hand, is much shorter since $^1\Sigma_g^+$ is more reactive and quickly decays into $^1\Delta_g$ or $^3\Sigma_g^-$. With the availability of InGaAs NIR detectors, detection of $^1\Delta_g$ emission becomes much more feasible (Zebger et al., 2004). We have shown that the use

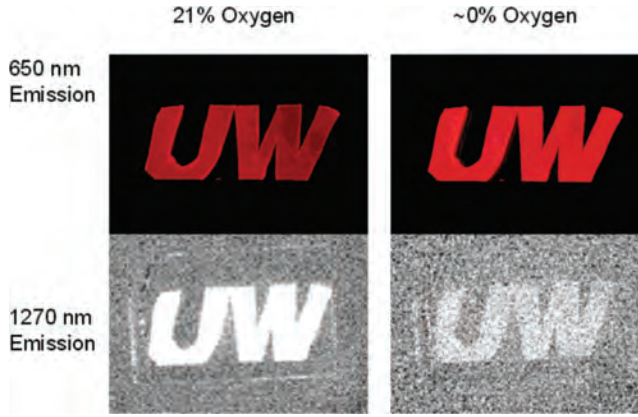


FIGURE 16.9 Images of 650 and 1270 nm emissions of UW logo. (See the color version of this figure in Color Plates Section.)

of $^1\Delta_g$ emission at 1270 nm can be a probe for direct oxygen concentration measurement (Khalil et al., 2005). For example, the emission collected from a PtTFPP film excited at 400 nm has a 650 nm phosphorescence band and a 1270 nm singlet oxygen emission band. Luminescence images at 650 nm and 1270 nm of a PtTFPP film shaped in the form of “UW” taken at two oxygen concentrations are given in Figure 16.9. Figure 16.9 shows the 1270 nm oxygen emission under air (21% oxygen) is bright and much weaker near 0% oxygen. On the other hand, images at 650 nm have little luminescence under air and brighter luminescence under vacuum. Thus, the two images would appear to be exchanged. This study shows qualitatively that the concentration of singlet oxygen over the surface of the film can be measured by the oxygen emission intensity as well as by the luminescence quenching of the PtTFPP.

16.5.5 Error Analysis

The sources of uncertainty for PSP measurements have been investigated and modeled by Liu and Sullivan (2004). These error sources include temperature, illumination, model displacement, model deformation, sedimentation, paint photo-degradation, stray light, and camera shot noise. While each of these error sources can be a significant problem in certain situations, those errors associated with model displacement and deformation as well as temperature are the major contributors in a large low-speed tunnel.

16.5.5.1 Stern–Volmer Equation The relationship between temperature, pressure, and intensity is given by the aerodynamic Stern–Volmer equation. In this equation T_{ref} is the reference temperature, P_{ref} is the reference pressure, and I_{ref} is the paint intensity at the reference condition

$$\frac{I_{\text{ref}}}{I} = A(T) + B(T) \frac{P}{P_{\text{ref}}} \quad (16.12)$$

The Stern–Volmer coefficients are temperature dependent because temperature affects both nonradiative deactivation and oxygen diffusion in a polymer. The Stern–Volmer

coefficients can be approximately expressed as a linear function of temperature as

$$\begin{aligned} A(T) &= A(T_{\text{ref}}) \left[1 + \frac{E_{\text{nr}}}{RT_{\text{ref}}} \left(\frac{T - T_{\text{ref}}}{T_{\text{ref}}} \right) \right] \text{thermal quenching} \\ B(T) &= B(T_{\text{ref}}) \left[1 + \frac{E_{\text{p}}}{RT_{\text{ref}}} \left(\frac{T - T_{\text{ref}}}{T_{\text{ref}}} \right) \right] \text{oxygen diffusivity} \end{aligned} \quad (16.13)$$

In this equation, E_{nr} is the Arrhenius activation energy for thermal quenching, E_{p} is the activation energy for oxygen diffusion, and R is the gas constant. These equations suggest that the temperature dependence of $A(T)$ is due to thermal quenching while the temperature dependence of $B(T)$ is related to the diffusivity of oxygen in the polymer binder. The Stern–Volmer coefficients are determined in the paint calibration process and these coefficients are necessary components in the error analysis process.

16.5.5.2 Uncertainty in Temperature and Pressure-Sensitive Paint Measurements

Sources of uncertainty for temperature- and pressure-sensitive paint measurements have been investigated and modeled by Liu and Sullivan. A functional relationship between the system components was developed by Liu and the elemental error sources and their sensitivity coefficients in the error propagation were evaluated. These error sources include temperature, illumination, model displacement and deformation, sedimentation, photo-degradation, and camera shot noise. To determine the overall variance in pressure for a single element the elemental error is computed and multiplied by its sensitivity coefficient. The total error is computed by summing each elemental error using

$$\frac{\text{var}(P)}{P^2} = \sum_{i=1}^M S_i^2 \frac{\text{var}(\zeta_i)}{\zeta_i^2} \quad (16.14)$$

To facilitate the error analysis of the data to be presented, a brief discussion of each error mechanism identified by Liu and the impact that each has on a given experiment is included. Finally, a table consisting of error sources, quantitative estimates of variances, and the final impact that these errors will have on the data are presented.

16.5.5.3 Temperature Sensitivity of Pressure-Sensitive Paint Errors in pressure measurements due to temperature are the result of changes in the temperature distribution on a test surface between the *wind-off* and the *wind-on* image acquisition. The physical process that causes the unsteady temperature on the jet impingement surface and the means for minimizing this error will be discussed later. Assuming one has minimized the thermal drift of the test surface, the remaining quantity of interest is the sensitivity of the system to temperature. Liu defines the sensitivity coefficient of the pressure-sensitive paint measurement to temperature using the Stern–Volmer coefficients that were defined in Equation (16.2). Differentiation of these equations and rearranging yields

$$\frac{-T}{B(T)} \left[B'(T) + A'(T) \frac{P_{\text{ref}}}{P} \right] \quad (16.15)$$

Again, $A(T)$ and $B(T)$ are defined using Equation (16.2), and the primes denote differentiation.

The elemental variance of the temperature can be estimated using both theoretical and experimental means. A theoretical estimate of the maximum temperature variation on a model can be obtained by computing the recovery temperature of the fluid on the model surface. This is accomplished using one-dimensional isentropic flow equations and the definition of the recovery temperature. An upper limit on the temperature variation is obtained by assuming the flow is laminar and the recovery factor (r) is 0.84. For an adiabatic model at Mach 0.2 $\text{var}(T)$ is about 0.4 K whereas at Mach 1.0 $\text{var}(T)$ is closer to 9 K.

An experimental measurement of the temperature variation can also be conducted using temperature-sensitive paint. The experimental variance for an isothermal model is generally about 10–20 times smaller than the theoretical prediction. The small temperature variation is the result of an experimental setup that seeks to minimize the temperature variation on the model surface. A temperature variation of 1 K will be used for the error analysis.

16.5.5.4 Illumination and Model Deformation/Displacement The relationship between surface illumination and paint luminescence is linear; therefore, any change in surface illumination will result in an equal change in paint luminescence. Errors in pressure measurements caused by variations in surface illumination can stem from several sources. Consider utilizing a point source for the illumination of a surface. The relationship between illumination intensity at a point on the surface and the distance between the source and the point of interest is an inverse function of the distance squared. Any movement of the painted surface or illumination source will result in a change in the distance between these two points, and thus a change in the illumination intensity at the surface. This movement can result from deformation of the model surface, or physical displacement of the model or illumination source. Another source of illumination errors is the temporal stability of the illumination source. Any variation of the intensity of the illumination source between the *wind-off* and the *wind-on* images will register as an error in illumination. The sensitivity coefficient for the illumination errors as defined by Liu is

$$\varphi = 1 + \frac{A(T)}{B(T)} \frac{P}{P_{\text{ref}}} \quad (16.16)$$

In this equation, $A(T)$ and $B(T)$ are defined in Equation (16.2), P is the experimental pressure, and P_{ref} is the reference pressure. The quantity in Equation (16.6) is a very common sensitivity coefficient and is therefore defined as φ .

In general, illumination errors are a major contributor to the overall error in pressure-sensitive paint measurements. The use of lifetime and binary paints has been developed specifically to mitigate this error source. If a tradition PSP is used, it is essential that the model surface, illumination source, and camera are all mounted rigidly so as to eliminate relative movement. The rigid nature of the experiment setup results in experimental system where errors due to displacement or deformation are negligible.

The remaining source of illumination error is the temporal stability of the illumination source. LED arrays have been developed that provide very stable illumination and are recommended for high quality PSP testing.

16.5.5.5 Detector Noise One significant source of noise for radiometric pressure-sensitive paint measurements using CCD cameras identified by Liu is noise from the detector. For a CCD camera this includes both read noise and shot noise. The read noise for most scientific grade CCD cameras is well below the shot noise. If this is not the case, the camera is not suitable for PSP measurements. The remaining discussion of detector noise will focus on shot noise of the detector. For a CCD array, the quantity of interest is the number of photoelectrons collected. In the absence of other sources of detector noise (e.g., read noise), the uncertainty in the intensity measurement is an inverse function of the square root of the number of photoelectrons that the measurement is based on. It is worth noting that, theoretically, the detector can be eliminated as a source of error; simply acquiring multiple frames and summing the images will drive the noise to zero. The shot noise will decrease with the square root of the number of shots acquired, or the total number of photons the measurements is based on.

$$\text{Shot noise} \propto \frac{1}{\sqrt{Nn_{\text{pe}}}} \quad (16.17)$$

In this equation, n_{pe} is the average number of photoelectrons collected by a camera pixel during each exposure, and N is the number of images acquired.

Unfortunately, acquiring an infinite number of images is not practical in an experimental environment. Test parameters such as fluid or surface temperature, photo-degradation and sedimentation of the painted surface, and illumination stability are difficult, or impossible to control in practical situations. As these parameters change during the data acquisition process, they begin to introduce more noise than is being eliminated by the additional images. There is a tradeoff between the time required to acquiring multiple images, and thus drive down shot noise, and the stability of the test parameters. For this reason, a limited number of images must be acquired at each test condition. Data sets consisting of 1–128 images are generally used for PSP experiments. This number is usually set by a limit on the time on condition allowed for data acquisition. The full-well capacity of the CPO.1600 camera is about 40,000 photoelectrons. Assuming the average pixel reached 80% of the full-well capacity, a set of 128 images will yield a shot noise of about 0.05%. Combining this shot noise with the sensitivity coefficient for detector noise

$$S = \left[1 + \frac{A(T)}{B(T)} \frac{P}{P_{\text{ref}}} \right] \left[\sqrt{1 + A(T) + B(T) \frac{P}{P_{\text{ref}}}} \right] \quad (16.18)$$

the noise level for pressure-sensitive paint measurements using FIB is about 70 Pa. This represents a best-case noise level for a data set with no low pass filtering.

16.5.5.6 Photo-Degradation and Sedimentation Photo-degradation of the luminescent probe and sedimentation of the painted surface introduce bias errors into the pressure measurements. Photo-degradation is an optical reaction that is a function of the illumination intensity, spectral content of the illumination, and time. Sedimentation is caused by dirt and oils in the flow that accumulate on the painted surface. Over time the sediments screen the paint from the illumination light and mask the paint luminescence from the detector. If sedimentation and photo-degradation occurred uniformly over the test surface, this error could be corrected using an *in situ* pressure tap. Unfortunately, these effects are a function of the local conditions. To minimize errors due to sedimentation and

photo-degradation it is necessary to minimize the time interval between the *wind-off* and the *wind-on* images.

It is noted that photo-degradation and sedimentation are not eliminated by the use of a binary PSP. The photo-degradation rate of the two dyes in a binary paint is rarely equal, and therefore, the difference in the photo-degradation rates is effectively the photo-degradation rate of the paint. In the case of a lifetime-based measurement, photo-degradation has almost no effect, other than to lower the PSP signal, and therefore, the shot noise. The impact of sedimentation is mitigated substantially by the use of either a binary or lifetime-based PSP system. It is possible that the optical density of the sediment is not identical for each channel of a binary PSP. It is also possible that sedimentation could occur rapidly, thus having an impact between the acquisition of individual images; however, these effects are generally negligible.

A worst case estimate of the impact of photo-degradation and sedimentation for a traditional PSP is based on the test facility and experimental hardware. The data acquisition time is generally several minutes and the photo-degradation rate is about 1% per hour at the exposure levels seen in most wind tunnels. Generally, the photo-degradation rate is about 1% per hour at the exposure levels seen in most wind tunnels. Sedimentation varies widely with wind tunnels, and therefore, a very rough estimate of 1% per hour is applied here as well. Assuming data is acquired at five to ten test conditions between *wind-off* data sets, the time required to complete data acquisition is about 5 min. This results in photo-degradation and sedimentation of about 0.09%. The sensitivity coefficient for sedimentation and photo-degradation is again, φ , defined in Equation (16.6).

16.5.5.7 Spectral Variability and Filter Leakage The output of some illumination sources, such as flash-lamps, experience spectral variations over time. Variations in the spectra of the illumination source can result in variations of the paint luminescence. The variation in the paint luminescence is due to variations in the absorption characteristics of the luminescent molecule. The spectral output of the LED array is, however, very stable over the life of the array, but there is a warm-up period lasting several minutes during which the spectral output of the LED array will shift by less than 0.1 nm. Following this warm-up period, the output is stable. Spectral variations of the illumination source of this magnitude are of no significant consequence. It is noted that spectral variation can be a significant problem for other illumination sources. Care must be taken to filter the source and ensure a stable illumination spectrum.

The purpose of the filter on the detector is to eliminate the blue/UV excitation from the red-shifted emission of the paint. Errors due to filter leakage can result if the illumination source includes spectral content at the wavelength of paint luminescence. Generally, this spectral content is minimized by using a short pass filter on the illumination source. Most LEDs, for example, do have some spectral content at wavelengths beyond 600 nm; this contribution is minimized by adding an appropriate short-pass filter to the LED array. The variance for spectral variability is assumed negligible for the LED lamp, while the filter leakage is estimated to be 0.01%. This is combined with the sensitivity coefficient for spectral variability and filter leakage φ , defined in Equation (16.6) to determine the uncertainty in the final pressure measurement.

16.5.5.8 Reference Pressure and Paint Calibration Parameters The quantity P_{ref}/P appears directly in the Stern–Volmer equation. Any error in measuring the reference pressure, the pressure that the *wind-off* images are acquired at, will result in an error in the

final pressure measurement. The sensitivity coefficient for this error is unity. This error is based on the accuracy of the pressure transducer available in the laboratory during the experiments and is less than one-tenth of a percent.

The uncertainty in determining the Stern–Volmer coefficients $A(T)$ and $B(T)$ are calibration errors. These errors can be estimated by performing repeated calibrations and computing the variance of the quantities $A(T)$ and $B(T)$. The variance computed from such tests in the controlled calibration chamber is generally less than 1%. The use of a priori calibrations often results in bias errors in field tests. If these errors are the result of variations in the Stern–Volmer coefficients, then the a priori calibrations are not valid. A more likely explanation is that other errors, such as temperature, illumination, and model movement are responsible for the inaccuracies. The respective sensitivity coefficient for $A(T)$ and $B(T)$ are $1 - \varphi$ and 1.

16.5.5.9 Sensitivity Coefficients A summary of the physical source of the each error and the corresponding sensitivity coefficient for that error is given in Table 16.3. Some means of estimating the variance of each error source is also necessary. The physical

TABLE 16.3 Estimation of Elemental Errors for Pressure-Sensitive Paint Measurements

Variable	Physical Origin	Sensitivity Coefficient	Elemental Variance	Estimated Error
		S_i	$\frac{\text{var}(\zeta_i)}{\zeta_i^2}$	
T	Temp. sens. of PSP	$\frac{-T}{B(T)} \left[B'(T) + A'(T) \frac{P_{\text{ref}}}{P} \right]$	$\left(\frac{\Delta T}{T} \right)^2$	$\Delta T \approx 1K$
$D_x(\Delta x)$	Deformation/ displacement	$\varphi = 1 + \frac{A(T) P_{\text{ref}}}{B(T) P}$	$\left(\frac{\Delta x}{x} \right)^2$	$\Delta x \approx 0$
$D_{q0}(\Delta t)$	Temporal variation in illumination	φ	$\left(\frac{\Delta I}{I} \right)^2$	$\frac{\Delta I}{I} \approx 0.1\%$
V_{ref}	Photo-detector noise ref. image	$\varphi \left(\sqrt{1 + A(T) + B(T) \frac{P}{P_{\text{ref}}}} \right)$	$\frac{1}{n_{\text{pe}}}$	$n_{\text{pe}} \approx 3.6 \times 10^6$
V	Photo-detector noise signal image	$-\varphi \left(\sqrt{1 + A(T) + B(T) \frac{P}{P_{\text{ref}}}} \right)$	$\frac{1}{n_{\text{pe}}}$	$n_{\text{pe}} \approx 3.6 \times 10^6$
$D_t(\Delta t)$	Photo-degradation and sedimentation	φ	$\left(\frac{\Delta D_t}{D_t} \right)^2$	$\frac{\Delta D_t}{D_t} \approx 0.1\%$
Π_t/Π_{fref}	Illum. spectral varia- bility and filter leakage	φ	$\left(\frac{\Delta \Pi}{\Pi} \right)^2$	$\frac{\Delta \Pi}{\Pi} \approx 0.05\%$
P_{ref}	Error in reference pressure	1	$\left(\frac{\Delta P}{P} \right)^2$	$\frac{\Delta P}{P} \approx 0.1\%$
A	Paint calibration	$1 - \varphi$	$\left(\frac{\Delta A}{A} \right)^2$	$\frac{\Delta A}{A} \approx 1\%$
B	Paint calibration	-1	$\left(\frac{\Delta B}{B} \right)^2$	$\frac{\Delta B}{B} \approx 1\%$

parameter used to estimate the variance of each source is also listed in Table 16.3. The sensitivity coefficients are a function of the paint calibration coefficients, the temperature, and the pressure. To demonstrate the advantages of a PSP with low-temperature sensitivity, a theoretical experiment is performed using several common paint formulations, Ruthenium in RTV and PtTFPP in FIB are included.

$$\begin{aligned} A(T) &= 0.56 \left[1 + 2.72 \left(\frac{T - 298}{298} \right) \right] \\ B(T) &= 0.43 \left[1 + 4.75 \left(\frac{T - 298}{298} \right) \right] \text{ Ruthenium in RTV} \end{aligned} \quad (16.19)$$

$$\begin{aligned} A(T) &= 0.25 \left[1 + 0.96 \left(\frac{T - 298}{298} \right) \right] \\ B(T) &= 0.75 \left[1 + 1.52 \left(\frac{T - 298}{298} \right) \right] \text{ PtTFPP in FIB} \end{aligned} \quad (16.20)$$

The relative importance of each error source can be evaluated by comparing the elemental errors. The error in pressure that would result from the estimated elemental variance listed in Table 16.3 was computed at a P_{ref}/P of unity for PtTFPP in FIB, Ruthenium in RTV with Silica Gel, and Ruthenium in RTV. The pressure errors, plotted in Figure 16.10, indicate that temperature and paint calibrations are the major sources of error for these measurements. The temperature error is minimized using PtTFPP in FIB. The errors associated with the Ruthenium in RTV paint compare poorly to both PtTFPP in FIB and Ruthenium in RTV with Silica Gel in each category. This poor performance is related to the inferior pressure sensitivity of Ruthenium in RTV.

Temperature variation has been identified as a major source of error for many PSP measurements. Further insight into the susceptibility to temperature errors for the paint is

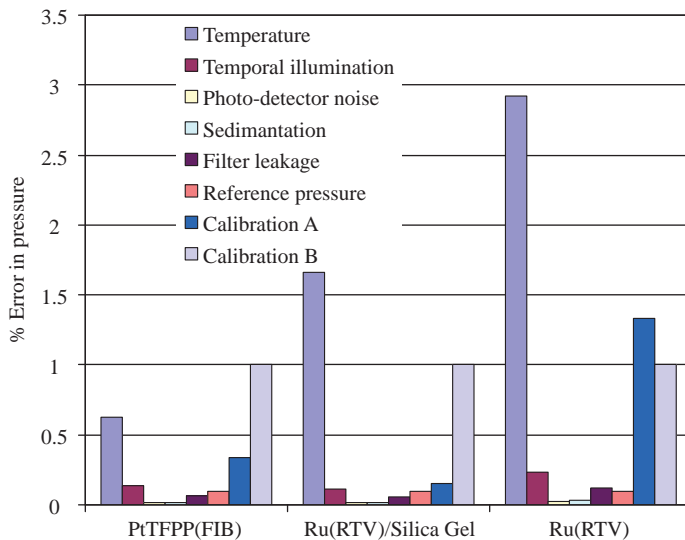


FIGURE 16.10 Elemental errors for pressure-sensitive paint measurement. (See the color version of this figure in Color Plates Section.)

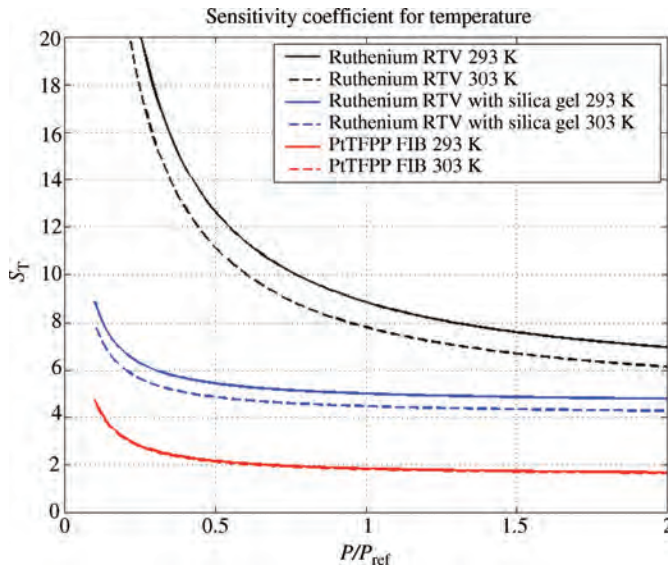


FIGURE 16.11 Sensitivity coefficient for temperature for three paint formulations. (See the color version of this figure in Color Plates Section.)

provided by plotting the sensitivity coefficient for temperature as a function of P_{ref}/P ; the result is shown in Figure 16.11. Again, PtTFPP in FIB provides the smallest sensitivity coefficient for temperature. It is also evident that the sensitivity coefficient for temperature of PtTFPP in FIB is a very weak function of temperature. The overall low-temperature sensitivity is the result of the low activation energy for the FIB polymer. In contrast, Ruthenium in RTV has a very high-sensitivity coefficient for temperature.

The goal of the preceding discussion is to introduce the tools for error analysis of pressure-sensitive paint measurements. Insight into the data acquisition procedure that will yield the best possible data is obtained by considering each error and the experimental mechanisms available to minimize each error. A good experimental setup will provide a very rigid framework for mounting the test surface, illumination source, and detectors to eliminate model movement and deformation. This can also be accomplished by using a lifetime or binary PSP system. Data acquisition should be performed quickly to minimize photo-degradation and sedimentation of the test surface. The detector should acquire multiple images quickly to minimize shot noise. Finally, the temperature of the test surface should be uniform throughout the test. In this case, a temperature compensating binary PSP can minimize the impact of both temperature and model movement.

16.6 PRESSURE-SENSITIVE PAINT MEASUREMENTS

16.6.1 Research Facilities

The use of pressure-sensitive paint for measurements in research-based experiments is well developed. The use of PSP is quite common in transonic tunnels, with measurements conducted at a variety of academic, government, and even commercial production facilities. Today, these measurements are conducted using both binary and lifetime-based PSP

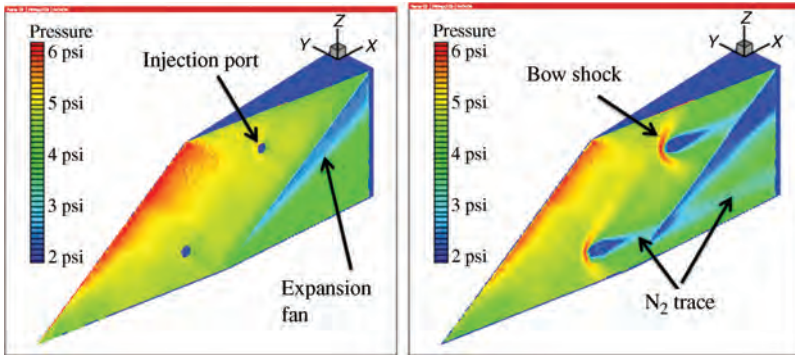


FIGURE 16.12 Pressure measurements on a supersonic flameholding pylon at Mach 2. (See the color version of this figure in Color Plates Section.)

systems. A few examples of the application of PSP systems in these facilities are included here. The goal is to demonstrate the application of PSP in different wind tunnel settings; this is by no means a comprehensive list of the users of PSP.

16.6.1.1 Supersonic Flameholding Pylons An example of binary PSP data acquired in a supersonic tunnel is an experimental study of a flameholder pylon in a direct-connect supersonic facility located at WPAFB/AFRL. Symmetric facility nozzles were used to produce supersonic flow at nominal Mach numbers of 2 and 3. Windows are located on both side walls and top wall for optical access and two sets of illumination source and CCD camera were aligned to obtain the side view and top view for PSP measurements. Each system was composed of a PCO.1600 CCD camera with a filter wheel, a 55-mm Nikon lens, and a laptop PC running OMS Acquire.

Surface pressure distributions on one pylon at Mach 2 are illustrated in Figure 16.12 (Crafton et al., 2009). A high-pressure zone at the thin leading edge of the strut due to the flow impingement and a low-pressure region behind the expansion are clearly observed. As these pylons are under investigation for super-sonic flame holders, there are several fuel injection ports evident on the surface. An interesting use of PSP is as a flow visualization tool. Nitrogen (N_2) was injected through the injection ports and the resulting pressure field is indicative of both pressure and mass fraction of injected gas. A pair of bow shocks are formed in front of the injection ports as would be expected. The trace of the injected N_2 near the pylon surface is identified by the low-pressure region behind the injection ports. These regions are not really low pressure, but the result of the replacement of oxygen by the injected N_2 . The presence of the gas on the surface would indicate that some of the injected gas should enter the stagnant subsonic zone behind the strut, and therefore, possibly be available for combustion and flame holding.

16.6.1.2 Transverse Jet Injection in a Supersonic Flow The jet-in-crossflow is a simple configuration and as such is often considered a canonical problem for both low- and high-speed flows. For high-speed flows, it has many interesting features that make it challenging to simulate, including flow separation (both upstream and downstream of the jet axis) and unsteadiness; furthermore, the degree of separation and unsteadiness appear to depend strongly on injection angle and pressure. PSP offers a means of studying the features on this flow with high spatial resolution.

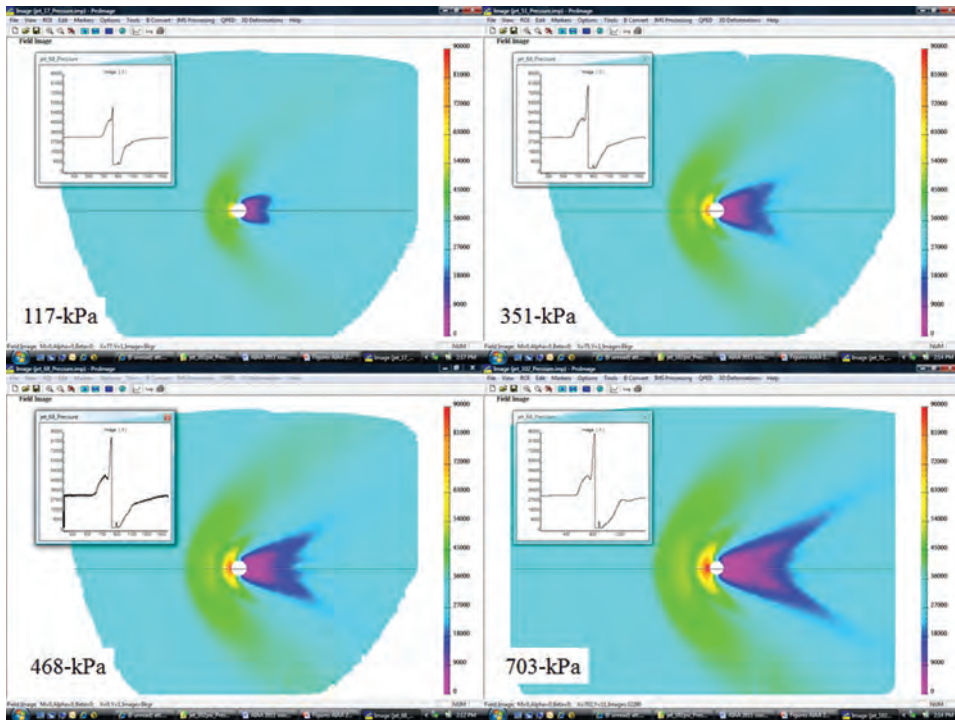


FIGURE 16.13 Mean pressure distribution for 3/16-in. injector block operating at several injection pressures. (See the color version of this figure in Color Plates Section.)

The experimental study was conducted in a supersonic flow facility operated within the Aerospace Propulsion Division, of the Propulsion Directorate, Wright-Patterson AFB (AFRL/RZA). Four injection blocks were tested in a Mach 2 flow at a series of injection pressures. Data were acquired with a binary PSP system composed of a PCO-1600 CCD camera with a 55-mm focal length lens, a filter wheel, two LM2X-400 LEDs, and a laptop PC running OMS Acquire. This particular experiment is part of an ongoing effort to build a database for high-fidelity CFD model validation and development.

A single pressure tap (in the field of view) was used for an *in situ* correction of the PSP data. Past experience in this facility indicates that the final binary PSP data should be accurate to within 200 Pa after the *in situ* bias correction. Examples of the pressure distribution using the 3/16 in. injector block at a series of injection pressures is shown in Figure 16.13 (Crafton et al., 2011). The flow features, such as the stagnation zone with high pressure just upstream of the jet, the bow shock upstream of the stagnation zone, and the low-pressure region behind the injector are expected. The flow is very symmetric about the injector. The pressure increases slowly from the bow shock toward the injector. There is a small pressure decrease, associated with a horseshoe vortex, followed by a stagnation zone just upstream of the jet. The low-pressure zone behind the injector is again symmetric, with the flow expanding to return to a flat pressure distribution downstream of the injector. Behind the jet, there are expansion fans emanating from the low-pressure zone. As the injection pressure is changed, the overall structure of the flow is unchanged. The bow shock, horseshoe vortex, and stagnation zone are still present in

front of the jet, and there is still a low-pressure zone behind the jet. The horseshoe vortex is not evident in the pressure distribution at the lowest injector pressure. As the injector pressure increases, the bow shock moves upstream and the stagnation pressure in front of the jet increases. The low-pressure zone and expansion fans behind the jet are also expanding and extending downstream.

16.6.2 Production Wind Tunnels

The use of PSP is becoming more common in large transonic tunnels, with production systems in use in several facilities such as AEDC (Ruyten and Sellers, 2004), ARA (Mosharov et al., 1998), and DLR (Engler et al., 2005). These measurements have been conducted using both binary and lifetime-based PSP systems. A demonstration of a lifetime and binary PSP system was conducted in the 9×8 ft transonic tunnel at ARA to compare these approaches to PSP. A binary PSP and data acquisition system was utilized on one wing and a lifetime PSP and data acquisition system was utilized on the other wing. The binary paint system was composed of BF-400 PSP, 3 LM2XDM-400 LEDs, and a PCO.1600 with a filter wheel. The lifetime system was composed of UF-400 PSP, 3 LM2XDMHP-400 LEDs, and a PCO.1600 with a pulse generator. Data were acquired at identical test conditions, and pressure tap data were collected at 48 stations along two span locations of the wing. Several of the pressure taps were used to anchor the absolute pressure of the PSP measurements and the remaining taps were used to estimate the accuracy of the PSP data by computing the deviation between the tap and PSP data.

Data were acquired at Mach 0.7 and 5° , the resulting pressure distribution, along with a comparison of the PSP and pressure tap data is shown in Figure 16.14 (Lifetime System) and Figure 16.15 (Binary System) (Lee et al., 2010). The pressure tap data were compared with the PSP data for each system, and the deviation between the taps and the PSP was about 400 Pa ($\sim 0.03 C_p$) for the data acquired at this test condition. It is noted that the hardware used for each system is quite similar, PCO.1600 CCD cameras, LED

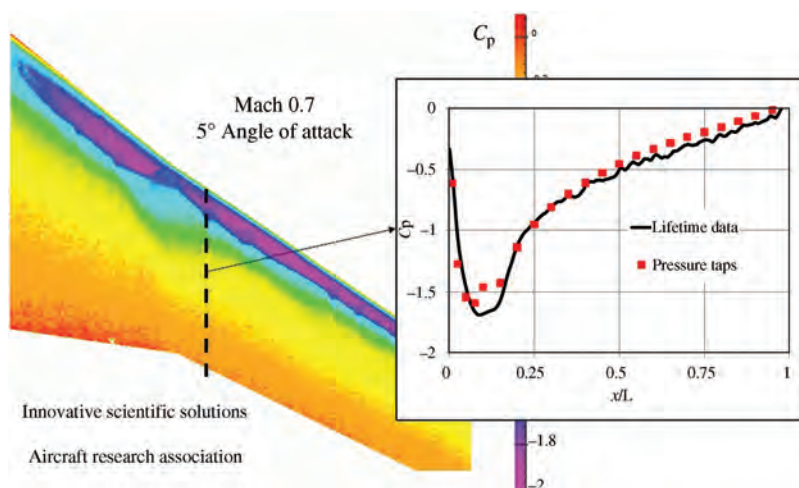


FIGURE 16.14 Pressure distribution using lifetime PSP on an airfoil in the ARA 9×8 ft transonic wind tunnel. (See the color version of this figure in Color Plates Section.)

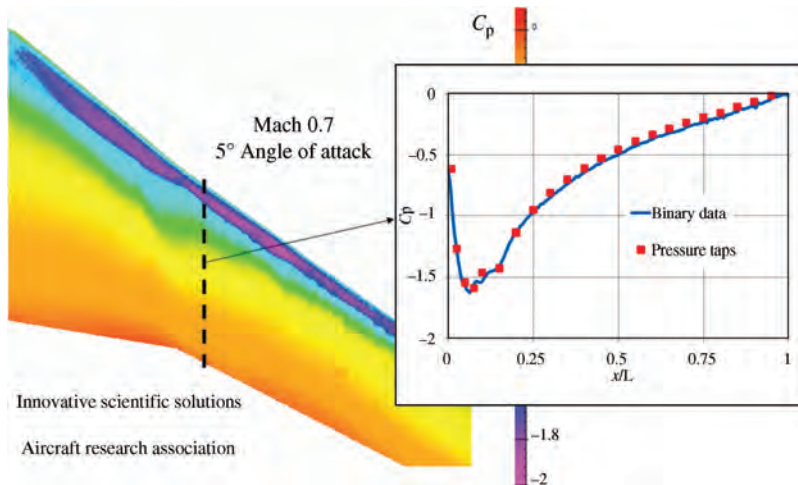


FIGURE 16.15 Pressure distribution using Binary PSP on an airfoil in the ARA 9×8 ft transonic wind tunnel. (See the color version of this figure in Color Plates Section.)

illumination, and PtTFPP-FIB-based paints, and therefore, it is not surprising that similar results were obtained. One of the advantages of PSP is the detailed surface pressure distribution that it can provide when the system is used to acquire full field images of the model. In this case, the data were analyzed and presented on a bitmap, and therefore, no loads calculations were conducted. Data from a later entry in 2009 with a 10-camera system and a full view of the model were presented by Vardaki et al. (2010). In this case, the deviation between the taps and PSP was about 500 Pa, and the integrated C_L and C_M data agreed with the balance to within 5%.

One of the first tunnels to deploy a PSP system integrated into the tunnel data acquisition system was the lifetime-based PSP system at Arnold Engineering Development Center (AEDC). The 16T PSP data acquisition operates in two-gate lifetime mode using PtTFPP FIB paint, 8 CoolSnap K4 Cameras, and 40 LM4X-400 LEDs to provide full coverage of the model. The system is tightly integrated into the tunnel to provide productive testing and near real-time data presentation of the results on the model mesh. In 2008, AEDC demonstrated their PSP system using the Facility Aerodynamics Validation and Operations Research (FAVOR) model. This model is for use as a standard check model for propulsion wind tunnel (PWT) 16T. The test article is a 5% scale model of the F-111 with new wings having a NACA 64-210 profile, a fixed sweep angle of 35° , and a span of 1.22 m. PSP data were acquired for demonstration of the techniques' capability to measure surface pressure and determine total vehicle loads and for use in CFD validation. The PSP data were compared and validated by conventional pressure data measured on the model and with total vehicle loads measured by the internal balance.

Testing was conducted from Mach 0.4 to 1.1 and from -3° to 24° angle of attack. The data at several Mach numbers and 10° are presented in Figures 16.16–16.18. These data illustrate excellent agreement with the conventional pressure data as well as having excellent symmetry between the left and the right wings. Similar to the ARA data, the PSP data were anchored by the conventional pressure measurements using an average offset. There still remains a residual error between the paint and each individual orifice

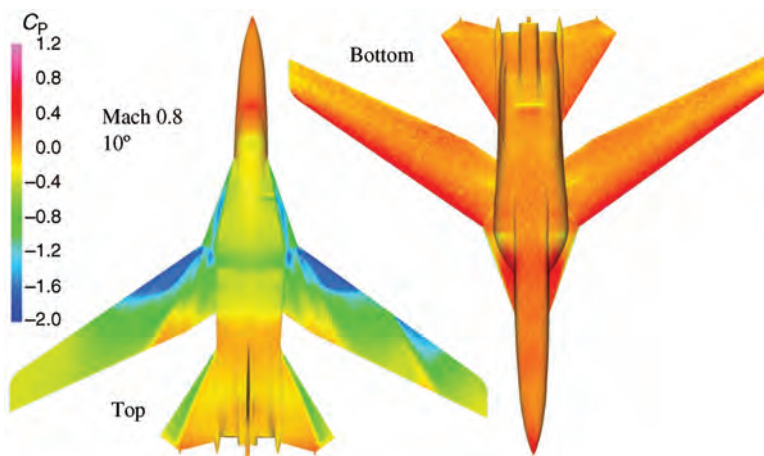


FIGURE 16.16 Pressure distribution at Mach 0.8 and 10° on the FAVOR model at 16T, AEDC. (See the color version of this figure in Color Plates Section.)

and the PSP measurement error can be estimated by computing the difference between the PSP and the conventional pressure measurements on the two wings and the limited fuselage locations. The average pressure error of all data points at each Mach number and for each pressure tap was computed. The average error for Mach 0.8, 0.95, and 1.1 was about 550, 440, and 400 Pa, respectively. Replicate runs were performed with similar results; these measurement errors are very good for PSP.

As full field data were acquired, a comparison of measured balance loads with the PSP integrated loads for the total vehicle were also computed. The PSP normal force and pitching moment agree very well with the balance measurements at all Mach numbers with the exception of the PSP pitching moment at Mach 0.4 for angles of attack less than 14° . The most likely reason is the very small and strong low-pressure region along the

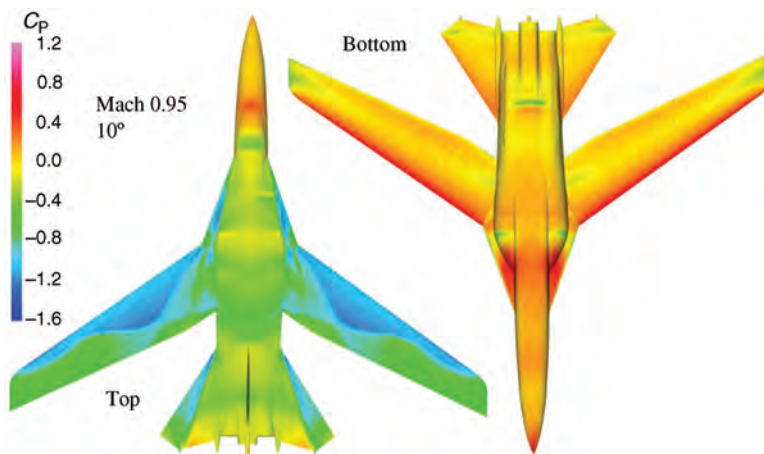


FIGURE 16.17 Pressure distribution at Mach 0.95 and 10° on the FAVOR model at 16T, AEDC. (See the color version of this figure in Color Plates Section.)

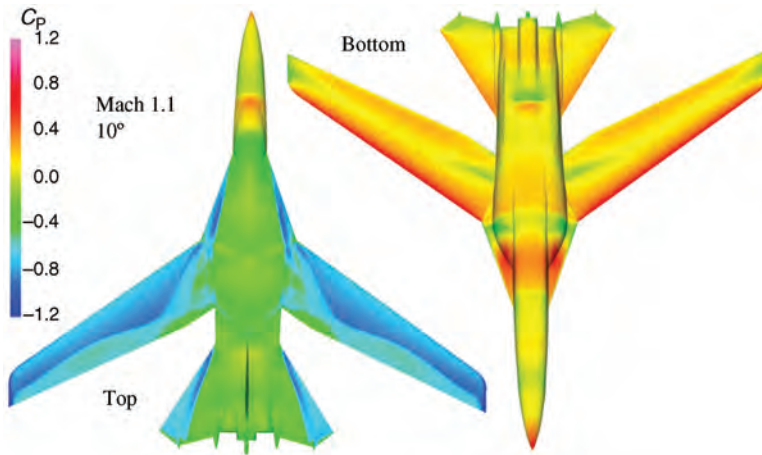


FIGURE 16.18 Pressure distribution at Mach 1.1 and 10° on the FAVOR model at 16T, AEDC. (See the color version of this figure in Color Plates Section.)

leading edge of the wings. A small error in the mapping of this gradient can significantly affect the integrated pitching moment. The side force and yawing moment should be near zero since the model is supposedly symmetric. While the balance side force is near zero, a significant yawing moment was measured, especially when asymmetric separation occurred over the wing. This separation is evident in the large rolling moment generated and coincided with a spike in the yawing moment. This must be a coupling yaw moment because the side force did not have a corresponding change. The PSP yaw and roll plane data typically tracked the balance data but with some offset. It is difficult to pinpoint the source of the error in the PSP data because small forces at large distances (i.e., wing tips) from the moment reference center can introduce large moment errors. The PSP axial force tracked the balance measurement for all conditions but was always low. This was as expected because the PSP does not measure skin friction force, which is measured by the balance. Generally, the integrated C_L and C_M data agreed with the balance to better than 10%.

16.6.3 Low-Speed Applications of Pressure-Sensitive Paint

Low-speed wind tunnel measurements have been a target for deployment of pressure-sensitive paint systems for over 15 years; however, quantitative pressure measurements in flows below Mach 0.3 have proven challenging. Over the past 7 years, core PSP hardware such as CCD cameras, LEDs, image processing software, and temperature compensating binary paints have been improved. Integration of this improved hardware into pressure-sensitive paint systems has resulted in several successful demonstrations of pressure-sensitive paint in low-speed wind tunnels. These demonstrations include several small tunnel tests where the deviation between the pressure taps and pressure-sensitive paint data is as low as 50 Pa. In larger production style tunnels, deviation between the pressure taps and pressure-sensitive paint data of close to 100 Pa has been demonstrated. This section includes an overview of the challenges associated with utilizing pressure-sensitive paint in low-speed wind tunnels and gives an outline of the equipment and experimental procedures that have produced high quality pressure data in low-speed tunnels.

16.6.3.1 Background Early investigations into the use of PSP in low-speed wind tunnels were conducted by Brown (2000) and Bell (2004). Brown focused on applying careful experimental procedures to mitigate sources of error such as temperature and model movement. While Brown demonstrated very accurate PSP measurement, the experimental procedures are not practical in a large production wind tunnel. Bell (2001) investigated the sensitivity of the lifetime and binary measurements (Bell, 2004). Bell investigated the sensitivity of the lifetime and binary⁵⁷ PSP techniques to the common sources of errors in PSP measurements. Bell offers an excellent overview of the experimental procedures necessary to acquire high quality PSP data, concluding that binary systems offer the best option for deploying a productive PSP system. This, along with the demonstrations of production-style binary PSP systems from DLR14 and TsAGI13 indicate that a binary approach may yield accurate, productive PSP measurements in large low-speed wind tunnels.

16.6.3.2 Automotive Mirror in the AFIT Wind Tunnel One of the target applications of a low-speed PSP system is automotive research. Of particular interest is the generation of acoustic noise by the interaction of the automobile with the surrounding fluid, for example, vortex shedding from mirrors or cross-flow. An experiment to demonstrate the use of PSP for automotive testing was conducted in July 2004 in the Air Force Institute of Technology (AFIT) low-speed wind tunnel. The AFIT tunnel is an open circuit wind tunnel with a 3 ft by 4 ft test section and a maximum speed of Mach 0.2. A full-sized automotive mirror was outfitted with six pressure taps and painted with PSP. Testing was conducted at 30, 45, and 60 m/s.

A second automotive experiment was conducted using a side mirror. In this case, six pressure taps were placed on the mirror, and the mirror was mounted in the tunnel. The mirror was painted with binary FIB and tests were conducted at speeds of 30, 45, and 60 m/s. In this case, data acquisition involved averaging 10 signal and reference images at wind-off and wind-on conditions. The data were processed by aligning the images using the registration markers, computing the ratio as described in Equation (16.3), applying a 16 pixel Gaussian low pass filter, and then converting to pressure using an a priori calibration. The processed data were then corrected using the pressure tap data. A simple bias correction was computed by minimizing the mean squared deviation between the pressure taps and the PSP data, and a bias correction of 479 Pa was applied to the data. With the correction applied, the mean squared deviation between the taps and the PSP data was just under 200 Pa.

The corrected data acquired from the system at 60 m/s is shown in Figure 16.19 (Crafton et al., 2006). The flow is from the right to the left in Figure 16.19 and the upstream face of the mirror is oriented such that the side of the mirror is approximately parallel to the flow. The pressure increases as the flow stagnates on the upstream surface of the mirror, then the pressure drops as the flow accelerates around the curved corner near the middle of the mirror. The pressure drop is most noticeable at the lower rounded corner of the mirror. On the downstream surface, the pressure is relatively flat, and close to tunnel static. At 60 m/s, the tunnel dynamic pressure is about 2200 Pa and the range of pressures between the downstream and upstream surfaces of the mirror is about 2200 Pa, a C_p of about 1. The C_p drops to about -1 on the long curved corner, and to about -1.7 on the rounded corner. The PSP and pressure tap data compare to within about 200 Pa. In this case, this is about 10% of the tunnel dynamic pressure and better than 4% of the dynamic range on the mirror surface.

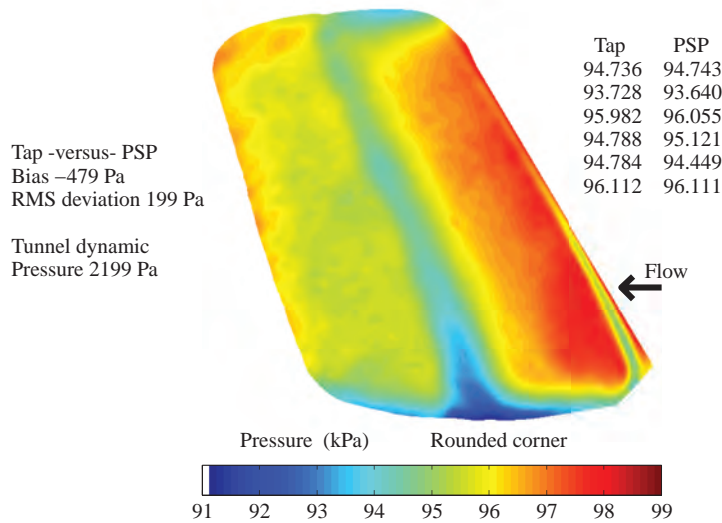


FIGURE 16.19 Pressure measurements on an automotive mirror. (See the color version of this figure in Color Plates Section.)

Several issues and system improvements were identified during this set of low-speed wind tunnel tests. Both tests were conducted using four LM2-400 LEDs. This illumination source produces about 200 mW of 400 nm light, for a total of 800 mW. Exposure times on these tests were over 1 s per image, and therefore, the on condition time was well about 30 s. A brighter illumination source would facilitate shorter exposures, and therefore, more images could be averaged during a given time on condition. More signal averaging in a shorter time would produce better signal-to-noise and therefore, more accurate data. It is also noted that long data acquisition times are prohibitive in large wind tunnels due to the cost of tunnel time. A stronger illumination source was developed for future tests.

16.6.3.3 MicroX in SARL Measurements of pressure using the PSP system were conducted on an 8% scale model of the Micro-X model in the Subsonic Aerodynamics Research Laboratory (SARL) in November 2005. The SARL is an open circuit tunnel at Wright Patterson Air Force Base with a 7 ft by 10 ft test section. Past PSP campaigns in the SARL have been compromised by errors from model movement, background lighting, sedimentation of the model, and temperature variations of the test surface (Crafton et al., 2005). In these tests, the mean squared deviation between the taps and PSP has never been better than 1000 Pa. The impact of these errors should be minimized by the use of the upgraded binary PSP system, careful design of the model, and careful experimental procedures.

The model was constructed of aluminum to minimize temperature gradients and a temperature compensating binary FIB paint was applied to compensate for errors due to model movement as well as any remaining thermal errors. Both wind-off and wind-on images were acquired while interlacing background shots to compensate for the dynamic background lighting encountered in this tunnel. The experimental setup included two views of the model, one concentrated on the top surface of the wings and canards and the other on the side body of the model. Each subsystem was composed of four LM2X-400

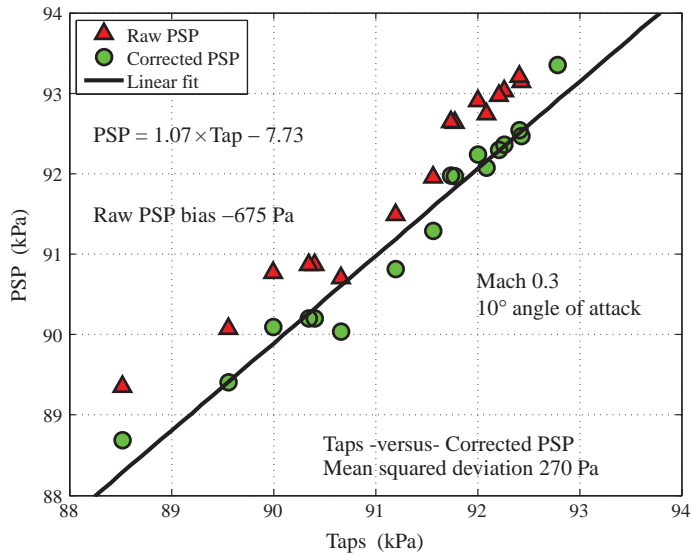


FIGURE 16.20 PSP versus taps in large tunnel. (See the color version of this figure in Color Plates Section.)

LEDs and a PCO.1600 with a filter switch as described in Figure 16.7. The exposure time for each system was approximately 1.5 s per image. The two-component system was armed to acquire data and then controlled using a single quantum composer to trigger the cameras and lamps from each system. Data acquisition consisted of collecting eight signal and reference images along with the interlaced backgrounds at each test condition. Data were acquired at Mach 0.3 and from 0° to 35° angle of attack in 5° increments. Finally, the model included pressure taps for anchoring and validating the PSP data and resection markers so that the final data could be mapped onto a surface mesh.

The data were processed by aligning the images using the registration markers, computing the ratio as described in Equation (16.3), applying a 16 pixel Gaussian low pass filter, and then converting to pressure using an a priori calibration. The processed data were then corrected using the pressure tap data and minimizing the mean squared deviation between the pressure taps and PSP data. Since this test program included a CFD mesh of the model and physical locations of the pressure taps, the pressure tap correction process is accomplished on the mesh upon completion of the resection process.

The PSP and pressure tap data for one test condition are presented in Figure 16.20. For this test condition, the bias error indicated by the taps was 675 Pa. After correcting the PSP data for the bias the residual root mean squared deviation between the taps and PSP serves as an indicator of the noise level in the PSP measurement for that condition. Features with amplitude lower than the mean squared deviation cannot be considered relevant. The residual deviation between the PSP and the taps was 270 Pa for this test condition, about 5% of the dynamic pressure. This level of uncertainty represented a significant improvement over past entries into this facility; however, the targeted error budget for the system is closer to 100 Pa. Further, system refinements are needed.

One of the more interesting data points in this test campaign that demonstrates the utility of PSP for measurements in large low-speed tunnels involves the development

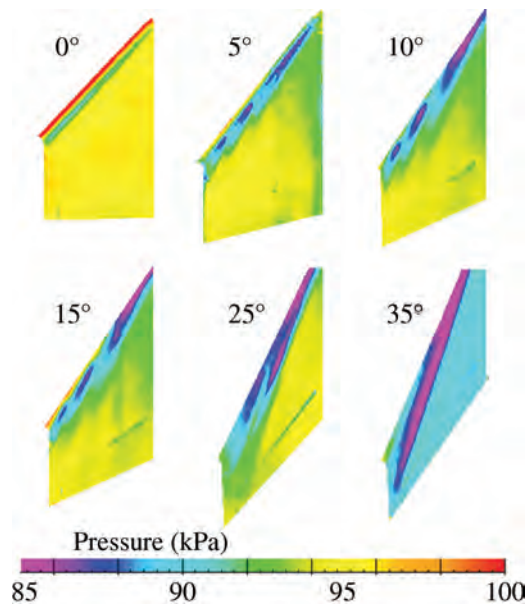


FIGURE 16.21 MicroX in SARL PSP data. (See the color version of this figure in Color Plates Section.)

of a multivortex system on the tail wings. This configuration includes a large tail at 70° dihedral. The pressure distribution on the top surface of the tail at six angles of attack, from 0° to 35° , is shown in Figure 16.21. At 5° angle of attack, the presence of three distinct vortices on the leading edge of the tail is evident. These vortices sweep along the leading edge of the tail from the root to the tip. The first vortex, located near the root of the tail, is expected as it seems to originate near the junction of the swept tail and the main body of the model. The second and third vortices are somewhat smaller and were not anticipated by the aerodynamics team. These three vortex structures persist through the 10 and 15° angles of attack; however, the second and third vortices seem to be weakening at 15° . By 35° only the expected junction vortex is present. This multivortex system, which was later determined to be caused by a slight defect in the model, would be very difficult to detect using traditional pressure taps.

16.6.3.4 Four Camera Binary PSP System in the ADD 3-m Tunnel (Lee et al., 2010)

In the summer of 2009, ISSI and the team at the Agency for Defense Development (ADD) installed the system in the ADD low-speed wind tunnel in Taejeon, Korea. The ADD low-speed wind tunnel is a closed circuit, temperature controlled tunnel with a 3 m by 2.25 m test section. The test section is created using Plexiglas, and therefore, provides excellent optical access. The tunnel operates from 10 m/s to 120 m/s and has an axial turbulence level of less than 0.05%. The model angle of attack can be set using a pitch sector and the yaw angle is set using ceiling/floor turntables. The pitch sector is mounted to the floor turntable and ceiling turntable is linked using a mechanical linkage. Optical access to the upper and lower surfaces of the model is obtained through windows in the turntables. The cameras and lamps for PSP are mounted onto the turntables above the windows, and therefore, the cameras and lamps move with the model as the yaw angle is changed. Four

camera/filter wheel sets were installed for this test: two on the ceiling turntable and the two under the floor.

The model for this test is a wing body type aircraft that has a 2 m span and a 1 m length. Testing focused on the upper and lower wing surfaces. The model was constructed of aluminum and included 26 pressure taps that were monitored and recorded during the test using a PSI8400 system with a ± 1 psid ESP module. Resection markers were placed on the model periphery for alignment of wind-off and wind-on images, and for mapping 2D images onto the 3D mesh describing model geometry. The binary PSP data acquisition system is comprised of a four subsystems controlled by a master computer using a local network. Each of the four subsystems includes a PCO.2000 camera, a filter switch, and two LM2XX-400 LEDs. The model is illuminated using a total of eight LM2XX-400 LED lamps. The LM2XX-400 is a water cooled LED lamp that provides 12 W of optical power at 400 nm from a 75 mm diameter head. Each camera images the model through a 24 mm Nikon lens and each filter switch includes a 645 nm long-pass filter for the signal channel and a 550 nm ± 40 nm band-pass filter for the reference channel.

The master PC arms each of the subsystems and controls triggering of subsystem components using a pulse generator that is connected to the PSP network. The master PC may be connected to the tunnel network and collect run conditions and model data, or it may run in a stand-alone mode. The subsystem PC controls the camera and filter switch and stores all data locally using a master file name sent from the master PC. Other information collected by the master PC such as tunnel conditions, model attitude, or pressure tap data can be organized and stored for use in post processing on the subsystem PC.

The wing/body model was painted with binary FIB PSP and installed in the tunnel. Two cameras were focused on the upper surface of the wing and two cameras on the lower surface. Data were acquired on the ADD model using the four channel system at tunnel speeds of 46, 60, and 80 m/s. Wind-off images were acquired at each wind-on model position to minimize image alignment errors. As the tunnel speed was low, signal averaging was used to drive down the shot noise and improve the overall signal-to-noise of the measurements. Exposure time was approximately 1 s, and 64 images of the signal and reference channels were averaged at each wind-off and wind-on condition. The data were processed by aligning the images using the registration markers, computing the ratio, applying a 16 pixel Gaussian low pass filter, and then converting to pressure using an *a priori* calibration. The processed data were then mapped to the mesh and the *in situ* correction using the pressure tap data were applied by minimizing the mean squared deviation between the pressure taps and the PSP data.

An example of the data from a Mach number sweep at 5° angle of attack is shown in Figure 16.22. The pressure distribution on the wing is clearly visualized at the 60 and 80 m/s conditions. Note that between 46 and 80 m/s the dynamic pressure, and therefore, the pressure variation on the wing, increases by a factor of over 3. It is difficult to see the pressure distribution on the wing at 46 m/s using the stretched scale necessary to display the Mach number sweep.

To demonstrate the PSP system capability at the 46 m/s speed, data from two beta conditions at this speed are displayed in Figure 16.23. In this case, the pressure map is stretched over a smaller range and the pressure distribution on the upper surfaced of the wing is clearly visualized. In fact, slight variations in the pressure distribution on the wing can be detected at the different test conditions, even at this low speed. The data in Figure 16.23 demonstrate that the binary PSP system can be used to obtain data in a larger tunnel at this low speed.

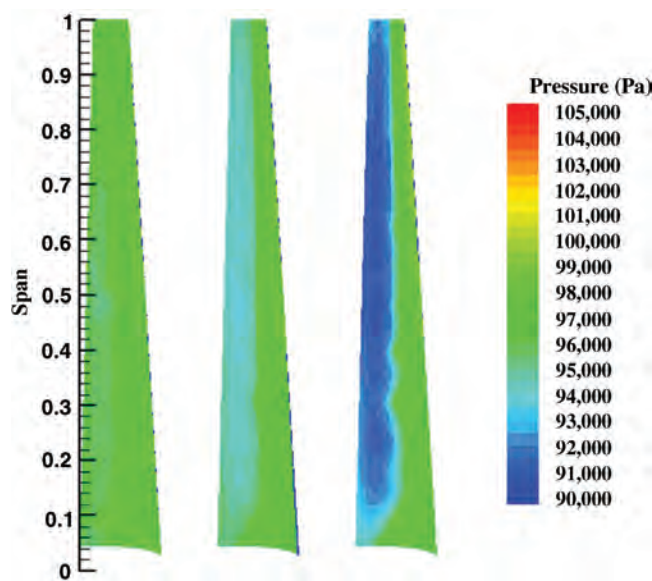


FIGURE 16.22 Binary PSP data at three Mach numbers. (See the color version of this figure in Color Plates Section.)

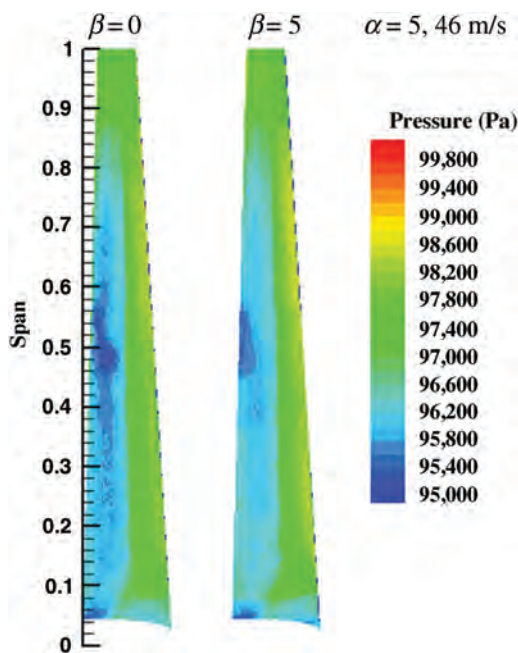


FIGURE 16.23 Binary PSP data at 46 m/s. (See the color version of this figure in Color Plates Section.)

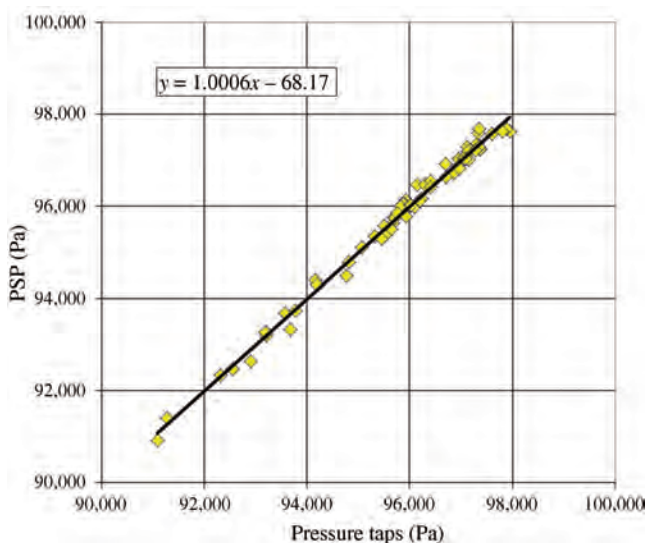


FIGURE 16.24 Comparison of PSP and pressure tap data. (See the color version of this figure in Color Plates Section.)

A linear curve fit to a plot of the corrected PSP data versus the pressure tap data at several test conditions is shown in Figure 16.24. Note that the slope is very close to 1, indicating that the *a priori* and *in situ* calibrations are in very good agreement. In this test campaign, the root mean squared deviation between the taps and the PSP was 148 Pa. While this may not serve as a formal means of evaluating the absolute accuracy and resolution of the PSP data, it yields some insight into the uncertainty level in the data. This noise level corresponds to 3.8% of the tunnel dynamic pressure at the 80 m/s test condition. This is approaching the target error budget of 3% for the low-speed system. It is noted that this error level is still about 11% of the dynamic pressure at the 46 m/s test condition. In any case, the data in Figures 16.22 and 16.23 represent very promising results for a low-speed PSP test in a larger wind tunnel.

16.6.3.5 Demonstration of the Color Camera System in the ADD Low-Speed Tunnel (Lee et al., 2010) The data presented in the previous sections were acquired using a single camera and a filter switch. While this has proven effective for low-speed PSP systems, further improvements to the system are under investigation. One such improvement involves the use of a color camera to replace the filter switch. The most significant advantage of the color camera system is that it acquires both the signal and the reference data simultaneously through a single optical system, and therefore, image alignment between the signal and reference channels is excellent. Other potential errors such as vibration of the model, variations in the model temperature, or variations in illumination that could occur between signal and reference images will be eliminated.

A preliminary demonstration of the color camera concept in a larger low-speed wind tunnel was conducted in the ADD low-speed tunnel test described in Section 16.6.3.4. During this test, a single Apogee 2000 color camera was installed in the wind tunnel, and data were acquired at a limited number of test conditions. Data were acquired using the same binary FIB PSP and illumination was provided by the LM2XX-400 LEDs. An

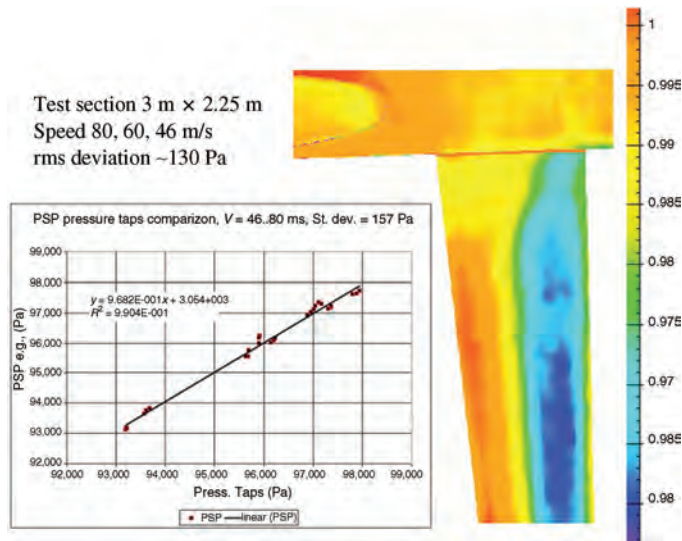


FIGURE 16.25 Binary PSP data using a color camera. (See the color version of this figure in Color Plates Section.)

example of data acquired using a color camera system is shown in Figure 16.25. The data produced by the color camera system compares favorably with the filter switch system. The root mean squared deviation between the PSP data and the pressure taps was 157 Pa for the color camera data. This is very close to the result obtained with the more mature filter switch system. As described earlier, the major limitation of the color camera system is related to the limited number of pixels available on this camera. The effective size of this camera is 600 pixels by 800 pixels and this limits either the field of view or the spatial resolution of the system. As cameras with higher pixel density are released onto the market, this type of color camera system should be investigated further.

16.6.4 Pressure-Sensitive Paint on Rotating Machinery (Juliano et al., 2011)

Experimental measurements of the mean and unsteady pressure distribution on rotating machinery such as compressors, propellers, and rotorcraft blades have proven difficult. These measurements are difficult due to the cost and complexity of installing a significant number of pressure transducers that operate in the rotating frame. Furthermore, pressure data from experimental measurements are available only at discrete points, and therefore, are unlikely to capture the distribution of spatial frequencies or peak pressures on the rotating surface. A means of acquiring measurements of the mean and unsteady pressure with high spatial resolution on the rotating surface would be of significant value. A demonstration of PSP measurements on a rotating surface has recently been performed; the experimental procedure and results are described here.

Lifetime-based PSP measurements have proven to be effective for minimizing model deformation errors in production PSP systems. These systems employ pulsed LED arrays and gated integration cameras with accumulation to acquire data from two gates in the decay curve of an excited pressure-sensitive molecule (Figure 16.7). Pulsed LED arrays produce about 100 μJ per pulse, and therefore, the use of the accumulation option on the

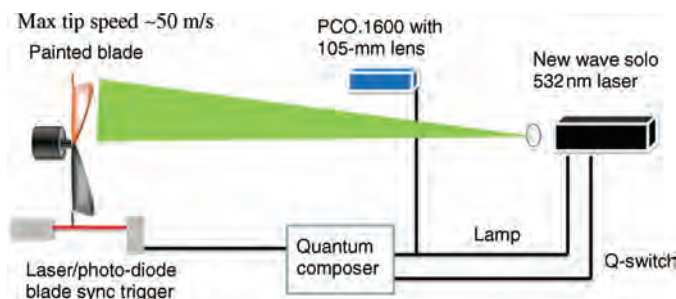


FIGURE 16.26 Experimental setup for PSP measurements on a model airplane propeller. (See the color version of this figure in Color Plates Section.)

cameras allows multiple pulses to be averaged on chip to improve the signal to noise of the measurement. This type of averaging requires that the flow be stable for several seconds as the signal is accumulated for each gate. The flow on a rotating surface, such as a helicopter rotor or aeroelastic surface is subject to unsteady fluid-structure interactions, and therefore, signal averaging is not an option. It is essential that all data from both gates be acquired on a single illumination pulse.

To accomplish this goal of single-shot PSP data acquisition, a single-shot lifetime-based PSP system has been developed and demonstrated. Excitation for the PSP is provided by a NewWave Solo 532 nm PIV laser. This laser produces 120 mJ per pulse, an increase in power of over 100 compared to an LED array. A PCO.1600 CCD camera operating in Frame Transfer mode is used for data acquisition. In this mode, the camera acquires data for a specific time using an electronic shutter. The frame is then shifted to the storage portion of the array and the CCD reexposes. These two frames are then read out and two-gate lifetime data, as described in Figure 16.7, are stored and processed. Timing may be controlled by externally triggering the camera to synchronize it with the laser and propeller. The camera and laser timing are controlled using a pulse generator.

A bench top experimental setup (Figure 16.26) was constructed using a model airplane propeller and electric motor (Electrify V-Pitch GPMG4501) to demonstrate single-shot PSP measurements on a propeller. The blade radius was 12.6 cm and the chord was 2.5 cm. At a rotation rate of 70 Hz, the tip speed of this blade is about 55 m/s; therefore, the theoretical maximum pressure differential to be expected on the blade is around 3600 Pa (assuming $\Delta C_p = 2$). This experiment is a very aggressive demonstration as it combines two of the most challenging aspects of PSP, low-speed flows and rotating machinery. One blade was painted with a Ruthenium in RTV with Silica Gel PSP. While this is not a fast-responding PSP, the goal of this experiment was to demonstrate the single-shot data acquisition system on a rotating airfoil. Unsteady measurements could be conducted using the same system with a fast-responding PSP.

The laser was positioned approximately 6 m from the rotor hub while the camera was at a distance of about 3.5 m. A 75-mm lens was used to expand the beam which then passed through a piece of ground glass to diffuse the beam. PSP systems using lasers as light sources can suffer from pulse-to-pulse variations in intensity and mode shape, and high frequency intensity variations caused by speckle. Both issues are negated here by acquiring both images from the same laser pulse, with the same camera, using the same optical elements. Each image contains the same illumination intensity pattern so a ratio of these images removes the intensity variations.

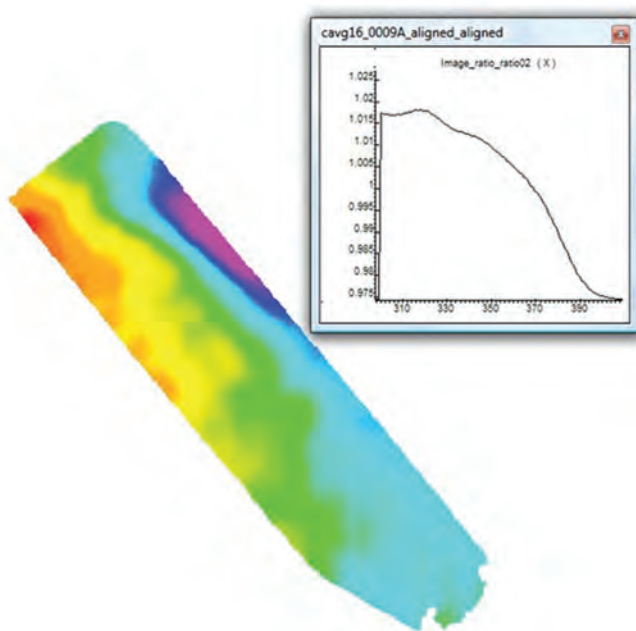


FIGURE 16.27 Binary PSP data using a color camera. (See the color version of this figure in Color Plates Section.)

The rotor was set to a low RPM to simulate a wind-off condition and a series of image pairs were acquired and averaged to minimize the wind-off shot noise. The rotor was then set to its highest operating RPM (55 m/s tip speed) and a series of single shot image pairs were acquired. The signal level in these single shot images is about 65% of the dynamic range of the camera, or about 25,000 photoelectrons. A quick calculation combining the shot noise ($\sim 0.6\%$ per image) and pressure sensitivity ($\sim 3\%$ per psi) suggest a best case noise level of about 1700 Pa for these data. Note that the dynamic pressure at the tip is only 1800 Pa, and therefore, the resulting data would have a noise level in C_p of about 1. The noise level can be improved with some low pass filtering (e.g., 16 pixel filter) to about 150 Pa ($C_p \sim 0.08$) with the sacrifice of some spatial resolution. The final wind-off ratio to wind-on ratio data are presented in Figure 16.27. Note the low-pressure region on the leading edge of the blade. The chord-wise variation in pressure across the blade would suggest a ΔC_p of about 2 for this propeller blade.

The data presented in Figure 16.27 are a significant accomplishment in that this data was acquired in single-shot mode, at low speed, on a rotating blade, and from a tunnel-scale operating distance. This data represent only the first phase of development for the proposed system. The system must integrate a fast-responding PSP and then it could be used to study a variety of unsteady flows such as aeroelasticity, flutter, supersonic inlets, or generation of acoustic noise.

16.6.5 Fast-Responding Pressure-Sensitive Paint

Typical paint formulations are comprised of an oxygen-sensitive fluorescent dye and a binder for physically attaching the dye to the model surface. Conventional formulations

typically use a polymer as a binder material. The disadvantage of the binder, however, is that it inhibits the interaction of the atmospheric oxygen and the embedded dye molecules. The response time of the paint to pressure is largely governed by the rate of diffusion of gas within the binder. Conventional, polymer-based paint formulations have response times on the order of 1 s, making them unsuitable for evaluating unsteady aerodynamic phenomena such as unsteady flows or aeroelastic phenomena.

The temporal-response characteristics of PSP are primarily governed by the thickness of the paint formulation and the diffusion coefficient of the binder material, according to the relation

$$\tau_{\text{diff}} \propto \frac{h^2}{D_m} \quad (16.21)$$

where the response time due to diffusion (τ_{diff}) increases with the paint thickness (h) squared and decreases with increasing diffusion coefficient (D_m). Some investigators have focused on decreasing the thickness of the paint in order to improve the response characteristics. This approach, however, has the disadvantage of sacrificing luminescent output from the paint and, thus, the signal-to-noise ratio (SNR). The paint formulation to be used in the proposed work has been developed based on the strategy of increasing the diffusivity of gas within the paint binder, as described by Gregory et al. (2008). Porous binders have been developed with the goal of enhancing the oxygen diffusion within the paint layer and, thus, improving the temporal response.

The difference between a conventional polymer-based PSP and a porous PSP is described schematically in Figure 16.28. For conventional PSP, oxygen molecules in a test gas must permeate into the binder layer for oxygen quenching. The process of oxygen permeation in a polymer binder layer produces slow response for a conventional PSP. On the other hand, the dye in a porous PSP is open to the test gas so that the oxygen

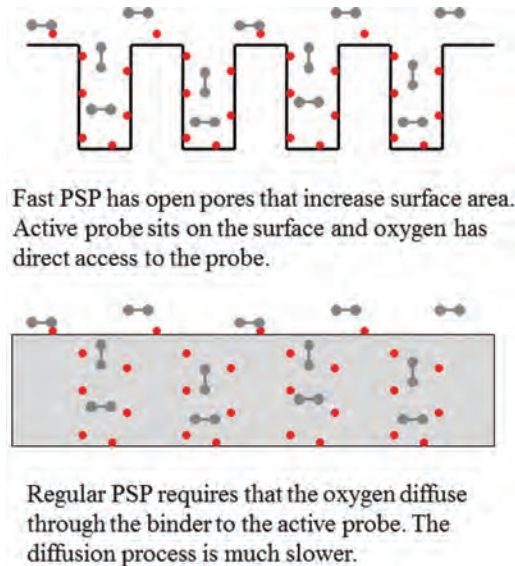


FIGURE 16.28 Comparison of the composition of porous PSP and conventional PSP. (See the color version of this figure in Color Plates Section.)

molecules are free to interact with the dye. The open binder creates a PSP that responds very quickly to changes in oxygen concentration and, therefore, pressure. A large effective surface area due to the porous surface improves luminescence intensity; thus, a higher SNR can be achieved. The drawback of the porous PSP approach is that the dye is too accessible to the oxygen. This results in near-complete quenching of all of the dye molecules at very low pressures. These formulations are effective for supersonic tunnels where the static pressure is below 15 kPa. For flows with higher pressures, the SNR ratio suffers.

Polymer/ceramic PSP (Scroggin et al., 1999) has been developed as a hybrid paint formulation that incorporates the advantages of both traditional and porous PSP. The polymer/ceramic formulation incorporates a high percentage of ceramic particles that provide the porous structure for rapid oxygen quenching, with a small amount of polymer to bind the paint to the surface. A dye is deposited onto the polymer/ceramic surface to complete the paint formulation. The resulting system is a fast-time-response paint layer with favorable SNR at higher pressure. Unlike anodized aluminum the polymer-based paint can be airbrushed onto a model; thus, paint application to complex surfaces is possible.

Experimental demonstrations of fast PSP have been conducted by several research teams. Gregory et al. (2007) have used these polymer/ceramic PSP formulations to measure oscillating pressure fluctuations with frequencies up to 20 kHz on a fluidic oscillator. Oscillating airfoils have been investigated by Fonov et al. (1999) using a thin coating of a binary PSP. In this experiment, the pressure distribution on a NACA-0012 oscillated at 20 Hz was investigated, thus demonstrating the capability of fast PSP on periodic flows. Nonperiodic PSP data were acquired by Kameda et al. (2005) on a delta wing in a Mach-0.6 flow. Here, Kameda detected the oscillation frequency of shocks on the delta wing at up to 170 Hz using frequency analysis of the PSP images. Nakakita (2007) demonstrated point-by-point frequency analysis of a fast PSP signal acquired on a cylinder in cross-flow.

These experiments demonstrate the potential of fast PSP as a tool for experimental studies and the evolution of the tool for wind tunnel settings. It is noted that the experiments included here incorporate many of the concepts previously demonstrated such as high-speed data acquisition and frequency analysis of the signal. The contribution of this work is in demonstrating the use of newer cameras and illumination sources to improve SNR of the PSP system. The improved cameras and LEDs are allowing fast PSP to be used as a tool in a study of fluids rather than focusing on the development of PSP capability.

16.6.5.1 Demonstration of Pressure-Sensitive Paint Measurements in a Short Duration Facility Another demonstration of the fast response PSP system involved the detection of the pressure distribution on an airfoil in a shock tube. The experimental setup (Figure 16.29) is composed of a small NACA-0012 airfoil at 2° angle of attack mounted near the end of a shock tube. The shock tube is fired with the endplate removed so there is no reflected shock, and therefore, the shock tube operates as a short duration transonic wind tunnel. After the shock passes by the airfoil, the flow is transonic with the actual velocity dependent on the pressure ratio of the diaphragm rupture. The flow is generally about 200 m/s in this facility and the duration of the flow, which is based on the arrival of the reflected expansion fan, lasts about 3 ms. This run time is sufficient to acquire a single PSP image on the airfoil using the fast response PSP.

The shock tube has a 1 m driver section mated to a 2 m driven section. The driven section has a 5 in. by 3 in. square cross-section with 5 in. by 5 in. windows on both sides at the 1 m and 1.5 m location. The NACA-0012 airfoil mounts onto a blank window in the

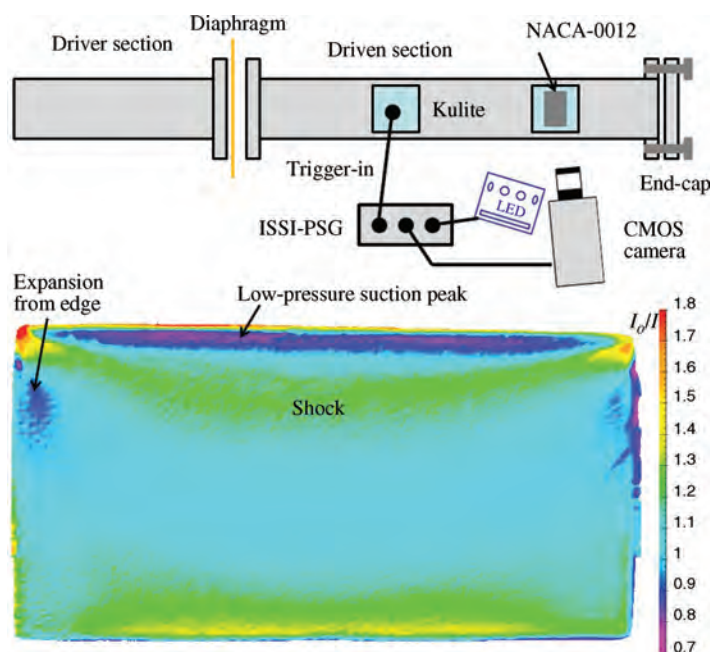


FIGURE 16.29 PSP measurements on an airfoil in the ISSI shock tube. (See the color version of this figure in Color Plates Section.)

downstream location and optical access is provided from the opposite window. A Kulite is mounted in a blank window in the upstream location to provide a trigger signal to the pulse generator. The pulse generator is programmed with a short delay (~ 1 ms) to allow the shock to pass and the transonic flow to establish itself on the airfoil and then the camera and LED are fired.

An example of the flow on the suction side airfoil with the flow at about 200 m/s is shown in Figure 16.29. The flow accelerates to a supersonic speed around the leading edge of the airfoil, and a shock forms at about 15% chord. Weak expansions can be seen on the sides of the airfoil as well. It is noted that this simple setup does not produce high quality flow, repeatable operation, or meaningful Reynolds numbers. The point of this setup is to demonstrate the use of PSP in a short duration facility such as a shock tunnel or Ludwig tube.

16.6.5.2 Unsteady Pressure Measurements (Crafton *et al.*, 2011) Recently, a demonstration of unsteady PSP measurements in a transverse jet injected into a supersonic flow was conducted at Wright-Patterson AFB (AFRL/RZA). Four injection blocks tested were tested in a Mach 2 flow at a series of injection pressures. This test included both fast PSP and traditional binary PSP for mean pressure measurements, presented in Figure 16.13; here, we present data from a fast-responding PSP. The fast PSP system utilized a Photron SA5 CMOS camera with an LM2XX-400 LED for illumination and a PtTFPP-PP fast PSP.

The fast PSP data from the 3/16 in. normal injector block were processed and an example of a sequence of images is shown in Figure 16.30. Here, the pressure distribution from the porous polymer is shown at three time steps along with the mean pressure distribution. In

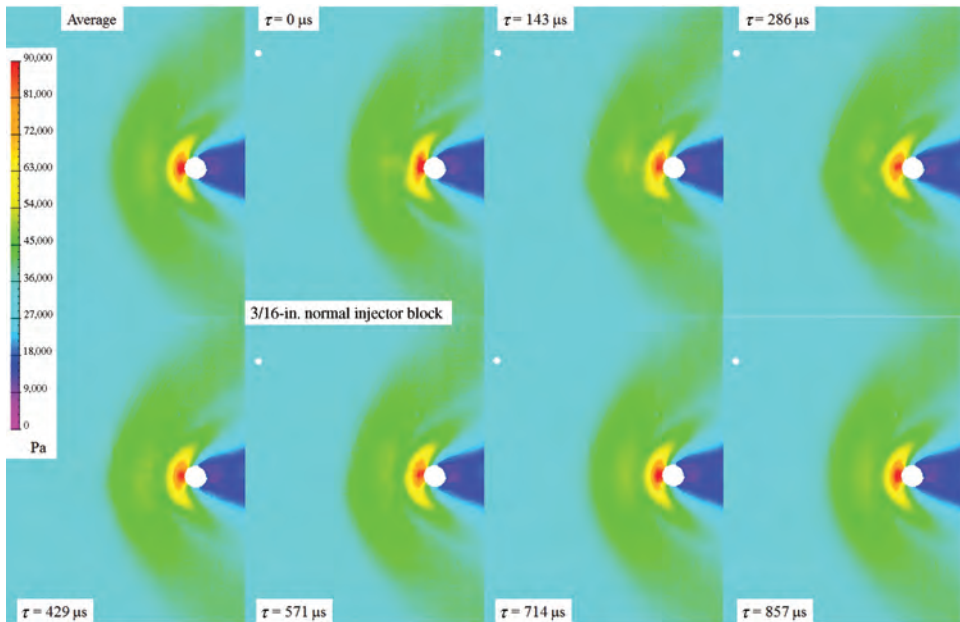


FIGURE 16.30 Mean and unsteady pressure distribution on the 3/16 in. normal injection block operating at 703 kPa. (See the color version of this figure in Color Plates Section.)

the first time step, the stagnation zone is compressed and asymmetric and the ridge of higher pressure associated with the front of the horseshoe vortex upstream of the stagnation zone is disrupted. It is possible that a feature in the flow, for example, a turbulent structure in the boundary layer, has moved through the flow and impacted the jet. As the system recovers, the stagnation zone expands and the pressure in the stagnation zone is lower than the average condition. The secondary pressure ridge is reforming as well. Finally, the flow returns to near its average position. This entire process occurs in less than 1 ms.

The data presented in Figure 16.30 represent only a fraction of the potential of the current fast PSP system. Using the fast PSP system, over 5000 samples similar to those in Figure 16.30 were acquired. Furthermore, the data were acquired in a continuous stream at 7 kHz, and therefore, can be presented and analyzed in a manner similar to traditional fast pressure tap data. The pressure data at a particular pixel can be plotted frame by frame, as shown in Figure 16.31, to produce a pressure-versus-time plot. In this case, the data from four specific locations are presented. The pressure in the freestream region, upstream of the bow shock, is nearly constant. Near the bow shock, the pressure fluctuates with an amplitude of several kPa. In the stagnation zone, the pressure fluctuations are similar to those near the bow shock but the amplitude of the fluctuations is up to 10 kPa. Behind the jet, the pressure drops, as does the amplitude of the fluctuations. It appears from the time trace that some of the larger scale pressure fluctuations are correlated in the bow shock, stagnation zone, and expansion zone regions.

The power spectrum of data presented in Figure 16.31 can be computed to investigate the frequency content of the flow. The amplitude of the power spectrum for each data set presented in Figure 16.31 was computed, and the resulting data are presented in Figure 16.32. From the data presented in Figure 16.32, it appears that the majority of

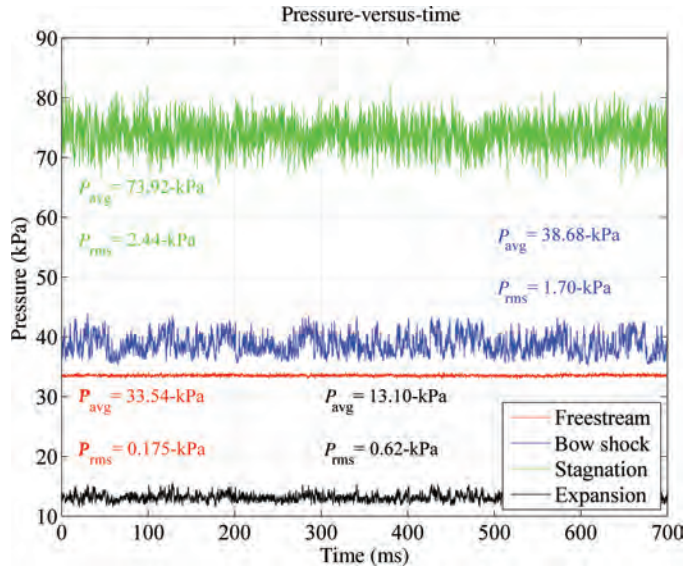


FIGURE 16.31 Fast PSP data as a function of time at four locations. (See the color version of this figure in Color Plates Section.)

the frequency content of the flow is below 500 Hz. There are no clear peaks in any of the spectra, suggesting that there is no fundamental frequency content to this flow. The freestream data contain no significant pressure fluctuations, and therefore, there is no significant information in the amplitude of the power spectrum. It is possible that this data set represents a noise floor of the detection scheme, or of the flow. The power

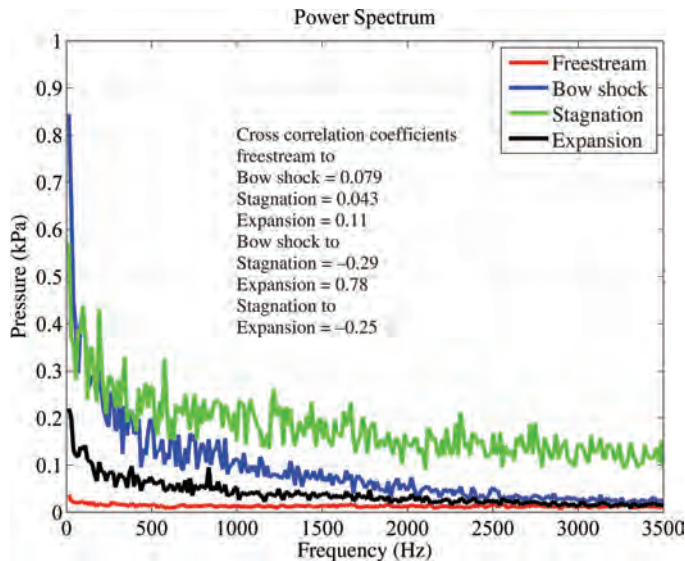


FIGURE 16.32 Amplitude of the power spectrum from the fast PSP pressure data presented in Figure 16.31. (See the color version of this figure in Color Plates Section.)

spectrum of the data near the bow shock and behind the jet both decay to a value near this noise floor by about 2 kHz. This would suggest that the 7 kHz data acquisition rate is fully resolving the frequency content in these regions of the flow. The power spectrum of the data near the stagnation point, by contrast, is indicating significant pressure fluctuations out to the 3.5 kHz limit of this data set. A data acquisition system with higher bandwidth would be of value for interrogation of this flow. In fact, several data sets were acquired at 25 kHz with a lower spatial resolution.

It is recognized that the data and analysis presented in Figures 16.31 and 16.32 is identical to the data and analysis that can be obtained with a few traditional fast pressure transducers. The major advantage of the current system is that there are approximately 1 million transducers available for the analysis. This data can be extracted at each pixel location and analyzed as shown in Figure 16.31. Quantities such as the mean pressure and deviation of the pressure from the mean at each spatial location can be computed. The original data set included over 1 million data points at 5,000 time steps. To mitigate the size of the data set, a 10-pixel binning filter was applied to the data. The resulting data set represents an array of 10,000 fast pressure transducers that have been sampled 5,000 times at a 7 kHz rate. While some spatial resolution has been sacrificed, the binning filter does improve the shot noise of the measurement by a factor of 10, and the size of the data set is manageable. The mean and standard deviation of the pressure was computed. Not surprisingly, the pressure fluctuations are largest in the stagnation zone where the amplitude is about 5 kPa. There is a second region of large pressure fluctuations associated with the bow shock. These may be related to fluctuation in the strength or the location of the bow shock. It is noted that these pressure fluctuations showed a strong correlation with the pressure fluctuations in the expansion region behind the jet (Figure 16.32). A final set of large pressure fluctuations are present near the location of the horseshoe vortex and just upstream of the stagnation zone. These fluctuations have a magnitude of about 2 kPa, similar to those in the bow shock. It is likely that there is some correlation between these pressure fluctuations given the similar structure and amplitude.

A final demonstration of the fast PSP system capability is the ability to analyze the frequency content of the flow with high spatial resolution. The data collected by the fast PSP system can be processed spectrally, as shown in Figure 16.32, at each pixel to produce a map of the amplitude of the power spectra at each frequency. An example of these maps at six frequency bins is shown in Figure 16.33. It is interesting to compare these maps, which represent the amplitude of the pressure fluctuations within a specific frequency range, with the overall pressure fluctuations presented in Figure 16.33. In the lowest frequency bin (~ 14 Hz), the flow shows a slight asymmetry in the stagnation zone, the bow shock, and the horseshoe vortex. There is no asymmetry present in the pressure fluctuation data in Figure 16.33. At the next frequency bin (~ 27 Hz); the flow is again symmetric at all locations. As the frequency increases, the amplitude of the pressure fluctuations is dropping in most locations. Around 287 Hz, a set of expansion fans begin exhibit fluctuations in the flow. By 438 Hz, the fluctuations of the expansion fans are effectively gone, but the flow now exhibits a bifurcated node structure at several locations. Interestingly, these node structures are only present at this frequency bin. At frequencies above about 2 kHz the only significant pressure fluctuations are associated with the stagnation zone in front of the jet. A sequence of several hundred of these maps can be combined into a movie, thus allowing the frequency content of the flow to be visualized spatially.

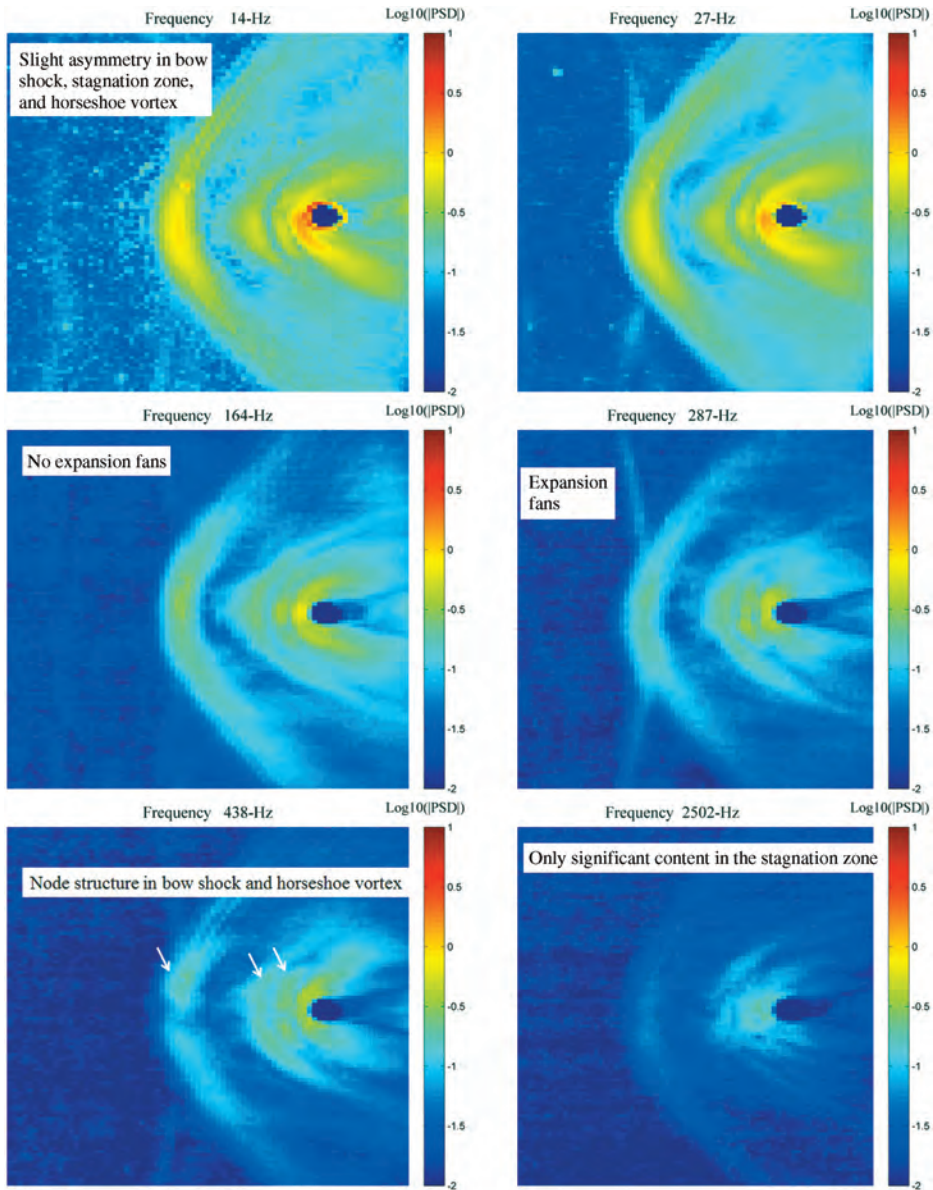


FIGURE 16.33 Map of the power spectrum amplitude at several frequencies. (See the color version of this figure in Color Plates Section.)

ACKNOWLEDGMENTS

We would like to acknowledge the support of funding from the National Science Foundation (Award numbers 0517782 & 0929864), the Air Force Office of Scientific Research (STTR AF04-T001; F49620-01-1-036), and the University of Washington Royalty Research Fund (Grant No. 65-1295), which has made the research performed at the UW possible.

REFERENCES

- Amao Y, Okura I. Optical oxygen sensor devices using metalloporphyrins. *Journal of Porphyrins and Phthalocyanines* 2009;13(11):1111.
- Baron A. *On time- and spatially-resolved measurements of luminescence-based oxygen sensors [Phd thesis]*. University of Washington, Seattle, WA; 1996.
- Baron AE, Danielson JDS, Gouterman M, Wan JR, Callis JB, McLachlan B. Submillisecond response-times of oxygen-quenched luminescent coatings. *Review of Scientific Instruments* 1993;64:3394–402.
- Bell J. Accuracy limitation of lifetime-based pressure sensitive paint measurements. 19th ICIASF. NASA Ames Research Center, Moffett Field, CA; 2001.
- Bell J. Application of pressure-sensitive paint to testing at very low speeds. AIAA 2004-0878; 2004.
- Bell JH, Schairer ET, Hand LA, Mehta RD. Surface pressure measurements using luminescent coatings. *Annual Review of Fluid Mechanics* 2001;33:155–206.
- Bencic TJ. Rotating pressure and temperature measurements on scale-model fans using luminescent paints [Technical Report]. Japan Aerospace Exploration agency; 1998.
- Birks JB. *Photophysics of Aromatic Molecules*. London: Wiley-Interscience; 1970.
- Borek C. et al., Highly efficient, near-infrared electrophosphorescence from a Pt-metalloporphyrin complex. *Angewandte Chemie International Edition* 2007;46:1109–1112.
- Brown O. Low-speed pressure measurements using a luminescent coating system [PhD thesis]. Stanford University; 2000.
- Bykov A, Fonov SD, Mosharov V, Orlov A, Pesetsky V, Radchenko V. Study result for the application of two-component PSP technology to aerodynamic experiment. AGARD. Seattle, USA; 1997.
- Carlson B, Gouterman M. U.S. Patent, 5965642 (March 11); 1997.
- Carlson B, Bullock JP, Hance TM, Phelan GD. Barometric sensitive coatings based upon osmium-complexes dissolved in a fluoroacrylic polymer. *Analytical Chemistry* 2009;81(1):262.
- Coyle L. Lifetime measurements on pressure sensitive paints : temperature correction, effects of environment, and trials on new luminescent materials. PhD thesis, University of Washington; 1999.
- Crafton J, Fonov SD, Hsu K, Carter CD, Gruber MR. Optical measurements of pressure and shear on a strut in supersonic flow. AIAA-2009-5033;2009.
- Crafton JW, Fonov SD, Goss LP, Jones EG, Reeder MF. Comparison of radiometric and lifetime based pressure-sensitive paints for low speed pressure measurements. AIAA-2006-1;2006.
- Crafton J, Fonov SD, Jones E, Fonov V, Goss L, Tyler C. Simultaneous measurements of pressure and deformation on a UCAV in the SARL. AIAA-2005-1028;2005.
- Crafton JW, Forlines R, Palluconi S, Hsu K, Carter CD. Investigation of transverse jet injection in a supersonic crossflow using fast responding pressure-sensitive paint. AIAA-2011-3522. 2011.
- Demas JN, DeGraff BA, Coleman P. Oxygen sensors based on luminescence quenching. *Analytical Chemistry* 1999;71:793A–800A.
- Di Natale C, Monti D, Paolesse R. Chemical sensitivity of porphyrin assemblies. *Materials Today* 2010;13(7–8):37.
- Engler R, Fey U, Henne U, Klein C, Sachs W. Quantitative wind tunnel studies using pressure- and temperature sensitive paints. *Journal of Visualization* 2005;8:277–284.
- Fischer LH, Stich MIJ, Wolfbeis OS, Tian N, Holder E, Schaferling M. Red- and Green-Emitting Iridium(III) Complexes for a Dual Barometric and Temperature-Sensitive Paint. *Chemistry—A European Journal* 2009;15(41):10857.
- Fonov SD, Engler RH, Klein C, Mihailov SV, Mosharov VE, Kulesh VP, Radchenko VN, Schairer E. Pressure sensitive paint for oscillating pressure fields measurements. 18th ICIASF; 1999; p. 24.1–24.4.

- Gewehr PM, Delpy DT. Optical oxygen sensor based on phosphorescence lifetime quenching and employing a polymer immobilized metalloporphyrin probe. 1. Theory and instrumentation. *Medical & Biological Engineering & Computing* 1993;31:2.
- Goss L, Trump D, Sarka B, Lydick L, Baker W. Multi-dimensional time-resolved pressure-sensitive-paint techniques: a numerical and experimental comparison. 37th Aerospace Sciences Meeting; Jan. 1999, Reno Nevada.
- Gouin S, Gouterman M. Ideality of pressure-sensitive paint. II. Effect of annealing on the temperature dependence of the luminescence. *Journal of Applied Polymer Science* 2000;77:2805.
- Gouterman MP. In: Dolphin D, editor. *The Porphyrins*. Vol. III. New York: Academic Press; 1978. p. 1–165
- Gouterman M. Oxygen quenching of luminescence of pressure sensitive paint for wind tunnel research. *Journal of Chemical Education* 1997;74:697.
- Gouterman M, Hall RJ, Khalil G-E, Martin PC, Shankland EG, Cerny RL. Tetra (pentafluorophenol) porpholactone. *Journal of American Chemical Society* 1989;111:3702.
- Gouterman M, Callis J, Dalton LR, Khalil GE, Mébarki Y, Cooper KR, Grenier M. Dual luminophor pressure sensitive Paint III: Application to automotive model testing. *Measurement Science and Technology* 2004;15:1986.
- Gregory JW, Sullivan JP, Raman G, Raghu S. Characterization of the microfluidic oscillator. *AIAA Journal* 2007;45(3):568–576.
- Gregory JW, Asai K, Kameda M, Liu T, Sullivan JP. A Review of Pressure-Sensitive Paint for High Speed and Unsteady Aerodynamics. Proceedings of the Institution of Mechanical Engineers, Part G. *Journal of Aerospace Engineering* 2008;222(2):249–290.
- Holmes JW. Analysis of radiometric, lifetime and fluorescence lifetime imaging. *The Aeronautical Journal Paper* 1998;2306:189–194.
- Juliano TJ, Kumar P, Peng D, Gregory JW, Crafton J, Fonov SD. Single-shot, lifetime-based pressure-sensitive paint for rotating blades. *Measurement Science and Technology* 2011;22:085403.
- Kadish K, Smith K, Guillard R, editors. *The Porphyrin Handbook*. Vol. 6–7. New York: Academic Press; 2000.
- Kameda M, Tabei T, Nakakita K, Sakaue H, Asai K. Image measurements of unsteady pressure fluctuation by a pressure-sensitive coating on porous anodized aluminum. *Measurement Science and Technology* 2005;16(12):2517–2524.
- Kavandi J, Callis J, Gouterman M, Khalil G, Wright D, Burns D, McLachlan B. Luminescent barometry in wind. *Tunnels Review of Scientific Instruments* 1990;61:3340–3347.
- Khalil G, Chang A, Gouterman M, Callis JB, Dalton LR, Turro NJ, Jockusch S. Oxygen pressure measurement using singlet oxygen emission. *Review of Scientific Instruments* 2005;76:054101.
- Khalil G, Costin C, Crafton J, Jones E, Grenoble S, Gouterman M, Callis J, Dalton L. Dual luminophor pressure sensitive paint I: ratio of reference to sensor giving a small temperature dependence. *Sensors and Actuators B* 2000;97(1):13–21.
- Khalil GE, Costin C, Crafton J, Jones G, Grenoble S, Gouterman M, Callis JB, Dalton LR. Dual luminophor pressure sensitive Paint I: Ratio of reference to sensor giving a small temperature dependency. *Sensors and Acutators B* 2004;97:13
- Khalil GE, Daddario P, Lau KSF, Imtiaz S, King M, Gouterman M, Sidelev A, Puran N, Ghandehari M, Brückner C. *Meso* -Tetraarylporpholactones as High pH Sensors. *Analyst* 2010;135(8):2125.
- Khalil G, Gouterman M, Green E. *Method for measuring oxygen concentration*. USA Patent No. 4,810,6551989.
- Khalil GE. Electronic spectra and structure of some group VB, VIB, VIIB metal porphyrins. M.Sc. Thesis, University of Washington; 1973.
- Khalil G et al. Synthesis and spectroscopic characterization of Ni, Zn, Pd, & Pt tetra(pentafluorophenyl)porpholactone with comparisons to Mg, Zn, Y, Pd, and Pt metal complexes of tetra(pentafluorophenyl)porphine. *Journal of Porphyrins and Phthalocyanines* 2002;6:135–145.

- Khalil GE et al. NIR Luminescence of Gadolinium Porphyrin Complexes. *Chemical Physics Letters* 2007;435:45.
- Khan AU. *Singlet Molecular Oxygen Spectroscopy: Chemical and Photosensitized* Vol. I, editor, Frimer AA, Florida: CRC Press;1985. p. 39–80.
- Klein C, Engler R, Fonov SD, Trinks O. Pressure sensitive paint measurements on a wing model in a low-speed wind tunnel. 18th ICIASF;1999.
- Lakowicz JR. *Principles of Fluorescence Spectroscopy*. Plenum Press; 1983. p. 1.
- Lee J, Kim S, Ko J, Chung I, Fonov SD, Forlines R, Crafton J. Pressure sensitive paint system for measurements in a large low speed wind tunnel. AIAA-2010-4917;2010
- Liu T, Sullivan JP. *Pressure and Temperature Sensitive Paints*. Berlin: Springer-Verlag; 2004.
- McDonagh C, Burke CS, MacCraith BD. Optical chemical sensors. *Chemical Reviews* 2008;108 (2):400.
- McGraw C, A thin film oxygen sensor for the study of insect flight. PhD thesis. Department of Chemistry, University of Washington; 2004.
- Mosharov VE, Radchenko VN, Fonov SD. *Luminescent Pressure Sensors in Aerodynamics*. Moscow: Central Aerodynamic Institute (TsAGI); 1998.
- Nakakita K. Unsteady pressure distribution measurement around 2D-cylinders using pressure-sensitive paint. AIAA-2007-3819;2007.
- Schaap AP, editor. *Singlet Molecular Oxygen*. Hutchinson & Ross New York: 1976.
- Scroggin AM, Slamovich EB, Crafton JW, Lachendro N, Sullivan JP. Porous polymer/ceramic composites for luminescence-based temperature and pressure measurement. Proceedings of the Materials Research Society Symposium;1999. Vol. 560, p. 347–352.
- Spellane PJ, Gouterman M, Antipas A, Kim S, Liu YC. *Inorg Chem* 1980;19:386.
- Vardaki E, Stokes N, Fonov SD, Crafton J. *Pressure sensitive paint measurements at the ARA transonic wind tunnel*. AIAA-2010-4796.
- Ruyten W, Sellers M. On-line processing of pressure-sensitive paint images. *Journal of Aerospace Computing, Information, and Communication* 2004;1(9):372–382.
- Wan JR. Fast response luminescent pressure sensitive coating and derivatives of Tetra(pentafluorophenyl)porpholactone. PhD thesis, University of Washington, Nov; 1993.
- Wasserman HW, Murray RW. editors. *Singlet Oxygen*. New York: Academic Press;1979.
- Wolffbeis OS. In: Wolffbeis OS, editor. *Oxygen Sensor, Fiber Optic Chemical Sensors and Biosensors*. Vol. 2. Boca Raton (FL): Chemical Rubber Company;1991.
- Xu Wy, McDonough RC, Langsdorf B, Demas JN, Degraff BA. Oxygen sensors based on luminescence quenching - interactions of metal-complexes with the polymer supports. *Analytical Chemistry* 1994;66(23):4133.
- Xu WY, Schmidt R, Whaley M, Demas JN, Degraff BA, Karikari EK, Farmer BL. Oxygen sensors based on luminescence quenching - interactions of pyrene with the polymer supports. *Analytical Chemistry* 1995;67(18):3172.
- Puklin E, Carlson B, Gouin S, Costin C, Green E, Ponomarev S, Tanji H, Gouterman M. Ideality of pressure-sensitive paint. I. Platinum tetra(pentafluorophenyl)porphine in fluoroacrylic polymer. *Journal of Applied Polymer Science* 2000;77:2795–2804.
- Jockusch S, Turro NJ, Thompson EK, Gouterman M, Callis JB, Khalil GE. Singlet molecular oxygen by direct excitation. *Photobiological Science* 2008;7(2):235–239.
- Zebger I, Snyder JW, Andersen LK et al. Direct optical detection of singlet oxygen from a single cell. *Photochemistry and Photobiology* 2004;79:319–322.

17

FLOW MEASUREMENT

JESSE YODER

17.1 New-technology and traditional technology flowmeters

- 17.1.1 New-technology flowmeters
- 17.1.2 Coriolis flowmeters
- 17.1.3 Magnetic flowmeters
- 17.1.4 Ultrasonic flowmeters
- 17.1.5 Vortex flowmeters
- 17.1.6 Thermal flowmeters
- 17.1.7 Traditional technology flowmeters
- 17.1.8 Differential pressure (DP) flowmeters
- 17.1.9 Positive displacement flowmeters
- 17.1.10 Turbine flowmeters
- 17.1.11 Open-channel flowmeters
- 17.1.12 Variable area flowmeters

17.2 Trends in flow measurement

Further readings

Flow is one of the most important variables measured in industrial and commercial environments. Whether the fluid is oil, gas, steam, air, or water, there are many occasions when it is vitally important to know how much flow is traveling through a pipe. Measurement of petroleum liquids such as oil has become especially important, as the price of oil has increased to nearly \$100 per barrel. And measuring natural gas at multiple points along the distribution chain for natural gas is also critical for custody transfer, where ownership of the natural gas changes hands. In addition, some flow is measured in rivers and streams, rather than in closed conduits, and this is called open-channel flow measurement.

17.1 NEW-TECHNOLOGY AND TRADITIONAL TECHNOLOGY FLOWMETERS

Flow is measured by flowmeters that either measure the volume of flow or its mass. A flowmeter is a device designed to measure the quantity of fluid flow in an open channel or closed pipe. Flowmeters are often divided into two broad categories

- New-technology flowmeters
- Traditional technology flowmeters.

The difference between new and traditional technology flowmeters has to do with when they were introduced, their accuracy and reliability, and the methods they use to measure flow.

New-technology flowmeters are distinguished by the following characteristics

1. They were introduced into the commercial market in 1950 or after.
2. They incorporate technical advantages that avoid some problems with earlier flowmeters.
3. They are the focus of more product development by suppliers than are traditional technology flowmeters.
4. Their level of performance, including reliability and accuracy, is better than that of the pre-1950 flowmeters.
5. They typically have been quick to adopt modern communication protocols such as highway addressable remote transducer (HART), Foundation Fieldbus, and Profibus.

Traditional technology flowmeters, by contrast, have been around for many years, in some cases as long as several 100 years. They have the following characteristics

1. They were introduced into commercial and industrial environments before 1950.
2. They typically require more maintenance than new-technology flowmeters.
3. Suppliers generally do less product development on traditional technology flowmeters than on new-technology flowmeters.
4. They typically are not as accurate or reliable as new-technology flowmeters.
5. They have been relatively slow to adopt modern communication protocols such as HART, Foundation Fieldbus, and Profibus.

17.1.1 New-Technology Flowmeters

New-technology flowmeters are so-called because they represent new methods of flow measurement that postdate the traditional methods of turbine, positive displacement, and differential pressure. Most new-technology flowmeters were introduced in the 1960s and 1970s, although magnetic flowmeters came onto the market even earlier than that. The following are the five types of new-technology flowmeters

- Coriolis
- Magnetic
- Ultrasonic

- Vortex
- Thermal

17.1.2 Coriolis Flowmeters

Coriolis flowmeters get their name from the French mathematician Gustave Coriolis. Coriolis showed in 1835 that it is necessary to take an inertial force into account when describing the motion of an object in a rotating frame of reference. This effect is often illustrated by using the rotating Earth as an example. An object hypothetically thrown from the North Pole toward a target on the equator will miss its mark because, by the time it arrives, the target will have rotated from its previous position with the rotation of the Earth. This same example is sometimes illustrated using a merry-go-round instead of the rotating Earth. Someone at the center of a merry-go-round who aims a ball at a target at the edge of the spinning object will miss the target if the merry-go-round is spinning fast enough because by the time the ball reaches the location of the target, the target will have moved on, along with the spinning merry-go-round.

The Coriolis effect as described in terms of the rotating Earth and a spinning merry-go-round is actually a perceptual effect rather than a force. A ball thrown from the North Pole or the center of a merry-go-round will appear to curve when viewed from the position the ball is thrown from. However, there is no actual force being exerted on the ball causing it to curve, apart from the force exerted on the ball when it is thrown. The ball only appears to curve because of the rotating frame of reference beneath it. Hence, this example describes only a Coriolis effect, not a Coriolis force (Figure 17.1).



FIGURE 17.1 A Coriolis flowmeter being calibrated at Colorado Engineering Experiment Station, Inc. (CEESI); photo by Flow Research.

The force that is involved with the motion of Coriolis flowmeters is less easy to understand, but it has to do with the fact that the meter is oscillating. Coriolis flowmeters are made up of one or more vibrating tubes, which are usually bent. The fluid accelerates as it moves through the tube toward the point of maximum vibration, and decelerates in the tube as it leaves this point. This motion causes the tubes to twist. Mass flow is directly proportional to the amount of twist in the tubes. Tube positions are sensed by position sensors.

Although the majority of Coriolis flowmeters have bent tubes, some have both single and dual straight tubes. KROHNE introduced the first commercially successful straight tube Coriolis flowmeter in 1994. Since that time, other vendors have begun offering straight-tube flowmeters. Straight-tube meters cause less pressure loss than bent-tube meters and also are less susceptible to clogging and easier to clean. This is especially important for food and other sanitary applications. While straight-tube meters are quite effective for liquid applications, suppliers have had very limited success in applying them to gas applications.

Suppliers have differentiated themselves by various tube designs. Most of the bent tube designs for Coriolis meters are proprietary to the manufacturers. In 1977, Micro Motion was the first company to introduce a commercially successful Coriolis meter, a bent-tube meter. Since that time, a number of other companies have entered the market, including Endress+Hauser, KROHNE, and Foxboro (Invensys). While most Coriolis meters have either single or dual tubes, whether bent or straight in design, in 2011 Endress+Hauser introduced a four-tube bent tube Coriolis meter.

For many years, between 85 and 90% of all Coriolis flowmeters line sizes were of 2 in. or less. Rheonik was the only company that manufactured Coriolis meters with line sizes above 6 in. In the past several years, three new suppliers, Micro Motion, Endress+Hauser, and KROHNE, have begun manufacturing Coriolis meters in large line sizes above 6 in. These large line size meters typically cost over \$50,000 and are mainly designed for custody transfer applications, especially in the oil and gas industry. Most of these large line size meters are 8, 10, or 12 in. However, Endress+Hauser's new four-tube Coriolis meter is designed for 14-in. line sizes.

17.1.3 Magnetic Flowmeters

Magnetic flowmeters were first introduced in 1952 in the Netherlands by a company called TOBI. Since 1952, they have come to dominate the water flow measurement market. They are used to measure the flow of conductive liquids and slurries, including black liquor and paper pulp slurries. Their main limitation is that they can only measure conductive liquids. As a result, they are not widely used in the petroleum industry, since hydrocarbons are nonconductive. Magnetic flowmeters do not create pressure drop and are highly accurate. Their purchase price can be relatively high, depending on line size. While some magnetic flowmeters are less than 1/4 in. in size, some can range up to 110 in. in size. The large size meters are typically used for water and wastewater applications.

Magnetic flowmeters, or magmeters as they are sometimes called, use Faraday's Law of Electromagnetic Induction. This principle states that a voltage is generated in a conductive medium when it passes through a magnetic field. The amount of this voltage is proportional to the length of the conductor, the magnetic field density, and the velocity of the conductive medium. Using Faraday's Law, these three variables are multiplied together to yield the magnitude of the voltage.

Magnetic flowmeters have wire coils mounted inside or outside the body of the flowmeter. Either alternating current (AC) or direct current (DC) is applied to these coils,



FIGURE 17.2 KROHNE magnetic flowmeters in Dordrecht, the Netherlands; photo by Flow Research.

which generates a magnetic field inside the meter body. As conductive liquid passes through the meter body, a voltage is generated due to the presence of the magnetic field. Electrodes on either side of the meter body detect this voltage. The flowmeter uses this voltage value to compute the flowrate.

Although magmeters can measure the flow of almost any type of liquid, they cannot be used to measure the flow of steam or gas. Almost 30% of magmeters are used in the water and wastewater industry, and many of these are relatively large in size. Because they can measure dirty liquids, they are widely used in the pulp and paper and metals and mining industries. They are also used in the chemical and food and beverage industries to measure the flow of liquids in manufacturing and food processing operations. Filling machines are another major application for magnetic flowmeters (Figure 17.2).

Because magnetic flowmeters are used to measure the flow of so many different types of liquids, different types of linings are used, depending on the type of application involved. Both hard and soft rubbers are used to measure dirty liquids, as they are very durable. Other liners include polytetrafluoroethylene, polyurethane, and Teflon. Sanitary liners are widely used in the chemical; food and beverage; and pharmaceutical industries to measure the flow of food, beverages, and chemicals.

17.1.4 Ultrasonic Flowmeters

Ultrasonic flowmeters were first commercially introduced in 1963 in Japan by Tokyo Keiki. Since 1963, ultrasonic flowmeters have come to be a dominant force in the flow measurement industry. Their advantages are many: they create very little pressure drop in the line, they are highly accurate, and they are very reliable over time. Suppliers have

advanced the technology of ultrasonic flowmeters over the years and have developed multipath ultrasonic meters that are highly accurate and are used for custody transfer applications. Ultrasonic flowmeters are mainly used for liquid and gas applications, although some are now being used to measure steam flow.

The dominant technology used in ultrasonic flowmeters is called transit time. Transit time meters have one or more ultrasonic transducers that send an ultrasonic pulse across the flowstream at an angle and back. A transducer on the other side of the pipe acts as a receiver. The flowmeter measures the “transit time” of the pulse from one side of the pipe to the other. When the signal travels with the flow, it travels faster than when it travels against the flow. The difference between these two transit times is proportional to flowrate.

When liquids contain a large number of impurities, Doppler ultrasonic meters are used. Doppler meters work like transit time meters except that instead of bouncing a signal off the other side of the pipe, they bounce a signal off the particles in the flowstream. These particles are traveling at the same velocity as the liquid flow. As the ultrasonic pulse bounces off the particles, a frequency shift in the signal occurs that is proportional to the velocity of the flow. A receiver detects this shift in frequency and uses it to compute flowrate.

Doppler ultrasonic meters are used to measure the flow of dirty liquids and slurries. Transit time meters are used mainly to measure the flow of clean liquids. However, in recent years, suppliers have made progress in enabling transit time meters to measure liquids with some impurities in them. Doppler meters are not as accurate as transit time meters, but measuring the flow of dirty liquids and slurries is a difficult measurement. Many end-users are now using transit time technology where formerly they would have used Doppler because of the improvements in the ability of transit time meters to handle somewhat dirty liquids (Figure 17.3).

Mounting type is an important consideration for ultrasonic flowmeters. Some ultrasonic meters use clamp-on technology, meaning that the transducers are clamped onto the pipe and the internal flow velocity is read by an external transmitter that sits somewhere on the outside of the pipe. Clamp-on flowmeters have a great advantage in that they are completely nonintrusive and have no impact on the flowstream. They are sometimes used as check meters to monitor the performance of another flowmeter and where it is desirable to



FIGURE 17.3 An ultrasonic flowmeter being tested at CEESI; photo by Flow Research.

measure flow at different locations. The chief disadvantage of clamp-on ultrasonic meters is that they are less accurate than inline ultrasonic meters. The ultrasonic signal has to pass through the pipe wall, and its thickness and material of construction can have an impact on the signal. Also, if there is any buildup on the side of the pipe, this can affect the internal diameter of the pipe and the accuracy of clamp-on ultrasonic meters. For these reasons, inline ultrasonic flowmeters are generally used where high accuracy is required.

Multipath ultrasonic flowmeters have been developed to increase the accuracy of ultrasonic flowmeters, and are some of the most accurate flowmeters made. Multipath meters contain three or more paths, with each path containing a pair of ultrasonic transducers. A path is simply the trail of an ultrasonic pulse as it travels across the pipe. Multipath meters have the advantage of measuring the transit time signal at multiple locations in the pipe, thereby increasing measurement accuracy. Most multipath ultrasonic meters have 4, 5, or 6 paths, but some meters have been developed with as many as 8 or 18 paths.

Multipath ultrasonic meters have received the approval of the American Gas Association (AGA) for use in custody transfer of natural gas. Because of their high accuracy and reliability, they are displacing turbine and differential pressure (DP) flowmeters that are also used for this purpose. Multipath ultrasonic flowmeters, which are always inline, are used for to measure the flow of natural gas in large pipelines at custody transfer points. Many of these pipelines are 24–42 in., or even larger. Until ultrasonic meters received AGA approval in 1998, turbine and DP flowmeters were mainly used for custody transfer.

17.1.5 Vortex Flowmeters

Vortex flowmeters are the most versatile type of flowmeter because they can reliably measure liquid, steam, and gas flows with relative ease. They are also able to withstand the high temperatures and pressures associated with steam flow, and hence are uniquely qualified to measure steam. More recently, vortex flowmeters have received approvals from the American Petroleum Institute (API) for use in custody transfer applications. This draft standard was first approved in 2007, and then revisited in 2010.

Vortex flowmeters make use of what is called the von Karman effect. According to this principle, fluid passing by a bluff body will alternately generate vortices that are used to measure flow. The vortex flowmeter counts the number of vortices generated by the bluff body, which is a broad object with a flat front that is inserted into the flowstream. Flow velocity is proportional to the frequency of the vortices. Flowrate is calculated by multiplying flow velocity times the area of the pipe.

Vortex flowmeters are widely used for steam flow measurement. Due to its high pressure and temperature, steam is the most difficult fluid to measure. How steam flow is measured also depends on the type of steam: whether it is wet, saturated, or superheated. Steam is measured for power generation purposes, and it is also often measured in process plants. Vortex meters can handle the high pressures and temperatures associated with steam. The other type of flowmeter widely used for steam flow measurement is differential pressure.

Vortex flowmeters offer reliable and accurate flow measurement at a reasonable price. However, they do not have the same compelling features that some other new-technology flowmeters have. They do not have the high accuracy of multipath ultrasonic or Coriolis flowmeters, for example. Neither do they have the ability of magnetic flowmeters to measure the flow of almost any type of water, dirty, or clean. Instead, vortex flowmeters offer reliable and accurate measurement of flow in a wide range of conditions.

17.1.6 Thermal Flowmeters

Thermal flowmeters are used almost entirely for gas flow measurement. They were developed in the 1970s by a group of researchers who had previously worked on hot-wire anemometers used to measure airflow in a laboratory environment. The laboratory research, conducted by Dr. John Olin and Dr. Jerry Kurz, was intended to investigate airflow turbulence and air velocity profiles. The group wanted to adapt the hot-wire anemometers to an industrial environment, but they were too fragile for that purpose. Drs. Olin and Kurz developed a thermal flowmeter that was more rugged and hardened than the hot-wire anemometers that worked well in laboratory environments.

Somewhat before the research done by Drs. Olin and Kurz, Fluid Components International (FCI) took a different approach to the development of thermal flowmeters. FCI developed thermal flow switches to detect oil flow through pipes in the oil patch. While FCI's thermal flow switches preceded the thermal flowmeters developed by Olin and Kurz, FCI did not actually create thermal flowmeters until 1981, when they put more sophisticated electronics on their switches, creating thermal flowmeters.

Thermal flowmeters work differently from other new-technology types. They actually introduce heat into the flowstream and measure how quickly this heat dissipates. Heat dissipation is measured using one of the two methods

- Constant temperature differential
- Constant current

The constant temperature differential method involves the use of two temperature sensors. One sensor is heated and the other measures the gas temperature. Mass flowrate is calculated based on how much electrical power is necessary to keep a constant difference in temperature between the two temperature sensors.

The constant current method also requires the use of two sensors. One sensor is heated and the other sensor measures the temperature of the flowstream. Electrical power is kept constant to the heated sensor. Mass flow is measured based on the difference between the flowstream temperature and the temperature of the heated sensors. Both methods use the principle that greater cooling results from higher velocity flows.

In the early 1990s, there was an environmental push to measure the amount of sulfur dioxide (SO_2) and nitrous oxide (NO_x). Thermal flowmeters found a niche in this market, which is referred to as continuous emissions monitoring (CEM). At this time, thermal flowmeter suppliers developed multipoint thermal flowmeters to more accurately measure the flow through large exhaust pipes and smokestacks. Single point thermal meters were not sufficiently accurate because they only measure flow at a point, while the multipoint meters measured flow at multiple points. Some multipoint meters measure flow at as many as 16 points.

Since 2008, the Obama administration has placed a renewed emphasis on monitoring greenhouse gas emissions. This has opened up new opportunities for thermal flowmeters in areas such as biomass gasification, measuring flue gas, monitoring flare gas flow, methane recovery from coal mines, and measuring landfill gases. These opportunities have given a substantial boost to sales of thermal flowmeters in the past several years. While there is still a need for CEM, these new applications are well suited to thermal flowmeters, which are almost exclusively used to measure gas flow.

17.1.7 Traditional Technology Flowmeters

Traditional technology flowmeters incorporate more traditional technologies than new-technology flowmeters. As a group, these flowmeters have been on the market longer than new-technology meters. Their maintenance requirements are typically higher than new-technology meters. And while some suppliers are bringing out new features and new products into the traditional technology flowmeter market, suppliers are less active in researching and developing traditional technology flowmeters than they are in bringing enhanced and upgraded new-technology flowmeters to market.

The following are the flowmeter types that fall into the traditional technology category:

- Differential pressure
- Positive displacement
- Turbine
- Open channel
- Variable area

17.1.8 Differential Pressure (DP) Flowmeters

Differential pressure flowmeters operate by placing a constriction in the flowstream that creates a difference in pressure upstream and downstream from the constriction. The constriction is called a primary element. DP flowmeters use the difference in pressure to generate a flow measurement. There are many types of primary elements, including the following

- Orifice plates
- Venturi tubes
- Flow nozzles
- Pitot tubes
- Other

DP flowmeters differ from other flowmeters in that the primary elements used with DP transmitters to measure flow are often sold separately from the DP transmitter itself. Some companies are in the business of manufacturing primary elements. They may leave it up to the customer to specify whose DP transmitter they prefer to use with the primary element.

An *orifice plate* is a round plate, usually made of steel, with a hole in it. The hole may be located at various places on the plate; not all are located in the center. Orifice plates are relatively inexpensive, but are generally good at handling dirty fluids. One chief disadvantage of orifice plates is that they create significant pressure loss because the part of the plate that does not have a hole, or “orifice,” in it blocks the flow. They are also subject to wear over time, which can degrade the accuracy of the flow measurement.

Primary elements include the following devices

A *Venturi tube* is an elongated tube that narrows at one end. Venturis can be used with clean liquids, gases, and steam, although they also can handle slurries and dirty liquids. They cause less pressure loss than orifice plates because of their relatively wide opening. Venturis are the best suited for high-speed flows. They tend to be relatively expensive, and can also be difficult to install.

A *Pitot tube* is a vertical flow element that has pressure ports both in the front and also in the back of the tube. Pitot tubes are designed for use with clean liquids, gases, and steam. They are subject to clogging when used with dirty fluids or fluids with particles. Pitot tubes cause relatively little pressure drop. Averaging Pitot tubes are quite common; these measure the flow at multiple ports and average the result, giving a more accurate measurement than single port Pitot tubes.

A *flow nozzle* is a shortened tube that curves inward to form a narrower opening than the internal diameter of the pipe. Flow nozzles can handle both clean and dirty liquids, as well as clean gases and saturated steam. Like Venturi tubes, they work best with high-speed flows, and they also do well in high temperature applications. Flow nozzles have less pressure loss than orifice plates.

Other primary elements include wedge elements, laminar flow elements, Dall tubes, and proprietary flow elements such as the V-Cone. Each type has its own advantages and disadvantages. Laminar flow elements are often used to measure low flows. The V-Cone is designed to minimize the need for upstream and downstream piping. It used to be manufactured exclusively by McCrometer, but now other companies such as Cameron have begun manufacturing their own version of the V-Cone.

Some companies have brought out products that integrate a DP flow transmitter with a primary element. Examples include Emerson Process Management's ProBar and ProPlate. The ProBar integrates a DP transmitter with an Annubar primary element, which is an averaging Pitot tube. The ProPlate integrates a DP transmitter with an orifice plate. One advantage of these products is that they can be factory calibrated with the DP transmitter already attached to and working with the primary element.

The roots of DP flow measurement go back several centuries. The groundwork was laid by Benedetto Castelli and Evangelista Toricelli with their idea that flowrate equals velocity times the area of the pipe, and with their work on pressure. Toricelli is best known for developing the barometer. In 1738, Daniel Bernoulli developed the hydraulic equation for flowrate that forms the basis of DP flow measurement. This has come to be known as Bernoulli's theorem, and it still forms the basis of DP flow measurement today.

DP flowmeters were the first type of meter to receive approval from the AGA for use in custody transfer applications. The first AGA report was published in 1930, and it was followed by a second report in 1935. These reports were updated in 1955 and issued as AGA-3. AGA-3 was accepted by the API as a standard in 1975. It was adopted as an ANSI/API standard in 1977 and updated in 1992 and 2000. For many years, these approvals gave DP flowmeters a significant edge over other flowmeters, including turbine and ultrasonic, in custody transfer applications.

17.1.9 Positive Displacement Flowmeters

The first positive displacement flowmeter was invented by Thomas Glover in 1843. Glover had some problems with liquid-sealed drum meters, which date back to the early 1800s. As a result, he created a positive displacement flowmeter with sheet metal enclosures and sheepskin diaphragms. Today positive displacement flowmeters have cloth or synthetic rubber diaphragms and are made from cast aluminum.

Positive displacement flowmeters work by capturing the flow in compartments of known quantity, emptying those compartments, and then counting how many times this is done. This need for counting explains the register that appears on positive displacement flowmeters; this register serves as a counter for determining how often the compartments

are filled and emptied. Positive displacement flowmeters are used to measure the flow of liquids and gases, but they cannot be used to measure steam.

There are a number of different types of positive displacement flowmeters that are classified by their design. The main types are as follows

- *Diaphragm*: a diaphragm to capture the flow.
- *Rotary*: one or more rotors to trap the flow.
- *Oval Gear*: two oval gears or rotors mounted inside a cylinder.
- *Helical Gear*: a rotor resembling the shape of a helix.
- *Nutating Disc*: a round disc located inside a cylindrical chamber; as the fluid enters the chamber, the disc wobbles or “nutates.”
- *Oscillating Piston*: a piston that rotates inside a cylindrical chamber.

Positive displacement flowmeters are important for measuring the following types of fluids:

- Water
- Oil
- Industrial liquids
- Gas

Positive displacement meters are used to measure commercial and industrial water, but they are also widely used to measure flow in residential applications. Positive displacement flowmeters are used to measure viscous fluids including fuel oil, hydraulic fluids, and lubricating oils. They are used for in-plant measurement of industrial liquids. And diaphragm meters in particular are widely used for billing applications involving gas used as fuel for businesses and industrial plants.

17.1.10 Turbine Flowmeters

The first inventor of the turbine flowmeter is generally believed to be Richard Woltman, who created this flowmeter in 1790. Woltman was a German engineer who wrote several books about hydraulic engineering. His life was dedicated to the Department of Ports and Navigable Waters of Hanover, Germany.

The word “turbine” comes from a Latin word meaning “spinning thing.” Turbine meters have a rotor with propeller-like blades that spins as fluid passes over it. The rotor has bearings and is mounted in a housing. The rotor spins in proportion to flowrate. Different methods are used to detect the speed of the rotor, including an electronic sensor and a mechanical shaft.

There are at least seven different types of turbine meters that vary according to the design of the spinning rotor:

- *Axial*: a rotor that revolves around the axis of flow.
- *Jet*: mainly used for water measurement, jet meters have an orifice that water is forced through, forming a “jet”; the two types are single jet and multi-jet.
- *Paddlewheel*: a lightweight paddlewheel spins in proportion to flowrate.

- *Pelton Wheel*: like paddlewheel meters but with a single size rotor with straight blades.
- *Propeller*: mainly used for dirty liquids, with helical-shaped blades that are longer than the blades of most other turbine meters.
- *Woltman*: also called “bulk” meters, are water meters for large volume applications. They have a gear train to convert the motion of the rotor into the rotation of a vertical shaft.
- *Compound*: because they incorporate 2 meter technologies; typically they have turbine technology for high flowrates and positive displacement technology for low flowrates.

Similar to positive displacement meters, turbine meters are used for water, oil, gas, and industrial liquid applications.

17.1.11 Open-Channel Flowmeters

Most flow measurement occurs in closed pipes. However, there is one type of flowmeter that is designed to measure flow in rivers, streams, and open conduits. Called an open-channel flowmeter, these meters are used for agriculture, irrigation, and wherever flow has to be measured outside of a closed pipe. Another way to draw the distinction is between flow that occurs under pressure and gravitational flow. Flow in closed pipes typically occurs under pressure, whereas flow in rivers, streams, and open conduits typically relies on the force of gravity to generate the flow.

Weirs and Flumes: There are several different methods of open-channel measurement. One common method involves the use of weirs and flumes. Weirs and flumes are hydraulic structures placed in an open channel that have a known depth-to-flow relationship. Water depth is then measured upstream of the weir or flume, and flow is determined in this way.

A weir is like a dam placed across an open channel. It is placed in the open channel in such a way that the water can flow over it. An equation is associated with each weir to determine flowrate based on the depth of the flow that is measured upstream of the weir.

A flume has an area or slope that is different from that of the open channel. It is a specially shaped portion of the open channel. Liquid depth is measured at different specific points in the flume. An equation is associated with each flume that takes flume size into account. This equation is used in calculating flowrate.

Area Velocity Method: Another common method for measuring flow in open channels does not involve the use of weirs or flumes. Instead, flow is measured by calculating the mean velocity of the flow at a cross-section. This value is then multiplied by the flow area. Normally, this method requires making two measurements: one to measure velocity and another to determine depth of flow. Flowrate Q is computed by multiplying velocity with area:

$$Q = V \times A$$

The area velocity method is used when using a weir or flume is impractical, and for temporary flow measurements. Examples where this method is used include sewer flow monitoring and influx and infiltration studies.

In addition to being used to measure flow in rivers and streams, open-channel flow is also used to measure flow in large water and wastewater pipes that are part of a water or

sewer system. Even though the flow may be occurring in pipes, the pipes are large and are open at either end. In open-channel flow, the pipes are often only partially filled. The liquid flowing through the pipes is not flowing under pressure, and this is classified as open-channel flow.

17.1.12 Variable Area Flowmeters

Variable area flowmeters play a unique role in the world of flow. Most of them are read manually, and they are sometimes used to indicate a flow/no-flow situation. Because they are manually read, their accuracy is typically low. Flowrate is read by comparing the position of a float in a tube in the meter to a dial on the outside of the tube that indicates flowrate.

The first variable area flowmeter was invented by Karl Kueppers in Aachen, Germany in 1908. Felix Meyer founded the company Deutsche Rotawerke GmbH in Aachen the following year. This company was the predecessor of the company known today as Rota Yokogawa. In 1921, KROHNE started manufacturing variable area flowmeters in Duisburg, Germany, where they are still manufactured today.

Most variable area flowmeters consist of a tapered tube containing a float. Typically, these meters are mounted vertically, and the fluid's upward force is counterbalanced by the force of gravity. At some point, the float remains constant and the flowrate can usually be read from a scale on the meter tube. The variable area meter tubes consist of glass, metal, and plastic. Plastic tubes are the lowest in cost, whereas metal tubes are used to handle high pressure applications. Some meters called purgemeters have been developed to handle low flow applications.

Although variable area meters have traditionally been manually read, some are now being produced with a transmitter output. Some even have the HART protocol available. This makes it possible to do control and recording with variable area meters, and turns the meter into something more than a visual indicator. Despite this added feature, the use of variable area flowmeters is still limited by their relatively low accuracy. Variable area meters can be used to measure liquids and gases, but they are not used to measure steam flow.

17.2 TRENDS IN FLOW MEASUREMENT

One important trend in flow measurement is the move from traditional technology to new-technology flowmeters. Many end-users are selecting new-technology flowmeters for their accuracy, reliability, lack of pressure drop, and reduced need for maintenance. Positive displacement and turbine flowmeters in particular have moving parts that are subject to wear. This wear can degrade measurement accuracy over time. Orifice plates, which are used with DP transmitters to measure flow, are subject to wear. They also cause substantial pressure loss in the line, because all the flow is forced to pass through one or more "orifices" or holes in the plate that are substantially smaller than the line size.

In custody transfer of natural gas, DP and turbine flowmeters are losing ground to ultrasonic meters. In 1998, ultrasonic flowmeters gained the approval of the AGA for use in custody transfer of natural gas. DP and turbine meters had gained this approval much earlier. The first approval of DP flowmeters for use in gas flow measurement came in 1930, whereas turbine meters first received AGA approval in 1981. So these meters had a substantial head start on ultrasonic meters. Despite this, multipath ultrasonic meters have been gaining in popularity for custody transfer of natural gas because they have high

accuracy, cause virtually no pressure drop, do not have moving parts, and are highly reliable over time.

In the area of petroleum liquid distribution, positive displacement flowmeters are losing ground to Coriolis meters. Petroleum liquids are often delivered on trucks to businesses, and also to trains, ships, and airplanes. This is typically a custody transfer measurement that has been done by positive displacement meters. Coriolis meters are taking their place in many applications because of their high accuracy and reliability. Also, the price of crude oil has risen substantially in the past year, and companies are willing to pay more to measure it. Although the initial purchase price of Coriolis meters is higher than that of positive displacement meters, their total cost of ownership is often less because they require little maintenance. For these reasons, many end-users are choosing Coriolis meters over positive displacement meters for delivery of petroleum liquids.

Traditional technology flowmeters are holding their own in some applications. Diaphragm and rotary positive displacement flowmeters are holding their own in utility and billing applications for the use of gas in commercial and industrial buildings. Turbine meters are still widely used as reference meters for calibration facilities and laboratories, due to their high accuracy. Both positive displacement and turbine meters have a hold on the residential water meter market, although magnetic flowmeters are beginning to be used for this purpose. And variable area meters are still used when a low-cost meter is wanted for a flow/no-flow indication.

The primary attributes that end-users look for in flowmeters include accuracy, reliability, and lack of required maintenance. This trend primarily benefits ultrasonic, Coriolis, and magnetic flowmeters. The use of vortex meters is also growing, especially for steam applications. And thermal flowmeters are not so much displacing other meters as they are being increasingly used for a wide range of new environmental applications involving the measurement of greenhouse gases. For all these reasons, the new-technology flowmeter market is growing at a faster rate than that of the traditional technology flowmeter market. And even though the traditional technology flowmeter market as a whole is still showing modest growth, the markets for positive displacement and turbine flowmeters appear to be declining slowly.

FURTHER READINGS

- Miller RW. *Flow Measurement Engineering Handbook*. 3rd ed. New York: McGraw Hill; 1996.
- Spitzer DW. *Industrial Flow Measurement*. Research Triangle Park (NC): Instrument Society of America; 1990.
- Upp EL. *Fluid Flow Measurement*. Houston: Daniel Industries; 1993.
- Yoder J. *The World Market for Flowmeters*. 4th ed. Wakefield (MA): Flow Research, Inc.; 2012.
- Yoder J. *The World Market for Natural Gas and Gas Flow Measurement*. A series of six studies including the core study plus modules A, B, C, D, and E. Wakefield (MA): Flow Research, Inc.; 2011.

18

HEAT FLUX MEASUREMENT

THOMAS E. DILLER

- 18.1 Introduction
- 18.2 Important issues
 - 18.2.1 Heat sink
 - 18.2.2 Potential disruption of surface-mounted gages
 - 18.2.3 Potential disruption of insert gages
- 18.3 Gages based on spatial temperature difference
 - 18.3.1 One-dimensional planar gages
 - 18.3.2 Insert heat flux gages
 - 18.3.3 Radiometers
- 18.4 Gages based on temperature change with time
 - 18.4.1 Thin-film methods
 - 18.4.2 Transient optical methods
 - 18.4.3 Coaxial thermocouple
 - 18.4.4 Null-point calorimeter
 - 18.4.5 Slug calorimeter
 - 18.4.6 Differential flame thermometer
- 18.5 Gages based on active heating methods
 - 18.5.1 Constant heat flux
 - 18.5.2 Constant surface temperature
- 18.6 Calibration and errors
 - 18.6.1 Heat flux gage calibration
 - 18.6.2 Error estimates

References

18.1 INTRODUCTION

As the cost of energy increases, the importance of predicting and controlling its movement is of increasing concern. For many years, the calculation of heat transfer has been considered a fundamental part of the engineering design of any thermal system. Whether for building environments, power production, or manufacturing processes, temperature control is often of paramount concern for physical comfort and the production of industrial materials and products. To effectively control temperature requires that energy be transferred in a known, and controlled manner. It is understood that good temperature measurements are essential and many companies offer extensive lines of temperature sensors for almost any environment. But it is equally important to know the details of the energy transfer itself.

Temperature is a fundamental property of a material and accurate measurement and calibration standards are readily available. The measurement of heat transfer, however, is a very different situation. First, it is much more challenging because heat transfer is the movement of thermal energy through a material and is not a property. Consequently, establishing measurement standards and calibration methods is much more challenging. Second, the whole concept of heat flux or thermal energy transfer per area is not as well understood by the general population. Even many engineers do not have a good physical understanding of a watt per square meter or a British thermal unit per square foot hour. Heat flux gage sales are a miniscule fraction of the thermal measurement market. Hopefully this article will encourage researchers and practitioners to use this important technology. Part of an engineer's education should be the experience of using heat flux gages, and measuring heat transfer.

Previous reviews have covered the history of heat flux instrumentation. The most comprehensive was done in 1993 (Diller, 1993). Combustion applications were covered by Arai et al. (1996), focusing on radiation heat transfer measurement. An emphasis on electronic cooling was given by Keltner (1997). Heat transfer measurements in buildings was presented by Flanders (1994). Childs et al. (1999) covered the broad range of sensors used for convection measurements, while Han et al. (2001) focused more specifically on gas turbine applications. A compilation of heat flux gage manufacturers with details of the available gages is also available (Diller, 1999).

The focus of the present review is the most useful current heat flux instrumentation, particularly heat flux gages. Although some heat flux gages are available commercially and may even be standard stock items, others are currently limited to research laboratories. The latter may be the commercial gages of the future. Optical methods are discussed briefly, but these are generally research methods that require sophisticated equipment and data processing techniques. General principles for the proper use of heat flux gages are discussed first, followed by the details of specific gages along with some of their typical applications. Three classifications of gages are considered based on measured temperature difference over space, the temperature change with time, and the power dissipated at a maintained temperature. In all of these cases it is important to understand the pathways of heat in and around the sensor, which means that how gages are mounted on or in materials can consequently be crucial. The measurement methods are also categorized according to the type of temperature measurement. The three common techniques are thermocouples (either individually or differentially), resistance temperature devices (RTDs), and optical (liquid crystals, infra-red, or thermographic phosphors). Calibration of heat flux gages is then discussed along with error analysis. Obtaining a signal from a heat flux gage is easy; properly interpreting the measurements is the challenge.

18.2 IMPORTANT ISSUES

There are a number of styles of gages commercially available, made by several companies in the United States and Europe. While many of these gages are easy to install and read the output, it is often crucial that a gage with the correct style and range is selected to give meaningful results for a specific application. For example, it is desirable to use different types of gages to measure the low heat fluxes typical of buildings and natural convection, the moderate heat fluxes of radiation from fires and forced convection in fluids, and the high heat fluxes of high speed combustion, and hypersonic flow.

As explained in this section, there are a number of problems and complications that should be taken into account while planning for any heat flux measurements. The most important issue is maintaining the proper heat flow and corresponding temperature distributions while making the measurements. The presence of a heat flux gage on or in a surface will necessarily alter the temperature field and heat flux to some extent. For considering this effect the gages can be categorized as either surface mounted or insert. The surface-mounted gages are flat and are attached to the surface with some type of glue or paste. The insert gages are usually cylindrical and are pressed into a hole flush with the surface. An additional issue, which is sometimes neglected, is where the heat goes, or the heat sink. This will be considered first, followed by the potential heat flow disruption. If these issues are not properly considered, even the most accurate heat flux gage will give erroneous results.

18.2.1 Heat Sink

For the transfer of thermal energy to occur there must be a heat source and a heat sink. Heat flux gages are designed to measure the net heat transfer between the heat source and the heat sink, usually normal to a surface between them. For example, if a heat flux gage is mounted to a surface that does not provide a good heat source or sink, it will usually measure no heat flux at steady-state conditions, irregardless of the surrounding temperatures.

Figure 18.1 illustrates the usual condition at the surface of a material. Conduction occurs through the material to the heat sink or source with convection and/or radiation exchange at the surface. If the material is low thermal conductivity, the heat flux at steady-state conditions will be very small. Therefore, high conductivity metals are usually desired for the material. Also, it is usually difficult to arrange conditions at the surface

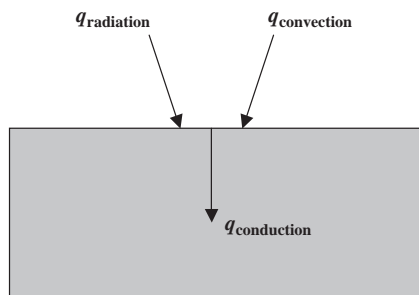


FIGURE 18.1 Surface heat balance.

which provide only convection or radiation. There is usually always some of both at least to some extent. The surface heat balance is then expressed as

$$q_{\text{conduction}} = q_{\text{convection}} + q_{\text{radiation}} \quad (18.1)$$

Alternatively, another material may be placed against the surface to provide conduction in place of convection and radiation.

18.2.2 Potential Disruption of Surface-Mounted Gages

The measured heat flux at the surface of the material on which the gage is mounted is assumed to be one-dimensional. Because of the thermal disruption caused by the placement of the gage on the surface, this may not be true, however. Wesley (1979) and Baba et al. (1985) have analyzed the thermal effects of a flat gage mounted on the surface of the material. They determined that the one-dimensional assumption is valid when

$$\left(\frac{k}{k_s}\right)\left(\frac{\delta}{R}\right) \gg 1 \quad (18.2)$$

where k_s is the thermal conductivity of the substrate, R is the radius of the gage, and δ and k are the effective thickness and thermal conductivity of the gage and adhesive. In addition, the gage should be mounted on the surface with a uniform layer of adhesive or thermal grease to minimize thermal contact resistance and keep the conditions one dimensional.

Measurements of convective heat flux are particularly sensitive to disturbances of the temperature of the surface. Convection is often expressed in terms of a heat transfer coefficient (h) and the fluid and surface temperatures.

$$q''_{\text{convection}} = h(T_{\text{fluid}} - T_{\text{surface}}) \quad (18.3)$$

Because the heat transfer coefficient in a boundary layer is increased locally by any surface temperature change, the effect of a small temperature change with location is further amplified at the gage location, as explained by Diller (1993). Therefore, any surface temperature disruptions caused by the gage should be kept much smaller than the surface to environment temperature difference causing the heat flux. For gages with one-dimensional heat transfer according to Equation (18.2), the corresponding criterion for surface disruption of temperature (5% of the fluid-to-surface temperature difference) is

$$h \frac{\delta}{k} \leq 0.05 \quad (18.4)$$

where h is the heat transfer coefficient between the fluid and surface. The important parameter is the ratio of the gage thickness (δ) to the gage thermal conductivity (k). As this ratio increases, the sensitivity of the gage increases, but the disruption caused by the gage also increases according to Equation (18.4). An optimum value for the particular application must be determined. If the gage material uniformly covers the entire surface, however, the value of $h\delta/k$ can be up to an order of magnitude larger because there are no local disruptions of the surface temperature. Independent measures of the substrate surface temperature and the gage surface temperature are advantageous in defining a value

of the heat transfer coefficient and insuring that the sensor does indeed provide a small thermal disruption.

Measurements of radiation heat transfer *from* the surface have similar sensitivities to surface temperature disturbances as convection. This is not true for radiation *to* the surface, however. This asymmetry of disturbance effects is due to the radiation heat flux being proportional to the fourth power of the absolute temperature. Consequently, a high-temperature heat source dominates the transfer exchange over the low-temperature heat sink. The surface coating properties of the gage (absorptivity, transmissivity, reflectivity, and emissivity), however, should be matched with those of the surrounding surface for either case with radiation present. If the coating is nearly black, a match of emissivity is usually sufficient.

18.2.3 Potential Disruption of Insert Gages

Insert heat flux gages appear to cause less disruption of the surface because they can be mounted flush with the surface. This does not insure a good temperature match with the surrounding surface, however. The local temperature and heat flux in the vicinity of the sensor is disturbed whenever the thermal properties, and heat sink of the gage are different than those of the surrounding surface. Gifford et al. (2010) have modeled the two effects as

$$\frac{q_g''}{q_p''} = \frac{1}{\frac{h_p}{h_g} + h_p R''} \quad (18.5)$$

where the heat flux measured by the gage (q_g'') is compared to that of the surrounding undisturbed surface (q_p''). The thermal resistance of the gage is represented by R'' , which causes a lower apparent heat flux from the gage. The resulting temperature disruption decreases the apparent heat transfer coefficient over the gage h_g relative to the surrounding undisturbed surface h_p . The effect has been demonstrated for laminar flow (Diller, 1993) and calculated for turbulent boundary layer flow (Kandular and Haddad, 2007). Moffat et al. (2000) have shown experimentally that this effect is larger than originally anticipated for turbulent boundary layers because of the laminar sublayer. It is counter-intuitive that this effect actually becomes larger as the gage size becomes smaller. Therefore, surface temperature disruptions caused by the gage should be kept much smaller than the surface to environment temperature difference causing the heat flux.

If the gage is not water cooled, it relies on the surrounding material to conduct away the heat it receives. Therefore, it must have a good thermal path to the surface into which it is mounted. It should also have good physical contact insured by a tight fit in the hole or threads with a method to tighten the gage into the surface. For steady-state measurements the gage should have an effective thermal conductivity as high or higher than the surrounding material because all of the heat must be dissipated into the surrounding surface. For unsteady heat transfer conditions, however, the gage properties should match those of the surrounding material to insure the same thermal transients. If the gage is water-cooled, the surrounding substrate can still affect the readings whenever forced or natural convection is present (Robertson and Ohlemiller, 1995). In either case, independent measures of the substrate and the gage surface temperatures are advantageous in defining a value of the heat transfer coefficient and for insuring that the gage does indeed match the surface temperature.

18.3 GAGES BASED ON SPATIAL TEMPERATURE DIFFERENCE

These heat flux gages operate by measuring a temperature difference over a spatial distance. There are several different geometries and temperature measurement methods. Because these gages measure continuously, the heat flux through the gage can be measured as long as the signal is monitored. They can be attached on the surface or inserted into the material. The different styles are discussed in separate subsections.

18.3.1 One-Dimensional Planar Gages

The simplest concept for heat flux measurement is the layered gage as illustrated in Figure 18.2. The temperature is measured on either side of a thermal resistance layer in the direction normal to the surface. The two most common methods of measuring the temperature difference are resistance temperature devices (RTD's) and thermocouples. A new ASTM standard is devoted to these heat flux gages (ASTM E2684–09, 2009).

The voltage output of the heat flux gage illustrated in Figure 18.2 is not only a function of the temperature difference, but also the thickness and thermal conductivity of the thermal resistance material, δ and k . At steady state the one-dimensional conduction equation reduces to

$$q'' = \frac{k}{\delta} (T_1 - T_2) \quad (18.6)$$

The transient response of the gage is a function of the thermal resistance layer thickness and the thermal diffusivity of the material. Hager (1965) has analyzed the one-dimensional transient response when mounted on a perfect heat sink and gives the time required for 98% response as

$$t = \frac{1.5 \delta^2}{\alpha} \quad (18.7)$$

where α is the gage thermal diffusivity. It should be noted that the sensitivity increases linearly with the thermal resistance layer thickness, but the response time increases as the square of the thickness. Consequently, sensitivity versus time response is one of the major tradeoffs in design of these gages.

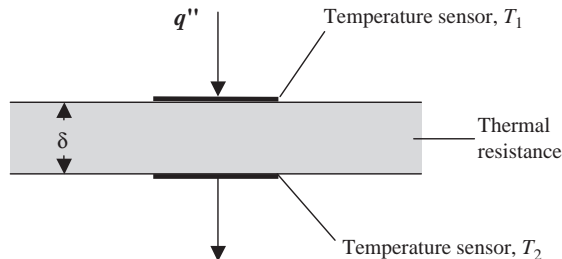


FIGURE 18.2 Layered heat flux gage.

18.3.1.1 Surface Mounted Gages Using RTD Sensors Any material that is an electrical conductor and changes electrical resistance with temperature can in theory be used as an RTD (resistance–temperature device). The temperature sensitivity for most metals is nearly constant over a wide temperature range, but is small (a fraction of a percent change in resistance per degrees Celsius). The advantage of an RTD is that it usually has a higher precision in measuring temperature than a thermocouple and it gives absolute temperature. It must be supplied with a small constant current, however, for measurements. Another disadvantage is that its resistance is also sensitive to strain and other factors, such as aging.

Epstein et al. (1986) produced a heat flux gage for turbomachinery research using a 25 μm thick sheet of polyimide (Kapton) with nickel RTD's deposited on either side. The nickel resistance element (1.0 mm by 1.3 mm) is immediately contacted to gold leads to isolate the voltage drop of the measurement at the sensor location. The leads from the bottom element are brought through the polyimide sheet so that all four leads can be taken to the edge of the sheet together. The nickel elements are either vacuum deposited with dc sputtering or electroless plated. Up to frequencies of about 20 Hz the gage responds directly to the heat flux, as indicated in Equation (18.6). For frequencies above 1 kHz the polyimide resistance layer appears infinitely thick, and the top resistance element (T_1) responds like a transient heat flux gage described in Section 18.4.1. To cover the entire range from dc to 100 kHz a numerical data reduction technique is used to reconstruct the heat flux signal. One of the advantages of these gages is that they can be wrapped onto curved surfaces, although the temperature calibrations change during this process, which may necessitate *in situ* calibration.

The heat flux gage developed by Piccini et al. (2000) is made by sputtering platinum RTD's onto one side of a 50 μm thick sheet of Upilex to measure T_1 . A thermocouple is mounted into the metal substrate onto which the sheet is glued for the second temperature measurement, T_2 . The heat flux at steady state is calculated from the temperature difference between the RTD and the thermocouple as shown in Equation (18.6). The thickness and thermal conductivity of the Upilex and the glue layer are determined from direct calibration. Analytical solutions of the unsteady heat transfer equations are used to determine the unsteady heat flux up to frequencies of 100 kHz, similar to the method used by Epstein et al. (1986).

For higher temperature applications, Talib et al. (2005) painted platinum films onto an enameled piece of 2 mm thick Inconel. This was mounted onto a surface with embedded thermocouples to create a heat flux gage of the same functionality as Piccini et al. (2000). This was used in a propane burner for fire testing. Fralick and Wrbanek (2002) have demonstrated a heat flux gage with platinum RTD's printed on both sides of a 1 mm thick slice of alumina for high temperature applications. This double-sided arrangement keeps both temperature measurements on the gage itself, which avoids the required *in situ* calibration of the previous two methods.

18.3.1.2 Surface Mounted Gages Using Thermocouple Sensors The temperature sensitivity of different combinations of materials for thermocouples is indicated by the Seebeck coefficient, S_T . Excluding the semi-conductor materials, the maximum sensitivity for any of the possible pairs is 50–100 $\mu\text{V}/^\circ\text{C}$. Kinzie (1973) lists details of many non-standard thermocouple pairs. Although a reference temperature is required for measuring absolute temperature, thermocouples generate an electrical output without the excitation current required of RTD's. For measuring temperature difference, however, a reference

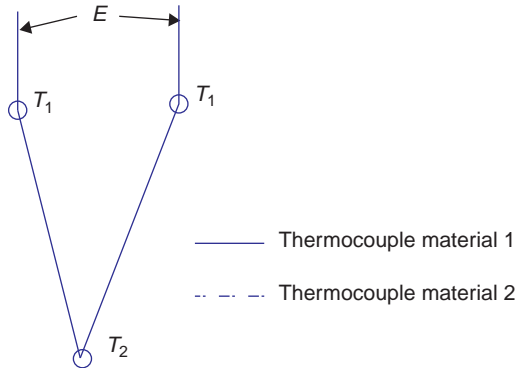


FIGURE 18.3 Thermocouple measurement of temperature difference.

temperature is not needed. This is illustrated in Figure 18.3 where the output voltage E is

$$E = S_T(T_1 - T_2) \quad (18.8)$$

If a number of these units are connected together in series, the result is a thermopile which increases the voltage output for a given temperature difference. An example of such a thermopile heat flux gage is illustrated in Figure 18.4. Here five pairs of thermocouples are shown combined in series across the thermal resistance layer. Each produces an output voltage that is proportional to the temperature difference, $T_1 - T_2$. The total output voltage is then proportional to the number of thermocouple pairs, N , and the Seebeck coefficient for the thermocouple pair, S_T .

$$E = NS_T(T_1 - T_2) \quad (18.9)$$

Using a thermopile, therefore, can easily overcome the greater sensitivity of individual RTD's. Thermocouples are also insensitive to strain and most other factors besides

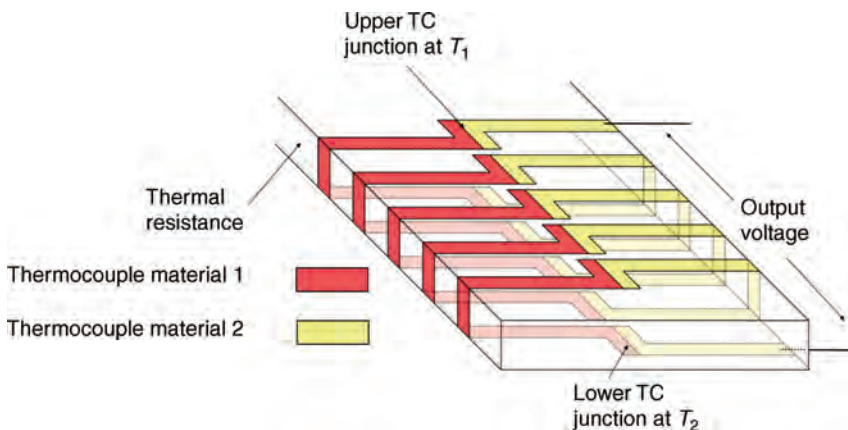


FIGURE 18.4 Thermopile heat flux gage.

temperature. Moreover, only two leads are required for a heat flux measurement, versus four for two RTD's. The corresponding heat flux sensitivity of the ideal layered gage is therefore

$$S_q = \frac{E}{q''} = \frac{NS_T\delta}{k} \quad (18.10)$$

To minimize the temperature disruption of the gage the thickness, δ , should be kept as small as possible, particularly if there is a large convection coefficient on the surface, as shown in Equation (18.4).

The values of δ used in different gages vary widely depending on the range of heat fluxes to be measured. Thermal resistance layers with thicknesses of 1 mm or more have generally been used for heat fluxes less than 1 kW/m². The time response is on the order of a second or more. Applications are typically conduction heat flux in building structures or insulation and natural convection. Bales et al. (1985) have published a book of articles discussing the design, calibration, and use of heat flux gages for building applications. One design uses welded wire to form the thermopile across a sensor about 1 mm thick with an upper temperature limit of 300°C. These are manufactured in a range of sizes by International Thermal Instrument Co. Applications include heat transfer in buildings and physiology. Sensors with higher sensitivity are made with semiconductor thermocouple materials for geothermal applications. One type of sensor is designed for operation at gage temperatures of up to 1000°C.

For heat fluxes of up to 100 kW/m² thermal resistance thicknesses of 25–100 μ m have been used. The corresponding time response is as low as 50 ms. Some improvement in the frequency response is possible with appropriate signal conditioning (Bauer and Heywood, 1997). To measure the temperature difference, $T_1 - T_2$, Ortolano and Hines (1983) used a thermopile constructed of thin metal foils built around a Mylar or Kapton sheet. It is very versatile because it easily conforms to most surface shapes. Applications have included many types of conduction, convection, and radiation. These gages (named HFS) are manufactured by the RdF Corp. and sold by several distributors. The same approach has been used by Hubble and Diller (2010) to produce an array of heat flux sensors on a Kapton sheet. Instead of connecting the differential thermocouples in series, however, they were isolated individually as single pairs that gave 10 heat flux readings on each sheet. With appropriate data processing a 95% time response of 36 ms was obtained.

Thick film technology was used by Van Dorth et al. (1983) to put over 500 thermocouple pairs on a heat flux sensor that was 15 by 30 mm in size. Platinum and platinum–gold conductors were printed over and under a thermal resistance layer of glass-like material. This gave good sensitivity for demonstrated heat fluxes up to 200 kW/m² and temperatures up to 500°C.

A smaller, but similar sensor design has been made with screen printing techniques of conductive inks (Langley et al., 1999). A copper/nickel thermocouple pair was used with a dielectric ink for the thermal resistance layer. The inks were printed onto anodized aluminum sheets for the substrate. Although the entire package is 350 μ m thick, the thermal resistance is low because of the high thermal conductivity of all of the materials. Sensitivities are sufficient to measure heat fluxes as low as 0.1 kW/m². The thermal time constant is about 1 s, and the upper temperature limit is approximately 150°C. The aluminum base provides some limited conformance to a surface.

Printed circuit techniques were previously used by Vatel Corp. to produce a heat flux gage by plating two thermocouple materials on either side of a 0.2 mm thick printed

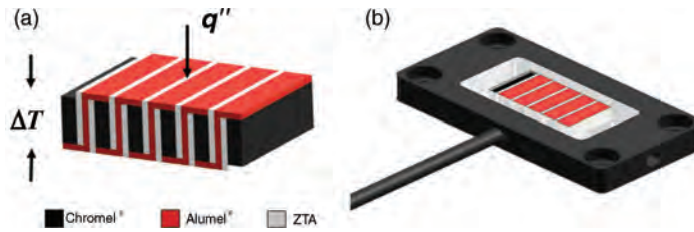


FIGURE 18.5 HTHFS, schematic of the measuring element (a) and, complete sensor (b).

circuit board. The thermocouples were connected in series by holes drilled and plated through the board forming many thermocouple pairs to give good sensitivity. The time constant for transient response was approximately one second.

An array of heat flux sensors was made with thermal vapor deposition by Ewing et al. (2010). Thermocouple pairs were formed on both sides of a $50\text{ }\mu\text{m}$ thick sheet of Kapton individually wired to give a separate heat flux and temperature output for each pair. There are a number of fabrication tricks to make a reliable sensor, including fluxless soldering of lead wires, vacuum depositing through holes in the sheet, and sealing the entire sensor to prevent oxidation with time.

A robust high temperature heat flux gage (HTHFS) was designed and fabricated by Gifford et al. (2010) by welding Chromel and Alumel strips as illustrated in Figure 18.5. The thermal resistance between the top and bottom thermocouples is formed by the elements themselves. The ceramic (ZTA) strips are included to provide electrical insulation between elements. An additional thermocouple wire welded to the top of the gage provides the surface temperature, which is useful for interpreting the heat flux signal. Long time operation and cycling to 1000°C was demonstrated. The time constant was less than one second. The measurement chip can also be mounted in an insert type housing which can be air or water cooled for gage temperature control.

Hubble and Diller (2010) developed a simple data processing method to extend the operation of the HTHFS so that it could be used on any substrate material. All of the sensors in this section are designed for the heat to go through the sensor into the material on which they are mounted as described in Section 18.2. The calorimeters described in Section 18.4 actually absorb the heat into the sensor itself, using the rate of temperature rise as the measure of heat transfer. The HTHFS can operate in both modes simultaneously—as a differential sensor and a calorimeter. Combining the heat transfer at the surface gives what is termed the “hybrid heat flux”

$$q'' = q''_{\text{differential}} + \frac{1}{2} q''_{\text{calorimeter}} \quad (18.11)$$

This gives the correct heat flux whether the gage is mounted on a good heat sink (high conductivity material) or a good insulator (low conductivity material). It also increases the time response of the gage by about an order of magnitude. The differential heat flux is proportional to the usual temperature difference measured across the gage and the calorimeter heat flux is proportional to the time rate of temperature change of the gage.

A thin-film version of a high temperature heat flux gage has been developed by MesoScribe Technologies, Inc. that can be deposited directly onto test surfaces using a

Direct Write Thermal Spray process (Theophilos et al., 2010). Up to eighty thermocouple pairs (Type K or N) are scribed around a layer of dielectric serving as the thermal resistance for the heat flux signal and a separate thermocouple gives the surface temperature. A layer of yttria-stabilized zirconia is first deposited on a metal surface to provide electrical isolation. It can also be deposited over the gage to encapsulate it for better durability. All of the layers are on the order of 0.1 mm thick, which gives a good output signal and good survivability at high gage temperatures.

An anisotropic thermoelement has been used by Mityakov et al. (2002) to produce a heat flux gage with ten to one hundred times the voltage output of the usual thermopile gages. This is attributed to the transverse Seebeck effect which is experienced in single crystal bismuth strips. The strips are wired in series and mounted on a mica base. The output voltage is generated at the surface of the material, which can be as thin as 0.1 mm.

18.3.1.3 Surface Mounted Wire-Wound Gages The wire-wound gage is similar to the thermopile layered gages except for the method of producing the thermocouple junctions around the thermal resistance layer. A fine wire of one of the thermocouple materials, for example constantan, is wrapped around the thermal resistance layer N number of turns. One-half of the wire is then electroplated with the other thermocouple material, for example copper. The result is a set of thermocouple junctions where the electroplating stops on the top and bottom of the thermal resistance layer. It is different from the thermopile gages of Section 18.3.1.2, however, because one of the wires is continuous all around the sensor. This forms a thermopile with the wire on one side and the wire and plating in series on the other side. There are N pairs of thermocouple junctions corresponding to the N windings. The theoretical output is therefore less than a true thermopile gage and is a function of the plating thickness in addition to the overall geometry and thermoelectric output of the materials.

A major use of these gages is low heat flux measurements on the surface of equipment to measure heat losses and insulation effectiveness. A summary of the theory is given by Hauser (1985) and a review of the applications by van der Graaf (1989). Thicknesses of the thermal resistance layer vary from 0.5 to 3.0 mm, with time constants from 1 to 30 s. Some of the sensors are flexible for wrapping around pipes and curved surfaces. Applications include building structures, insulation, geothermal and medicine, with heat fluxes generally less than 1 kW/m^2 . Concept Engineering offers this type of gage commercially.

18.3.2 Insert Heat Flux Gages

These heat flux gages are mounted through a hole in the material flush with the surface. The heat sink is provided either by the material in which it is mounted or by water cooling through channels in the gage. A new ASTM standard is devoted to these heat flux gages (ASTM E2683-09, 2009).

18.3.2.1 Insert Gages Using Thin-Film Thermocouple Sensors A differential thermopile as illustrated in Figure 18.4 can be deposited directly onto a substrate to give design and manufacturing flexibility. Such a thin-film device has been described in detail by Diller and Onishi (1988) and was first produced by Hager et al. (1991, 1993) using sputtering techniques. Microfabrication methods are used to deposit hundreds of thermocouple pairs around a silicon monoxide layer to create the differential thermopile.

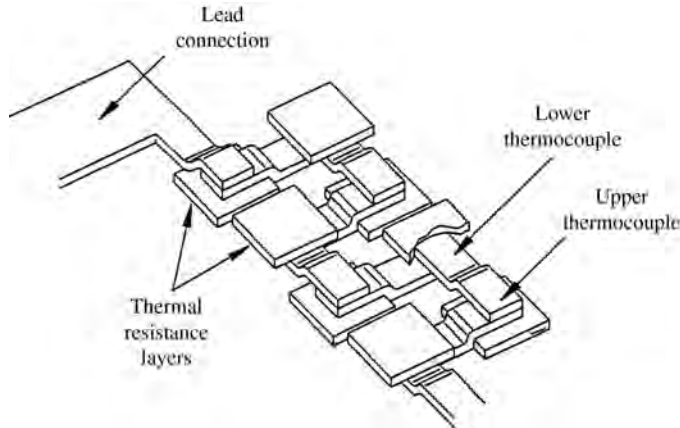


FIGURE 18.6 Heat flux microsensor pattern (Holmberg and Diller, 1995).

Either photolithography or stencil masks can be used to define the patterns. The resulting physical and thermal disruption of the surface due to the presence of the sensor is extremely small.

While the original version of this heat flux microsensor (HFM) placed the temperature sensors almost directly over top of each other across a single thermal resistance layer as illustrated in Figure 18.4, it is not a requirement. The bottom temperature sensors simply need to be at a uniform temperature and the top temperature sensors need to be at a temperature dictated by the heat flux perpendicular to the surface. This can be accomplished on a high conductivity substrate (aluminum nitride) by separate thermal resistance pads ($1\text{ }\mu\text{m}$ layer of silicon monoxide) underneath the top thermocouples. The pattern is illustrated in Figure 18.6 (Holmberg and Diller, 1995). The bottom temperature sensors can then be placed directly on the substrate with or without thermal resistance pads on top. Leads are taken down the side and attached to wires on the side or behind the sensor substrate. The substrate is then press fit into a high conductivity metal housing with an outside diameter of about 6 mm. A thin-film RTD or thermocouple is also deposited on the surface for independent temperature measurement of the sensor surface. The applicable range for heat flux measurement is from 1 kW/m^2 to 10 MW/m^2 . Because the sensor is so thin, the thermal response time is less than $10\text{ }\mu\text{s}$ (Holmberg and Diller, 1995), giving a good frequency response well above 10 kHz .

18.3.2.2 Insert Wire-Wound Gages The Schmidt–Boelter version of the wire wound gage is illustrated in Figure 18.7. The thermal resistance layer (wafer) is usually made of a high thermal conductivity material, for example anodized aluminum ($\sim 0.5\text{ mm}$ thick). The nonconductive coating is necessary to provide electrical insulation with the bare wire of the thermopile. The wire is usually constantan with copper plating on one side of the wires top and bottom. This forms a type of differential thermopile, but the segmented plating on the wire gives an output less than for a true thermopile (Kidd and Nelson, 1995). The entire wafer is placed in contact with a heat sink and surrounded by potting material to give a smooth surface to the top of the gage.

Schmidt–Boelter gages are manufactured by Medtherm Corp. in a range of sizes down to 1.5 mm diameter. They are designed to measure moderate to high heat fluxes

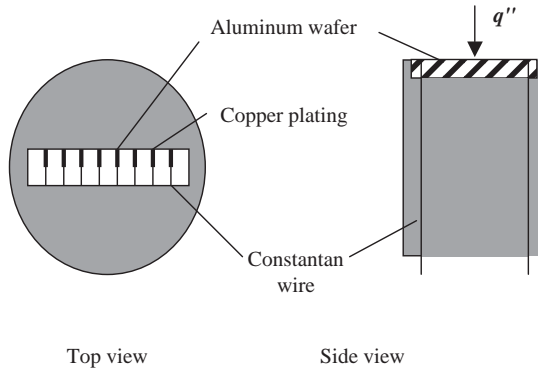


FIGURE 18.7 Schematic of an insert wire-wound heat flux gage.

(10–1000 kW/m²). One advantage of the Schmidt–Boelter gage is that it can be contoured to a curved surface. One of the drawbacks is that one-dimensional heat transfer is not really maintained. Two-dimensional effects can be significant and the potting material affects the gage sensitivity. The transient response is typically 25–50 ms, but is usually not first-order. Techniques for improving and correcting the time-response are presented by Kidd and Adams (2001) for making time-resolved measurements. A Schmidt–Boelter design having a first-order response has been patented (Hevey and Langley, 2001).

An alternative design was produced by Long et al. (2004) by winding constantan wire onto an annular disc of Macor. Copper plating of the outer radius of the wire produced a circular thermopile for use in turbomachinery research. Epoxy resin was used to cover the wires and appeared to substantially increase the thermal resistance of the gage.

18.3.2.3 Circular Foil Gage (Gardon) The circular foil gage was originated by Robert Gardon (1953) and consequently is often called a Gardon gage. A sketch of the circular foil gage is shown in Figure 18.8. Constantan is usually used for the disk material and is attached to a hollow copper body. A copper wire is attached to the center of the foil to

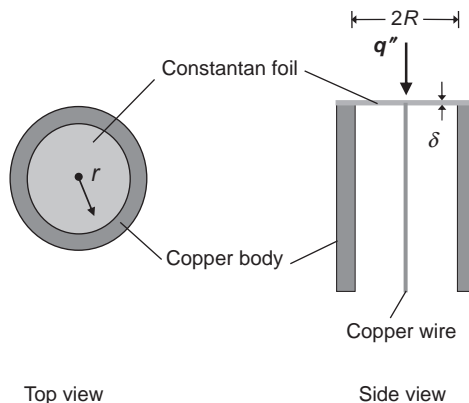


FIGURE 18.8 Schematic of a circular foil gage.

create a differential thermocouple between the center and edge of the disk. Unlike the previous gages discussed, however, the heat transfer in the gage is not in the direction that it enters the surface. The thermal energy is collected by the disk and transported laterally to the heat sink connected to the edge of the disk. The recorded temperature difference is, therefore, a function of the total heat transfer to the disk, but also is a function of the distribution of heat flux over the disk surface (Diller, 1993; Kuo and Kulkarni, 1991; Liechty et al., 2007; Agnone, 2009).

The usual application for Gardon gages is radiation to a water-cooled gage. Because there is negligible radiation from the gage, the heat flux is uniform and the temperature distribution across the foil is parabolic

$$T - T_w = \frac{q''}{4k\delta}(R^2 - r^2) \quad (18.12)$$

where R is the foil outer radius, r is the radial coordinate, and T_w is the temperature of the wall, which is assumed equal to the temperature at the outer edge of the foil. When convection is present, however, the convective heat transfer is proportional to the difference in temperature between the fluid and the gage. Because of the temperature distribution radially across the gage, the heat flux is no longer uniform. For the same total heat transfer, therefore, the gage will give less output than for radiation. If the convection heat transfer coefficient is uniform over the gage, the analytical correction of the heat flux for convection is approximately (Borell and Diller, 1987)

$$\frac{q''_{\text{corrected}}}{q''_{\text{uncorrected}}} = \frac{1}{1 - \frac{3S_q h}{4S_T}} \quad (18.13)$$

where S_q is the sensitivity of the gage as calibrated by radiation, h is the convection heat transfer coefficient, and S_T is the thermoelectric constant for the thermocouple pair. The correction is proportional to the gage sensitivity and the heat transfer coefficient. Consequently, the sensitivity of the gage should be kept low to limit the size of the correction and associated uncertainties. A general rule for specifying copper constantan gages is

$$\frac{q''_{\text{convection}}}{q''_{\text{fullscale}}} \leq \frac{T_{\text{fluid}} - T_w}{3200^\circ\text{C}} \quad (18.14)$$

where the full-scale heat flux is that corresponding to an output of 10 mV. The measurement error becomes even larger if the heat transfer coefficient is nonuniform over the gage surface, as can occur in strong shear flows. In addition, the surface temperature disruption caused by the presence of a circular foil gage affects the local thermal boundary layer and changes the local heat transfer coefficient, as discussed in Section 18.2.3.

These corrections are all based on the assumption that the gage makes perfect contact with the wall. Because the gage is normally slip fit into place, there may be an additional temperature drop between the heat sink of the gage and the wall. This would cause an additional drop in gage output and would require an additional correction of the results.

Most Gardon gages are made with a water cooling jacket as part of the sensor to keep it from overheating when in high heat flux environments. Water cooling should never be used when convection is being measured. The problem is that this causes a local

“cold spot” and results in an erroneously high heat flux measurement, the opposite of the foil temperature disruption effect just discussed.

Circular foil gages are manufactured and sold by Medtherm Corp. and Vatell Corp. Most are used for radiation measurements and are water cooled. The ASTM standard (ASTME511-07, 2009) for these gages provides a good summary of the best practice for their use.

18.3.3 Radiometers

With the proper surface coating any heat flux gage can be used to measure radiation, although specialized radiometers have been developed to increase the accuracy of radiation heat flux measurements (Murthy et al., 2003). One example is the Kendall radiometer (Kendall and Berdahl, 1970), which is a blackbody cavity with a thermopile and an installed electric heater. This allows the electric power measurement to serve as the standard to provide a self-calibrated device for total irradiance. Consequently, these instruments can be used as the primary standard for heat flux calibration of other gages. Several models are available from Medtherm Corp.

To eliminate the effects of convection a transparent window or lens can be placed over the end of a heat flux gage to transmit only the desired radiation transfer. The reduced angular view of the gage with a window must be included in the calibration, along with the transmissivity of the window. Without a window the effects of natural convection are particularly a problem at low heat flux levels (less than 15 kW/m^2) (Robertson and Ohlemiller, 1995). An additional problem in dirty environments is deposits on the window that reduce the transmissivity. One solution is to blow air past the window. A different approach without a window is the transpiration radiometer which blows air through a porous plug which composes the exposed surface of the gage. This not only blows off the fluid boundary layer with any contaminants, but the air is heated as it passes through the plug. The measured radiation is then related to the temperature change of the air as it passes through the plug (Martins et al., 2000, 2006).

Brajuskovic et al. (1991) have developed a slightly different version to operate in boilers. Instead of going through the plug, the air is directed through slots around a solid plug. The air-flow keeps particles from reaching the heat flux element and it takes away the thermal energy. The path of the heat transfer is radial through the plug, however, as in a circular foil gage. The heat flux is proportional to the temperature difference between the center and edge of the plug.

18.4 GAGES BASED ON TEMPERATURE CHANGE WITH TIME

The time–temperature history of a material can be used to infer the corresponding heat flux. This involves solving the conduction heat transfer problem of the temperature gage and the adjoining material. Because the solution starts with the measured temperature as a function of time and calculates backward to find the needed heat flux boundary condition, it is in the form of an inverse heat transfer problem.

18.4.1 Thin-Film Methods

Transient thin-film gages are designed to measure the surface temperature of a semi-infinite material (substrate) in response to the applied surface heat flux. The requisite

transient conduction solutions for the one-dimensional semi-infinite model are given in numerous references. The starting point (Schultz and Jones, 1973) is the solution of the one-dimensional heat diffusion equation for a step change in surface temperature from a uniform initial temperature

$$q'' = \frac{\sqrt{k\rho C}}{\sqrt{\pi t}}(T_w - T_0) \quad (18.15)$$

The material properties of the substrate enter as the square root of the product of $k\rho C$ (thermal conductivity, density, and specific heat). A number of data reduction schemes have been developed based on variations of this solution to determine the heat flux from the surface temperature history, represented by measured temperatures at specified times. One common method is given by

$$q''(t_n) = \frac{2\sqrt{k\rho C}}{\sqrt{\pi}} \sum_{i=1}^n \frac{T_i - T_{i-1}}{\sqrt{t_n - t_i} + \sqrt{t_n - t_{i-1}}} \quad (18.16)$$

The summation for the heat flux at each given time is based on the current and all previous temperatures. One limitation of this equation is that constant thermal properties have been assumed for the substrate. A more serious problem, however, is the inherent instability of using a temperature record to infer the heat flux. This is akin to taking the derivative of the temperature signal, which greatly amplifies any noise present in the signal and is particularly important when making detailed measurements of time-resolved heat flux. To increase sensitivity and solution stability a variety of inverse heat conduction techniques have been developed (Walker and Scott, 1998; Oldfield, 2008).

The advancement of thin-film deposition techniques has made the fabrication of large numbers of resistance elements for temperature measurement on the surface of models much simpler. All that is required is the application of a thin temperature measuring layer (usually an RTD) on the model with appropriate lead connections. The most popular methods have been to apply a thin layer of platinum paint or to directly sputter onto an electrically insulated substrate. With a typical thickness of $0.1 \mu\text{m}$ the response of the RTD's is of the order of $0.01 \mu\text{s}$. This causes negligible effect on the heat flux measurements for times greater than a few microseconds. Therefore, results that are properly processed are essentially instantaneous even for high-speed flows. There is a limit on the length of the useful test data, however, due to the semi-infinite substrate assumption used in the modeling. The upper time limit for a 1% error in the heat flux at the surface is

$$t = 0.3 \frac{\delta^2}{\alpha} \quad (18.17)$$

For a typical insulating substrate (MACOR) of $\delta = 1 \text{ mm}$ thickness the corresponding maximum time is nearly 400 ms. This makes the technique useful for many types of short-duration flows, such as shock tubes or high-speed blow-down facilities.

To measure the resistance of the RTD to determine the transient temperature history a small current is used to drive a bridge circuit. The constant current source must be large enough to provide a measurable output voltage, but small enough to keep the self-heating effects negligible. This optimization is made easier if the sensor resistance is a high value.

For a resistance of at least 100 ohms, a current of 1–2 mA provides good results. One of the advantages of vacuum deposited resistance films is that the value of resistance is easier to control during fabrication.

Calibration of the sensors for heat flux measurements requires two steps. First, the resistance versus temperature relationship must be determined for each sensor. This can simply be done by putting the model in an oven and monitoring the resistance of each gage over a set temperature range. The second part of the calibration is to determine the thermal properties of the substrate. The most popular method is to measure the response of the sensor to a pulse of energy provided by an electrical discharge through the resistance film or a pulsed radiation source to the surface. It is important to measure the value of $k\rho C$ for each model because properties can vary significantly for many of the substrate materials between batches.

A group at Oxford has developed a technique to apply the thin-film gages to metal substrates using a vitreous enamel coating to provide electrical insulation (Doorly and Oldfield, 1986). Although this gives much greater flexibility for experimental application of the method, it also complicates the data analysis. One method is to use an analog circuit plus digital processing of the signals, including filters and a frequency boosting circuit (Ainsworth et al., 1989). The calibration is also more complicated because of the multiple layers making up the substrate. A further refinement has been to place multiple RTD's on a single polyimide sheet which is then glued to the metal surface (Guo et al., 1998).

Another group has developed round Pyrex inserts that are placed in metal surfaces for measuring time-resolved heat flux in short duration turbine research test facilities. The data reduction includes both analog circuits and digital data processing, including Fourier transforms to optimize the frequency response and accuracy of the time-resolved results (Dunn et al., 1986). Concern over the thermal disruption caused by the Pyrex inserts was investigated and reported to be insignificant for the short 20–25 ms test times of the experiments (Dunn et al., 1997). Moffat et al. (2000) have investigated the more general case of longer times when the temperature rise of the insert becomes substantially more than the surrounding metal material.

18.4.2 Transient Optical Methods

A number of optical methods have been developed to measure the transient surface temperature for determination of heat flux. These include liquid crystals, infrared radiation measurement, temperature-sensitive paints and thermographic phosphors. The advantage is that cameras can be used to give full coverage of the surface, although the results are usually limited to steady-state convection. The usual method involves a flow facility where either the model is injected into the flow at a different temperature or the flow is suddenly initiated with the model at a different initial temperature.

Infrared cameras offer the most potential measurement capability. Once the surface emissivity is calibrated, the data processing is relatively easy (Ekkad et al., 2004). High speed versions are currently available that allow time-dependent measurements of heat flux from the transient temperature signals (Ahn et al., 2010). Proper surface coatings should be used. Viewing angles are important along with reflections from other surfaces.

Thermographic phosphors emit radiation in the visible spectrum when illuminated with ultraviolet light, which can be related to temperature at specific wavelengths (Khalid and Kontis, 2008). Their main advantage is that they offer the possibility to operate in

elevated temperature environments (Bizzak and Chyu, 1995). A CCD camera is required to record the optical images and the data processing and calibration are challenging (Walker, 2005).

The temperature-sensitive paint method uses a Europium complex as the luminophore with a mercury–xenon arc lamp for illumination (Kurits and Lewis, 2009). A CCD camera is used to collect the images as a function of time. Because the paint layer has a low thermal conductivity, a two-layer model to relate the measured surface temperature to the heat flux is usually required.

Liquid crystals are chiral-nematic molecules that reversibly change color as a function of temperature. They are limited to a temperature range of about 25–40°C and are available commercially from Hallcrest in a convenient micro-encapsulated form. A hue capturing technique is used with a temperature calibration to reduce the data to temperatures as a function of time (Ireland and Jones, 2000). The liquid crystal method is used almost exclusively with convection testing in air to obtain the distribution of heat transfer coefficients as defined in Equation (18.3) (Wagner et al., 2005; Das et al., 2005). Multiple step changes of the air temperature can even be used to obtain additional information, such as adiabatic wall temperatures (Talib et al., 2004).

18.4.3 Coaxial Thermocouple

Coaxial thermocouples are rugged sensors designed to be insert through a material to measure the surface temperature of the model wall as a function of time, the same as thin-film gages. Consequently, the same equations can be used to determine the corresponding heat flux. The physical concept is simple (Kidd, 1990), as illustrated in Figure 18.9. One thermocouple material forms the center wire, which is surrounded by an electrical insulator. The second thermocouple material forms a sheath around these two layers. The final assembly is often then drawn down to a smaller diameter. A second insulating material is sometimes placed around the assembly to isolate the thermocouple electrically from a metal substrate. The completed unit is press fit into the model. The actual thermocouple junction is formed at the very end by plating a thin layer of one of the materials, vacuum deposition, or by simply lightly sanding to mix the two materials together and bridge the thin insulating layer. Because the thickness of the top junction layer can be on the order of 10 μm , the initial response time can be much less than 1 ms. Smith et al. (2002) include a

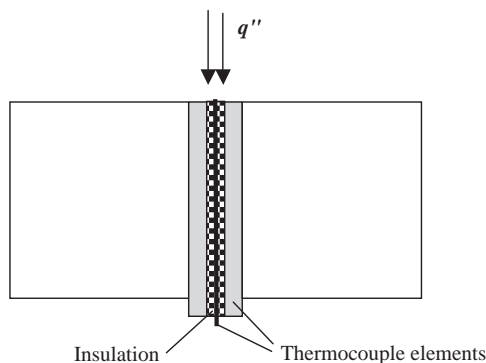


FIGURE 18.9 Schematic of a coaxial thermocouple.

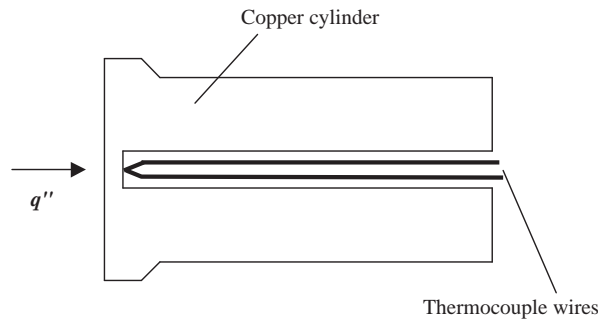


FIGURE 18.10 Null-point calorimeter schematic.

second thermocouple with the use of a parameter estimation method to reduce the temperature data to heat flux.

The most important parameter in specifying the coaxial thermocouple gages is to match the material thermal properties with those of the wall. As seen from Equation (18.16), the property of importance is the product $k\rho C$, which for the coaxial thermocouple is taken as the average of all the layers. Usually this can be matched quite well by judicious choice of the thermocouple pair and the thickness ratios.

18.4.4 Null-Point Calorimeter

In very high heat flux environments gages that measure surface temperature cannot physically survive for long. One solution is to make the gage body from high conductivity copper and move the thermocouple back away from the surface, as illustrated in Figure 18.10. The goal of the design is to produce the same temperature history at the null point (thermocouple location) as would occur on the surface of a semi-infinite slab of the model if the gage were not present. Consequently, the gage is called a null-point calorimeter. The proper placement of the thermocouple relative to the surface and the size and depth of the hole in the gage behind it are based on analytical modeling. An ASTM Standard (ASTM E598-08, 2009) gives the design details and guidelines on its proper use. The typical value for the useful measurement time for a copper body is between 1 and 300 ms. The usual thermocouple pair is chromel/alumel. Its attachment to the gage is one of the major fabrication difficulties. To accomplish a test within the recommended test time, the model is often swept through the flow. This also minimizes the time that the model and gages are exposed in high heat flux environments and extends their useful life. Because the null-point temperature on the back of the gage is designed to match the undisturbed wall temperature, the front side of the gage is necessarily hotter. Although this creates a “hot spot” effect, these gages are often used at the stagnation point of models where the effect is small (Kidd, 1992).

18.4.5 Slug Calorimeter

A calorimeter is a device used for measuring the quantity of absorbed thermal energy as a function of time. In all of the other sensors in this section the temperature distribution as a function of time is used for a specific solid shape and size. A slug calorimeter uses one

temperature measurement on the back surface to represent the entire mass of the slug, which is assumed to change temperature at a uniform rate. If a control volume is taken around the slug, application of energy conservation yields

$$q = mC \frac{dT}{dt} + q_{\text{losses}} \quad (18.18)$$

where q is the heat transfer at the gage surface and mC is the thermal mass of the slug. The losses are minimized by insulating all of the surfaces other than the front surface. If the losses can be neglected and the heat flux is constant with time, the gage temperature increases linearly with time.

$$T = T_0 + \frac{q''A}{mC} t \quad (18.19)$$

The result is a straight line on a temperature versus time plot, with the slope of the line giving the desired heat flux. The ASTM Standard (ASTM E547-08, 2009) also recommends calculating the slope for the cooling process after the heat source is removed as an indication of the losses during heating. To neglect losses, the rate of cooling should be less than 5% of the rate of heating. A method for accounting for losses, however, has been demonstrated (Hightower and Olivares, 2008), which allows for accurate measurements when larger losses are present.

There are several inherent problems in using slug calorimeters. Because the thermocouple is mounted on the back of the slug, the temperature measurement is not the average temperature of the slug. The material of the gage should be the same as the wall to minimize nonuniform temperature effects. The heat losses are usually hard to control in models with high heat flux conditions. Although slug calorimeters are rarely used currently, they can provide an important standard in the calibration process.

18.4.6 Differential Flame Thermometer

The differential flame thermometer (Keltner et al., 2010) consists of two inconel plates separated by a layer of insulation with a thermocouple mounted on each plate. The plates are 120 mm square and the insulation is 19 mm thick. The resulting time constant for the system is about 90 s. The application is in a radiation dominated furnace or combustor to measure the heat flux and steady-state temperature. Inverse heat transfer methods are used to convert the two temperature records into heat flux. It should be noted that this device is different from a standard heat flux gage because it is suspended in the furnace or environment and does not have a heat sink.

18.5 GAGES BASED ON ACTIVE HEATING METHODS

Because electric power dissipation can be easily and accurately measured, it is attractive to incorporate electric resistance heating into the measurement of heat transfer. At steady-state conditions the electric power supplied to the system must equal the heat transfer from the surface plus the heat losses. One drawback is that the transfer is always *from* the surface. Consequently, this method is most useful in well-controlled laboratory

experiments. Because of the time constraints and power limitations, it is not used in high heat flux or high temperature situations. This eliminates most high-speed flows. In addition, it is difficult to maintain truly steady-state conditions. Because the thermal capacitance of the surface material is usually large, very small time rates of temperature change may cause large errors in the measured heat transfer and it may take a long time to achieve steady-state conditions under which the temperature is no longer changing.

The techniques used can be grouped according to the two boundary conditions that normally result, constant heat flux, and constant wall temperature. In constant heat flux cases the experiment usually runs until a steady-state temperature distribution is achieved. For constant wall temperature manual or active control of the power is required to stabilize the system at the prescribed constant temperature.

18.5.1 Constant Heat Flux

Measurements of convective heat transfer coefficients with electrically heated constant heat flux surfaces have been done for many years. The most common method is to use a heater with uniform resistance, insulate the back surface, and minimize lateral conduction by the design. A thin metal sheet is often used as the heater to provide uniformity of resistance (and heat flux) and to minimize the lateral conduction. Several centimeters of good insulation as the backing material usually insures acceptable heat loss from the back surface. Unfortunately, the response time for this arrangement is typically very long, depending on the surface convection and substrate properties. With a surface heat transfer coefficient of $100 \text{ W/m}^2 \text{ K}$ and a good substrate insulator the time from a cold start to reach steady-state conditions (within 2%) is typically over 1 h. The time limitation is the result of transient heat transfer into the insulation backing layer.

Surface temperature measurements are not needed to determine the heat flux, but they are required to quantify the conditions under which the heat flux occurs. Temperature is particularly necessary for the determination of the heat transfer coefficient. Although any of the methods for measuring temperature are possible candidates, only thermocouples, resistance temperature devices, infrared cameras, and liquid crystals are commonly used. Their use is discussed in that order.

When thermocouples are used to measure the surface temperature their attachment and layout is important because of the potential problems caused by heat transfer through the wires. Because the wires have a much larger thermal conductivity than the insulation, the measured temperature can be significantly lowered. Running the wires parallel to the surface for a distance and using small diameter wire can minimize the effect.

If the temperature is measured with an RTD, it may also be used for supplying the electrical power. A thin-film resistance element sputtered onto quartz was used by Samant et al. (1984) for boiling studies of refrigerant 113. Because the surface of the heated film was only 0.25 mm by 2.0 mm, significant heat was transferred laterally through the quartz substrate, giving the heater a larger effective surface area by an estimated 12–15%. Normally the edge effect is much larger, but the high heat fluxes in boiling kept it to a minimum.

Infra-red optical techniques have been used extensively with large heater foils to measure the temperature distributions on constant heat flux surfaces to infer the convective heat transfer coefficients (Buchlin, 2010). Images are taken with an infrared camera, which is calibrated for the surface emissivity using surface thermocouples (Colban et al., 2006).

Praisner and Smith (2006) used liquid crystals to record the instantaneous heat transfer distributions for water flow over a surface. A stainless steel foil was painted black and heated electrically. The liquid crystals were applied to the underside of the test plate where the temperature distributions were read with a thermal imaging system. The frequency response of the system for determining the surface heat transfer coefficients was greater than 100 Hz. Heating can also be supplied by an infrared source on the opposite side from the liquid crystals (Ochoa et al., 2005).

18.5.2 Constant Surface Temperature

The design and operation of constant surface temperature experiments are substantially different than those for constant heat flux. Instead of measuring the temperature that results from an imposed heat flux, the electrical power is measured that is required to maintain a set temperature. This typically requires a control system to maintain the temperature constant spatially and/or temporally, depending if steady or unsteady measurements are desired. Although electric resistance heating is almost exclusively used, providing heat transfer *from* the surface, a device capable of providing a measured heat transfer in either direction has been constructed using Peltier devices (Shewen et al., 1989).

Many time-averaged measurements of spatial distributions of convection heat transfer have been made using segmented plates and heaters. If the plates are made of a thick, high-conductivity material, the surface will be nearly isothermal, even if the heat transfer coefficient has large spatial variations. Some thermal isolation is required between segments, often provided by insulation strips. Usually the individual heaters are adjusted manually to achieve a uniform temperature, although a good set of PID automatic temperature controllers can speed the stabilization process by an order of magnitude. On-off type controllers are generally not adequate to maintain a steady-state condition.

Figure 18.11 illustrates the typical geometry for segmented plate heaters. The measured heat flux occurs from the top surface of the plates that have a thickness of d . The bottom side of the plate is well insulated. The heat loss from the edges usually causes the major errors. The two modes of heat loss (VandenBerghe and Diller, 1989) are shown as q'_1 , the loss per unit width through the insulation by convection from the insulation surface, and q'_2 , the loss per unit width by conduction through the insulation to the adjoining plates. The reason for having the insulation is to thermally isolate the plates to allow

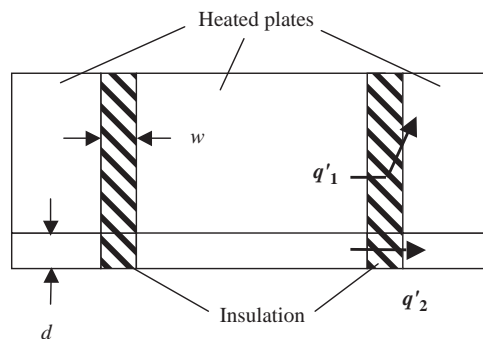


FIGURE 18.11 Segmented-plate constant wall temperature system.

individual heat transfer measurements from each. The potential error represented by q'_2 arises from conduction between the plates when the temperatures are not perfectly matched. It was demonstrated that the one-dimensional conduction solution

$$q'_2 = \Delta T \frac{kd}{w} \quad (18.20)$$

is almost always sufficient to represent this measurement uncertainty. The thermal conductivity of the insulation is k , the thickness of the insulation is w , and the temperature offset between any two plates is represented by ΔT . As expected, the conduction error decreases as the insulation width w increases. Conversely, the correction for the convection loss q'_1 increases with increasing w , and the centerline temperature of the insulation is further depressed below the wall temperature of the plates. The optimal design is where these competing effects are minimized.

Most experiments are designed with an insulation thickness much larger than needed for negligible conduction effects. The convection loss through the insulation and temperature disruption of the surface over the insulation, therefore, are much larger than necessary. In addition, there is a limit to how small the errors can be made for heater plates of any given size. As the measurement plate becomes smaller, the area of the edges increases relative to the surface area of the plate. Consequently, the errors become larger relative to the measured heat transfer as the plate area decreases, which limits the theoretical resolution of the method. Furthermore, if guard plates are not used on all edges, the measurement accuracy is even more limited. Only if the measurement plate is large and the heat transfer coefficient high is the relative convection loss through the insulation on the sides small.

One of the most detailed descriptions of the use of the segmented constant temperature plate method is the review of the turbulent boundary layer experiments at Stanford by Moffat and Kays (1984). The method was used in a well-regulated wind tunnel and continually refined for over 25 years. To achieve the quoted accuracy of $\pm 2\%$ required meticulous detail to experimental procedure and refinement of the data reduction programs. The latter is a lengthy process to eliminate any small errors and to build up an experience base to include all of the small corrections needed.

Small gages for local heat flux measurements have been produced to actively match the surrounding wall temperatures while measuring power dissipated at the surface. Kraabel et al. (1980) used a cone-shaped copper plug to minimize the gap at the surface while still minimizing the thermal interaction with the housing. A differential thermocouple was used to control the power input to a thermistor heater. The design concept is illustrated in Figure 18.12. The experimental uncertainty for steady-state measurements was given as 2%. It should be noted that this low uncertainty is only possible with the very small gap at the surface, cone shape for the sensor and differential thermocouple for the control. Such experiments, however, are more time consuming to perform.

Measurement of time-resolved heat flux requires active control of the power. Thin-film gages deposited on a low conductivity substrate are used to minimize the system time response. A thin-film resistance element can serve as both the heater and temperature measurement device. There are three important issues that must be addressed for accurate time-resolved heat flux measurements. The first is that the thin film gage temperature must be uniform over the entire gage and match the temperature of the surrounding surface. Temperature controllers usually have a small temperature droop. To match the

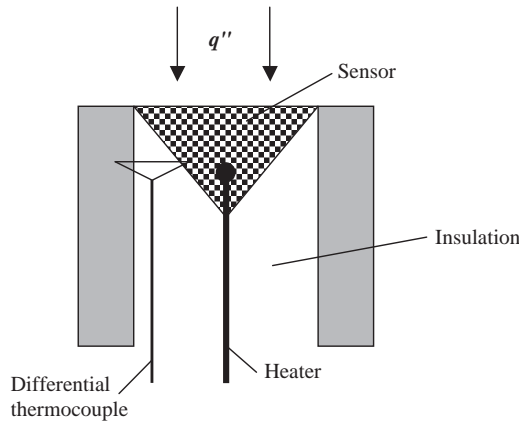


FIGURE 18.12 Heated gage concept.

temperature increases the instability of the system, which requires careful tuning of the system. The third problem is the consequent attenuation of the frequency response because of the thermal capacitance of the substrate material, even if it is a good thermal insulator. Higher frequencies may still be seen in the signal, but the amplitude is much smaller than the actual value. A good test of the system is to use a chopped heat input over a range of frequencies from high to a simple on–off. The measured magnitude should be the same independent of the frequency.

To overcome these problems and experimentally observe the temperature of the heater film Campbell et al. (1989) embedded a thermocouple made with $76\text{ }\mu\text{m}$ wire at the surface of the substrate underneath the sensor film. The low conductivity substrate of the sensor was mounted flush with the surrounding aluminum surface. The heated thin film covered the entire surface of the substrate and was actively powered with a modified hot-wire bridge. To match the sensor and surrounding surface temperatures a second matched thermocouple was embedded in the surrounding surface to form a differential thermocouple. A slow response differential controller was used to match the temperatures by driving the differential to zero. Consequently, the hot wire bridge controlled the power to the sensor during fast surface heat flux transients and the second controller matched temperatures over long times by adjusting the power to the surrounding material. This system was capable of maintaining the temperature match within $\pm 0.1^\circ\text{C}$. The compensation tuning of the bridge had to be adjusted very close to the instability point for proper response and no measurable temperature change of the surface. The resulting heat flux uncertainty was estimated to be $\pm 5\%$.

Dinu et al. (1998) performed a numerical solution of the transient response of their actively heated thin film. There are three components to these probes: the heated thin film, the substrate on which it is deposited, and a protective coating layer overtop. The sensor was assumed to be mounted into an isothermal plate, which defined all of the boundary conditions except the surface. The results of their model clearly indicated the importance of precise control over the sensor temperature. With only a 0.5°C higher temperature the heat flux was 20% higher. Even larger errors were seen in the amplitude of a sinusoidal heat flux.

Lee et al. (2001) created an array of actively heated gages consisting of 96 individual thin-film elements, each $270\text{ }\mu\text{m}$ by $270\text{ }\mu\text{m}$. The entire assembly is only 2.7 mm on

a side. It was used for measuring time-dependent boiling events. Because of the very high heat flux associated with boiling, the losses are relatively small and the controller stability is not as crucial as the gages used in air convection.

18.6 CALIBRATION AND ERRORS

There are two basic approaches to calibration of heat flux sensors. The first is to use a system that provides a comparison with a primary standard. The second uses another heat flux sensor as a secondary standard for the comparison. If the gage operating temperature will be outside the normal room temperature range, the sensitivity should be measured over the range of gage temperatures for the intended use because it may be temperature dependent. In addition, the mode of heat transfer can be important to the heat flux sensor response. Therefore, identification and separation of convection from radiation is an issue to address. Errors can be separated into the measurement uncertainty and the bias error of the calibration.

18.6.1 Heat Flux Gage Calibration

There are three methods for absolute radiation calibration of heat flux sensors as illustrated in the facilities at NIST (Murthy et al., 2000). They are all designed for water-cooled sensors that are kept at room temperature for the calibration. One uses an electrical substitution radiometer as the primary standard (Murthy et al., 2001) in a narrow-angle facility. The second is a spherical cavity which uses the temperature of the wall as the standard along with the Stefan–Boltzmann law for the blackbody radiation to the sensor (ISO/DIS 14934, 2009).

$$q''_{\text{inc}} = \sigma T^4 \quad (18.21)$$

The heat flux sensor should be located at the bottom of the sphere or the sphere should be in a vacuum to minimize the effects of convection. The third method is insertion into a graphite tube blackbody cavity with a pyrometer on the opposite side of the center partition for the primary standard. This method has been tested both experimentally and analytically (Abdelmessih and Horn, 2010; Murthy et al., 2004, 2006). It also has some convection effects and is dependent on the sensor placement in the cavity and time response of the temperature controller of the blackbody during its insertion.

The narrow-angle facility has minimal convection effects because the sensor is located outside of the blackbody high-temperature source. However, because of the narrow-angle radiation beam that heats the sensor, the maximum heat flux is relatively low (50 kW/m²) and it has shorter wavelength (higher required source temperature) than the other methods. In addition, most coatings have some angular dependence of the absorptivity (Alpert et al., 2003). The result is a small, but consistent difference in calibration results between the narrow-angle and wide-angle facilities (Murthy et al. 2006).

For all three cases the resulting calibration for the gage sensitivity S_{rad} is based on the incident radiation to the gage surface, q''_{inc} .

$$S_{\text{rad}} = \frac{E}{q''_{\text{inc}}} \quad (18.22)$$

The emissivity of the surface ε is required to calculate the absorbed radiation.

$$q''_{\text{abs}} = \varepsilon q''_{\text{inc}} \quad (18.23)$$

The gage surface is assumed to be gray with the absorptivity equal to the emissivity, ε , and both values constant with angle and wavelength for the conditions of use. The gages are kept at room temperature, using water cooling if necessary. This minimizes any emission from the gage surface, which is neglected.

Once the secondary standard gage has been calibrated against a primary standard, it is used to calibrate other gages by comparison under the same incident heat flux. One method is with simultaneous measurements on either side of an electrically heated flat plate. The other common method is using sequential substitution under a quartz lamp bank. In all cases the size and range of the two gages should be matched to minimize the differences in the radiation experienced.

If the gages are to be used at elevated temperatures, the calibrations are more difficult. First, the gage will experience substantial convection. Second, the emitted energy from the gage will become important. These require corrections and raise the uncertainty of the calibration. Pullins and Diller (2010) have minimized these problems by designing an enclosure that mounts the gage to be calibrated on the hot surface and the secondary standard in a water-cooled cold plate. Convection is minimized by putting the cold plate horizontal and beneath the hot plate. Because the enclosure is insulated, essentially all of the energy from the hot plate necessarily goes to the cold plate, as measured by the secondary standard. Corrections are needed only for the small non-uniformities of temperature of the hot and cold surfaces.

Convection calibrations are more difficult, but have been done to heat flux levels of about 5 kW/m^2 using electrical power measurement as the standard (Borell and Diller, 1987; Holmberg and Womeldorf, 1999). The gage sensitivity is based on the convective heat flux, q''_{conv} , provided by a wind tunnel (Holmberg and Womeldorf, 1999) or a large stagnating air jet (Borell and Diller, 1987)

$$S_{\text{conv}} = \frac{E}{q''_{\text{conv}}} \quad (18.24)$$

The standard is the electrical power input into a constant temperature plate with guard heaters on all sides and operating at steady-state conditions, as discussed in Section 18.5.2. The resulting sensitivity corresponds to the absorbed radiation. Consequently, the usual incident radiation calibration must be adjusted to correspond to the convection sensitivity by the emissivity of the original surface coating.

$$S_{\text{conv}} = \frac{S_{\text{inc}}}{\varepsilon} \quad (18.25)$$

Conduction calibrations are time consuming for the thin flat gages because of the time required to reach steady-state conditions. Guarded hot plate calibration systems are commercially available and can be traced to standards at NIST. An ASTM standard covers the calibration method (ASTM C1130-07, 2009).

$$S_{\text{cond}} = \frac{E}{q''_{\text{cond}}} \quad (18.26)$$

18.6.2 Error Estimates

Heat flux gage calibrations usually are repeatable to within 2–3%. The bias error introduced into the measurement uncertainty will be larger depending on how close the measurement conditions are to the calibration conditions. The precision uncertainty is usually dominated by the voltage measurement uncertainty. This can be minimized by properly specifying the heat flux range of the gage and using quality amplifiers and/or data-acquisition units. For radiation the coating properties may contribute a substantial uncertainty. Taken together the combined uncertainty for heat flux gage measurements that are performed carefully is often between 5% and 10% (Murthy et al., 2005).

REFERENCES

- Abdelmessih AN, Horn TJ. Experimental and computational characterization of high heat fluxes during transient blackbody calibrations. *Journal of Heat Transfer* 2010;132:023304.
- Agnone AM. Solution of the nonlinear heat equation for circular-foil heat flux gauges. *Journal of Thermophysics Heat Transfer* 2009;23:33–40.
- Ahn J-W, Maingi R, Mastrovito D, Roquemore AL. High speed infrared camera diagnostic for heat flux measurement in NSTX. *Review of Scientific Instruments* 2010;81:023501.
- Ainsworth RW, Allen JL, Davies MRD, Doorly JE, Forth CJP, Hilditch MA, Oldfield MLG, Sheard AG. Developments in instrumentation and processing of transient heat transfer measurements in a full-stage model turbine. *ASME Journal of Turbomachinery* 1989;111:20–27.
- Alpert RL, Orloff L, de Ris JL. Angular Sensitivity of Heat Flux Gages, in ASTM STP 1427, 2003; p. 67–80.
- Arai N, Matsunami A, Churchill SW. A review of measurements of heat flux density applicable to the field of combustion. *Experimental Thermal and Fluid Science* 1996;12:452–460.
- ASTM C1130-07, Standard practice for calibrating thin heat flux transducers. *Annual Book of ASTM Standards*, 4.06. 2009.
- ASTM E2683-09, Standard test method for measuring heat flux using flush-mounted insert temperature-gradient gages. *Annual Book of ASTM Standards*, 15.03. 2009.
- ASTM E2684-09, Standard test method for measuring heat flux using surface-mounted one-dimensional flat gages. *Annual Book of ASTM Standards*, 15.03. 2009.
- ASTM E511-07, Measurement of heat flux using a copper-constantan circular foil heat-flux gage. *Annual Book of ASTM Standards*, 15.03. 2009.
- ASTM E547-08, Standard test method for measuring heat-transfer rate using a thermal capacitance (slug) calorimeter. *Annual Book of ASTM Standards*, 15.03. 2009.
- ASTM E598-08, Standard test method for measuring extreme heat-transfer rates from high-energy environments using a transient, null-point calorimeter. *Annual Book of ASTM Standards*, 15.03. 2009.
- Baba T, Ono A, Hattori S. Analysis of the operational error of heat flux transducers placed on wall surfaces. *Review of Scientific Instruments* 1985;56:1399–1401.
- Bales E, Bomberg M, Courville GE. *Building Applications of Heat Flux Transducers*, ASTM; 1985.
- Bauer WD, Heywood JB. Transfer function of thin-film heat flux sensor. *Experimental Heat Transfer* 1997;10:181–190.
- Bizzak DJ, Chyu MK. Use of laser-induced fluorescence thermal imaging system for local jet impingement heat transfer measurement. *International Journal of Heat Mass Transfer* 1995;38:267–274.

- Borell GJ, Diller TE. A convection calibration method for local heat flux gages. *Journal of Heat Transfer* 1987;109:83–89.
- Brajuskovic B, Afgan N. A heat flux meter for ash deposit monitoring systems–II. Clean heat flux-meter characteristics. *International Journal of Heat Mass Transfer* 1991;34:2303–2315.
- Buchlin JM. Convective heat transfer and infrared thermography (IRTH). *Journal of Applied Fluid Mechanics* 2010;3:55–62.
- Campbell DS, Gundappa M, Diller TE. Design and calibration of a local heat-flux measurement system for unsteady flows. *ASME Journal of Heat Transfer* 1989;111:552–557.
- Childs PRN, Greenwood JR, Long CA. Heat flux measurement techniques. *Proceedings of the Institution of Mechanical Engineers* 1999;213C:655–677.
- Colban W, Gratton A, Thole KA, Haendler M. Heat transfer and film-cooling measurements on a stator vane with fan-shaped cooling holes. *Journal of Turbomachinery* 2006;128:53–61.
- Das MK, Tariq A, Phanigrahi PK, Muralidhar K. Estimation of convective heat transfer coefficient from transient liquid crystal data using an inverse technique. *Inverse Problems in Science and Engineering* 2005;13:133–155.
- Diller TE, Onishi S. Heat Flux Gage. U.S. Patent No. 4,779,994, Issued 25 October, 1988.
- Diller TE. Advances in heat flux measurement. In: Hartnett JP et al., editors. *Advances in Heat Transfer*, 23, Boston: Academic Press; 1993. p. 279–368.
- Diller TE. Heat Flux. In: Webster JG, editor. *The Measurement, Instrumentation and Sensors Handbook*, 34, Boca Raton: CRC Press; 1999. p. 34.1–15.
- Dinu C, Beasley DE, Figliola RS. Frequency response characteristics of an active heat flux gage. *ASME Journal of Heat Transfer* 1998;120:577–582.
- Doorly JE, Oldfield MLG. New heat transfer gages for use on multilayered substrates. *ASME Journal of Turbomachinery* 1986;108:153–160.
- Dunn MG, George WK, Rae WJ, Woodward SH, Moller JC, Seymour PJ. Heat-flux measurements for a full-stage turbine: part ii—description of analysis technique and typical time-resolved measurements. *ASME Journal of Turbomachinery* 1986;108:98–107.
- Dunn MG, Kim J, Rae WJ. Investigation of the heat-island effect for heat-flux measurements in short-duration facilities. *Journal of Turbomachinery* 1997;119:753–760.
- Ekkad SV, Ou S, Rivir RB. A transient thermography method for simultaneous film effectiveness and heat transfer coefficient measurements from a single test. *Journal of Turbomachinery* 2004;126:597–603.
- Epstein AH, Guenette GR, Norton RJG, Cao Y. High-frequency response heat-flux gauge. *Review of Scientific Instruments* 1986;57:639–649.
- Ewing J, Gifford A, Hubble D, Vlachos P, Wicks A, Diller T. A direct-measurement thin-film heat flux sensor array. *Measurement Science and Technology* 2010;21: 105201, 8.
- Flanders SN. Heat flux transducers measure “in situ” building thermal performance. *Journal of Thermal Insulation and Building Envs.* 1994;18:28–52.
- Fralick G, Wrbanek J. Thin Film Heat Flux Sensor of Improved Design. NASA TM-2002-211566, 2002.
- Gardon R. An instrument for the direct measurement of intense thermal radiation. *The Review of Scientific Instruments* 1953;24:366–370.
- Gifford A, HOFFIE A, Diller T, Huxtable S. Convection calibration of Schmidt–Boelter heat flux gages in stagnation and shear air flow. *ASME Journal of Heat Transfer* 2010;132: 031601, 9.
- Gifford AR, Hubble DO, Pullins CA, Huxtable ST, Diller TE. A durable heat flux sensor for extreme temperature and heat flux environments. *Journal of Thermophysics and Heat Transfer* 2010;24:69–76.

- Guo SM, Lai CC, Jones TV, Oldfield MLG, Lock GD, Rawlinson AJ. The application of thin-film technology to measure turbine-vane heat transfer and effectiveness in a film-cooled, engine-simulated environment. *International Journal of Heat Fluid Flow* 1998;19:594–600.
- Hager NE. Jr., Thin foil heat meter. *The Review of Scientific Instruments* 1965;36:1564–1570.
- Hager JM, Simmons S, Smith D, Onishi S, Langley LW, Diller TE. Experimental performance of a heat flux microsensor. *Journal of Engineering for Gas Turbines and Power* 1991;113:246–250.
- Hager JM, Onishi S, Langley LW, Diller TE. High temperature heat flux measurements. *AIAA Journal of Thermophysics Heat Transfer* 1993; 531–534.
- Han J-C, Dutta S, Ekkad S. *Gas Turbine Heat Transfer and Cooling Technology*, NY.: Taylor and Francis; 2001.
- Hauser RL. Construction and performance of *in situ* heat flux transducers. In: Bales E. et al., editors. *Building Applications of Heat Flux Transducers*, ASTM STP 885; 1985. p. 172–183.
- Hevey S, Langley L, Schmidt–Boelter Gage. U.S. Patent No. 6,186,661, Issued 13 February, 2001.
- Hightower TM, Olivares RA. Thermal Capacitance (Slug) Calorimeter Theory Including Heat Losses and Other Decaying Processes, NASA/TM-2008-215364, 2008.
- Holmberg DG, Diller TE. High-frequency heat flux sensor calibration and modeling. *ASME Journal of Fluids Engineering* 1995;117:659–664.
- Holmberg DG, Womeldorf CA. Design and Uncertainty analysis of a second-generation convective heat flux calibration facility. *HTD*, 364-4, N.Y.: ASME; 1999. p. 65–70.
- Hubble DO, Diller TE. Development and evaluation of the time-resolved heat and temperature array. *Journal of Thermal Science and Engineering Applications* 2010;2: 031003, 9.
- Hubble DO, Diller TE. A hybrid method for measuring heat flux. *Journal of Heat Transfer* 2010;132:031602, 8.
- Ireland PT, Jones TV. Liquid crystal measurements of heat transfer and surface shear stress. *Measurement Science and Technology* 2000;11:969–986.
- ISO/DIS 14934, Fire Tests—Calibration and Use of Heat Flux Meters, 2009.
- Kandular M, Haddad G. Two-dimensional thermal boundary layer corrections for convective heat flux gauges. *Journal of Thermophysics Heat Transfer* 2007;21:543–47.
- Keltner NR. Heat flux measurements: theory and applications (Chapter 8). In: Azar K. editor. *Thermal Measurements in Electronics Cooling*. Boca Raton: CRC Press; 1997. p. 173–320.
- Keltner NR, Beck JV, Nakos JT. Using directional flame thermometers for measuring thermal exposure. *Journal of ASTM International* 2010;7:102280.
- Kendall JM, Berdahl CM. Two blackbody radiometers of high accuracy. *Applied Optics* 1970;9:1082–1091.
- Khalid AH, Kontis K. Thermographic phosphors for high temperature measurements: principles, current state of the art and recent applications. *Sensors* 2008;8:5673–5744.
- Kidd CT, Nelson CG. How the Schmidt–Boelter gage really works. Proceedings of the 41st International Instrumentation Symposium, ISA, Research Triangle Park, 1995, pp. 347–368.
- Kidd CT, Adams JC. Jr., Fast-response heat-flux sensor for measurement commonality in hypersonic wind tunnels. *Journal of Spacecraft and Rockets* 2001;38:719–729.
- Kidd CT. Coaxial surface thermocouples: analytical and experimental considerations for aerothermal heat-flux measurement applications. Proceedings 36th International Instrumentation Symposium, ISA, Research Triangle Park, 1990, pp. 203–211.
- Kidd CT. High heat flux measurements and experimental calibrations/characteristics. NASA 1992; CP3161:31–50.
- Kinzie PA. *Thermocouple Temperature Measurement*. NY.: John Wiley; 1973.

- Kraabel JS, Baughn JW, McKillop AA. An Instrument for the measurement of heat flux from a surface with uniform temperature. *Journal of Heat Transfer* 1980;102:576–578.
- Kuo CH, Kulkarni AK. Analysis of heat flux measurement by circular foil gages in a mixed convection/radiation environment. *Journal of Heat Transfer* 1991;113:1037–1040.
- Kurits I, Lewis MJ. Global temperature-sensitive paint system for heat transfer measurements in long-duration hypersonic flows. *Journal of Thermophysics Heat Transfer* 2009;23:256–266.
- Langley L, Barnes A, Matijasevic G, Gandhi P. High sensitivity, surface-attached heat flux sensors. *Microelectronics Journal* 1999;30:1163–68.
- Lee J, Kim J, Kiger KT. Time and space resolved heat transfer characteristics of single droplet cooling using microscale heater arrays. *International Journal of Heat and Fluid Flow* 2001;22:188–200.
- Liechty BC, Clark MM, Jones MR, Larson RS, Woolford BL. Nonlinear thermal model of circular foil heat flux gauges. *Journal of Thermophysics and Heat Transfer* 2007;21:468–474.
- Long CA, Childs PRN, Greenwood JR, Tham KM. Manufacture and calibration of robust heat flux sensors for rotating turbomachinery. *Experimental Heat Transfer* 2004;17:181–197.
- Martins N, Carvalho MG, Afgan N, Leontiev AI. A radiation and convection fluxmeter for high temperature applications. *Experimental Thermal and Fluid Science* 2000;22:365–373.
- Martins N, Calisto H, Afgan N, Leontiev AI. The transient transpiration heat flux meter. *Applied Thermal Engineering* 2006;26:1152–55.
- Mityakov AV, Mityakov V Yu, Sapozhnikov SZ, Chumakov Yu. S. Application of the transverse Seebeck effect to measurement of instantaneous values of a heat flux on a vertical heated surface under conditions of free-convection heat transfer. *High Temperature* 2002;40:620–625.
- Moffat RJ, Eaton JK, Mukerji D. A general method for calculating the heat island correction and uncertainties for button gauges. *Measurement Science and Technology* 2000;11:920–932.
- Moffat RJ, Kays WM. A review of turbulent-boundary-layer heat transfer research at Stanford. *Advances in Heat Transfer* 1948–1983 1984;16:241–365.
- Murthy AV, Wetterlund I, DeWitt DP. Characterization of an ellipsoidal radiometer. *Journal of Research of the NIST* 2003;108:115–124.
- Murthy AV, Tsai BK, Saunders RD. Radiative calibration of heat-flux sensors at NIST: facilities and techniques. *Journal of Research of the National Institute of Standards & Technology* 2000;105:293–305.
- Murthy AV, Tsai BK, Saunders RD. Transfer calibration validation tests on a heat flux sensor in the 51 mm high-temperature blackbody. *Journal of Research of the National Institute of Standards & Technology* 2001;106:823–831.
- Murthy AV, Fraser FT, DeWitt DP. Experimental in-cavity radiative calibration of high-heat-flux meters. *Journal of Thermophysics and Heat Transfer* 2006;20:327–335.
- Murthy AV, Prokhorov AV, DeWitt DP. High heat-flux sensor calibration: a Monte Carlo modeling. *Journal of Thermophysics and Heat Transfer* 2004;18:333–341.
- Murthy AV, Fraser GT, Dewitt DP. A summary of heat-flux sensor calibration data. *Journal of Research of the NIST* 2005;110:97–100.
- Ochoa AD, Baughn JW, Byerley AR. A new technique for dynamic heat transfer measurements and flow visualization using liquid crystal thermography. *International Journal of Heat and Fluid Flow* 2005;26:264–275.
- Oldfield MLG. Impulse response processing of transient heat transfer gauge signals. *Journal of Turbomachinery* 2008;130:021023.
- Ortolano DJ, Hines FF. A Simplified Approach to Heat Flow Measurement. *Advances in Instrumentation*, 38(II), Research Triangle Park: ISA; 1983. p. 1449–1456.

- Piccini E, Guo SM, Jones TV. The development of a new direct-heat-flux gauge for heat-transfer facilities. *Measurement Science and Technology* 2000;11:342–349.
- Praisner TJ, Smith CR. The dynamics of the horseshoe vortex and associated endwall heat transfer—part 1: temporal behavior. *Journal of Turbomachinery* 2006;128:747–754.
- Pullins CA, Diller TE. *In situ* high temperature heat flux sensor calibration. *International Journal of Heat and Mass Transfer* 2010;53:3429–3438.
- Robertson AF, Ohlemiller TJ. Low heat-flux measurements: some precautions. *Fire Safety Journal* 1995;25:109–124.
- Samant KR, Simon TW, Stuart RV. Using thin-film technology to fabricate a small—patch boiling heat transfer test section. In: Jones OC, Farukhi NM, editors. *New Experimental Techniques in Heat Transfer*. NY.: ASME; 1984. p. 33–38.
- Schultz DL, Jones TV. Heat transfer measurements in short duration hypersonic facilities. *AGAR-Dograph* 165, 1973.
- Shewen EC, Holland KGT, Raithby GD. The measurement of surface heat flux using the Peltier effect. *ASME Journal of Heat Transfer* 1989;111:798–803.
- Smith TB, Schetz JA, Walker DG. Development and ground testing of heat flux gages for high enthalpy supersonic flight tests. *AIAA Paper No.* 2002–3133, 2002.
- Talib ARA, Neely AJ, Ireland PT, Mullender AJ. A novel liquid crystal image processing technique using multiple gas temperature steps to determine heat transfer coefficient distribution and adiabatic wall temperature. *Journal of Turbomachinery* 2004;126:587–596.
- Talib ARA, Neely AJ, Ireland PT, Mullender AJ. Detailed investigation of heat flux measurements made in a standard propane–air–fire-certification burner compared to levels derived from a low-temperature analog burner. *Journal of Engineering for Gas Turbines and Power* 2005;127:249–56.
- Theophilos TS, Longtin JP, Sampath S, Tankiewicz S, Gambino RJ. Integrated heat-flux sensors for harsh environments using thermal-spray technology. *IEEE Sensors Journal* 2010;6:1126–33.
- Van der Graaf F. Heat flux sensors. In: Gopel W. et al., editors. *Sensors*. 4, N.Y.: VCH; 1989. p. 295–322.
- Van Dorth AC. Thick film heat flux sensor. *Sensors and Actuators* 1983;4:323–331.
- VandenBerghe T, Diller TE. Analysis and design of experimental systems for heat transfer measurement from constant temperature surfaces. *Experimental Thermal and Fluid Science* 1989;2:236–246.
- Wagner G, Kotulla M, Ott P, Wegand B, von Wolfersdorf J. The transient liquid crystal technique: influence of surface curvature and finite wall thickness. *Journal of Turbomachinery* 2005;127:175–182.
- Walker DG, Scott EP. Evaluation of estimation methods for high unsteady heat fluxes from surface measurements. *Journal of Thermophysics and Heat Transfer* 1998;12:543–551.
- Walker DG. Heat flux determination from measured heating rates using thermographic phosphors. *Journal of HeatTransfer* 2005;127:560–570.
- Wesley DA. Thin disk on a convectively cooled plate—application to heat flux measurement errors. *ASME Journal of Heat Transfer* 1979;101:346–352.

19

HEAT TRANSFER MEASUREMENTS FOR NONBOILING TWO-PHASE FLOW

AFSHIN J. GHAJAR AND CLEMENT C. TANG

- 19.1 Introduction
- 19.2 Experimental setup for horizontal and slightly inclined pipes
- 19.3 Instruments for measurement and data acquisition
- 19.4 Heat transfer experiment procedures
- 19.5 Verifying the functionality of the experimental setup
 - 19.5.1 Frictional pressure drop in single-phase flow
 - 19.5.2 Heat transfer in single-phase flow
- 19.6 Experimental results of two-phase flow
 - 19.6.1 Flow patterns and flow maps
 - 19.6.2 Heat transfer in horizontal and slightly upward inclined pipe flows
- 19.7 Concluding remarks
- Nomenclature
- References

19.1 INTRODUCTION

In many industrial applications, such as the transport of oil and natural gas in pipelines and wellbores, the knowledge of nonboiling two-phase, two-component (liquid and permanent gas), heat transfer is required. During the transport of two-phase hydrocarbon fluids from an oil reservoir to the surface, temperature of the hydrocarbon fluids changes due to the difference in temperatures of the oil reservoir and the surface. The difference in temperature results in heat transfer between the hydrocarbon fluids and the Earth surrounding the oil well. In such situation, the ability to estimate the flowing temperature profile is necessary to address several design problems in petroleum production engineering (Shiu and Beggs, 1980).

In subsea oil and natural gas production, hydrocarbon fluids may exit the reservoir with a temperature of 75°C and flow in subsea surroundings of 4°C (Trevisan et al., 2006). As a result of the temperature difference between the reservoir and the surroundings, the knowledge of heat transfer is critical in order to prevent gas hydrate and wax deposition blockages (Furuholt, 1988). Wax deposition can cause severe problems including the reduction of inner pipe diameter causing blockage, increase in surface roughness of the pipe leading to restricted flow line pressure, decrease in production, and various other mechanical problems (McClaflin and Whitfill, 1984). Here are a couple of examples of the economical losses that were caused by wax deposition blockages: (1) a direct cost of \$5 million in removing the blockage from a subsea pipeline, and (2) the cost of an oil platform abandonment by Lasmo Company (UK) that amounted to \$100 million (Singh et al., 2000).

Schemes such as coil-spring wire insert, twisted tape insert, and helical ribs have been used to promote turbulence in pipes for the purpose of enhancing heat transfer. Although such heat transfer enhancement schemes have been proven to be effective, they do come with drawbacks, such as fouling, increases in pressure drop, and even blockage. Celata et al. (1999) presented an approach to enhance heat transfer in pipe flow by injecting small amount of gas into the liquid flow to promote turbulence. In the experimental study performed by Celata et al. (1999), a uniformly heated vertical pipe was internally cooled by water, while heat transfer coefficients with and without air injection were measured. The introduction of a small airflow rate into the water flow resulted in increases of the heat transfer coefficient up to 20–40% for forced-convection, and even larger heat transfer enhancement for mixed-convection (Celata et al., 1999).

Furthermore, since the mechanisms of heat and mass transfers are analogous to one another, the knowledge of two-phase heat transfer can be applicable to two-phase mass transfer. In order to successfully apply the knowledge of two-phase heat transfer to solve mass transfer problems, the appropriate parallels between the two mechanisms have to be first sorted out. Mass transfer in two-phase flow can be found in many chemical processes. One conceivable application is predicting the rate of corrosion in pipes that transport two-phase flow. In cases when chemical reactions between the pipe surface and the two-phase fluids are heavily influenced by mass transfer, the ability to predict the mass transfer coefficient becomes very beneficial to engineers.

In the endeavor to understand and characterize nonboiling two-phase flow heat transfer in pipes, systematic experimental measurements of heat transfer data are necessary. The contents within this article present the experimental setup used for conducting measurements of heat transfer in two-phase flow. In addition, the experimental procedures to properly measure heat transfer data, as well as the functionality of the experimental setup are discussed. Finally, the measured heat transfer results and their characteristics are presented.

19.2 EXPERIMENTAL SETUP FOR HORIZONTAL AND SLIGHTLY INCLINED PIPES

The experimental setup for slightly inclined pipe orientations was constructed with the capability of using it to measure nonboiling two-phase pressure drop and heat transfer, and conduct flow visualization for all major flow patterns and upward inclination angles from 0° (horizontal) to 7°. A schematic diagram of the flow loop for the current

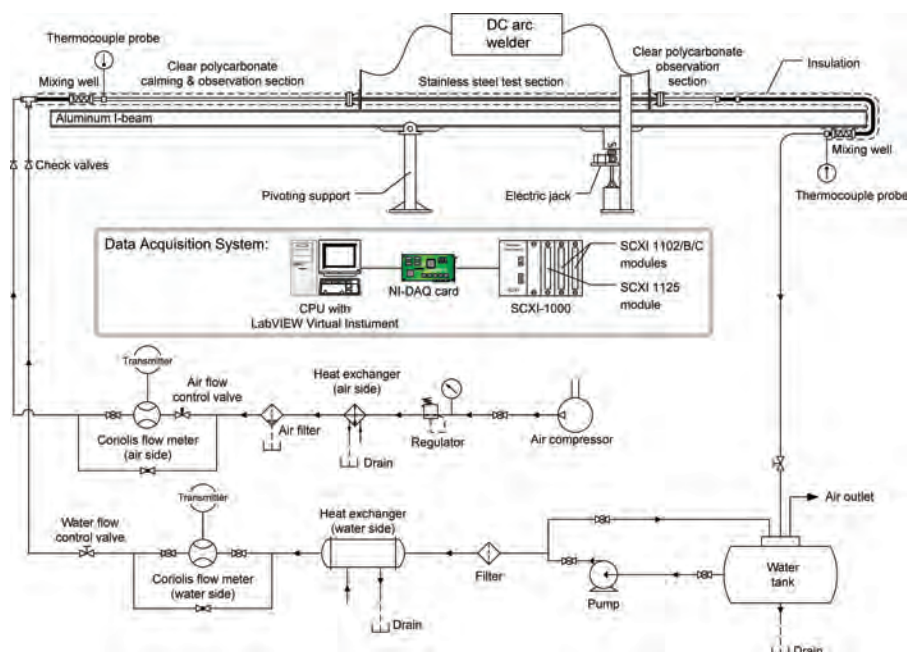


FIGURE 19.1 Flow loop of the experimental setup.

experimental setup is illustrated in Figure 19.1. The mixing section, the calming and observation section, and the stainless steel test section rest on top of an aluminum I-beam. The aluminum I-beam is supported by a pivoting mechanism that is attached to an electric jack. The I-beam is approximately 9.1 m long and the pivoting mechanism could bring the I-beam to an upward inclination of approximately 8° . Inclination angles were measured with a contractor's angle-measuring tool and with a more precise digital x - y axis accelerometer. The minimum resolution of the digital accelerometer is 0.5° .

The working fluids used in this study were air and distilled water. The reservoir used for storing the distilled water is a 208-L (55 gal) polyethylene tank. A Bell & Gossett (Series 1535) coupled centrifugal pump (size 3545 D10) is used to pump the distilled water from the reservoir through an Aqua-Pure AP12T water filter and an ITT Standard (Model BCF 4063) one shell and two-tube pass heat exchanger. The heat exchanger is used to remove heat added to the distilled water during the experiment as well as to maintain a constant inlet liquid temperature. The cooling fluid used by the heat exchanger is tap water taken directly from the wall tap.

From the heat exchanger, the distilled water flowed through a Micro Motion (Model CMF125) Coriolis flow meter. The Coriolis flow meter is connected to a Digital Field-Mount (Model RFT9739) transmitter that conditions the flow information for the data acquisition system. After passing the Coriolis flow meter, the distilled water then passes through a gate valve. The gate valve is used for regulating the amount of distilled water that is flowing into the mixing section. From the gate valve, the distilled water flows through a 25.4 mm I.D. hose, and then through a one-way check valve and into the air-water mixing section.

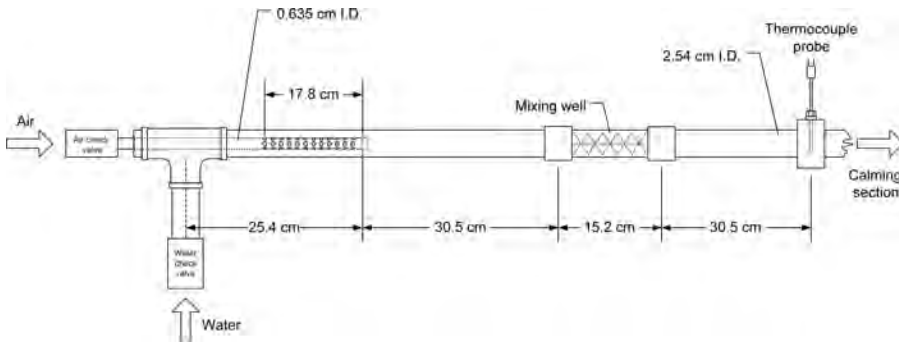


FIGURE 19.2 Schematic of gas–liquid mixing section.

Air is supplied from an Ingersoll-Rand T30 (Model 2545) air compressor. The air supplied from the air compressor is regulated by a Speedaire (Model 4ZM22) 12.7 mm regulator. The air is then cooled by passing through a copper coil submerged in a vessel of water from the wall tap. The same water from the wall tap that is used for cooling the distilled water is also used for cooling the air. This is to ensure that both air and distilled water have the same inlet temperature. The air is then filtered by a Speedaire (Model 4ZL49) 12.7 mm air filter to remove moisture from the air before the airflow rate is being measured.

From the air filter, the air flows through a Micro Motion (Model CMF100) Coriolis flow meter that is connected to another Digital Field-Mount (Model RFT9739) transmitter. After passing the Coriolis flow meter, the air then passes through a needle valve, which is used to regulate the amount of air flowing into the mixing section. From the needle valve, the air flows through a flexible hose, and then through a one-way check valve and into the air–water mixing section.

The air–water mixing section is attached upstream of the calming and observation section. A schematic of the air–water mixing section is illustrated in Figure 19.2. Similar type of mixing section has been successfully used in two-phase flow experimental studies by Ewing et al. (1999) and Kim (2000). Air and distilled water are converged in a 25.4 mm copper tee and are further mixed in the mixing well. After the mixing well, the temperature of the air–water two-phase flow is measured by an Omega TMQSS-125U-6 thermal probe. Exiting the mixing well, the air–water two-phase flow enters into the calming and observation section.

The 2.24 m long calming and observation section is made of clear polycarbonate pipe with an inner diameter of 25.4 mm. This gives the calming and observation section a length to diameter ratio (L/D) of 88. Air–water two-phase flow leaving the mixing section is allowed to calm and develop through the calming and observation section before entering the test section. The calming and observation section also serves as the section where flow pattern visualization is conducted. Two-phase flow exiting the calming and observation section enters the heat transfer test section.

The heat transfer test section is made of a schedule 10S 316 stainless steel pipe with an inner diameter of 27.9 mm and an outer diameter of 33.4 mm. The heat transfer test section is 2.64 m long, giving a length to diameter ratio (L/D) of about 95. A schematic of the heat transfer test section is illustrated in Figure 19.3. To measure the surface temperatures

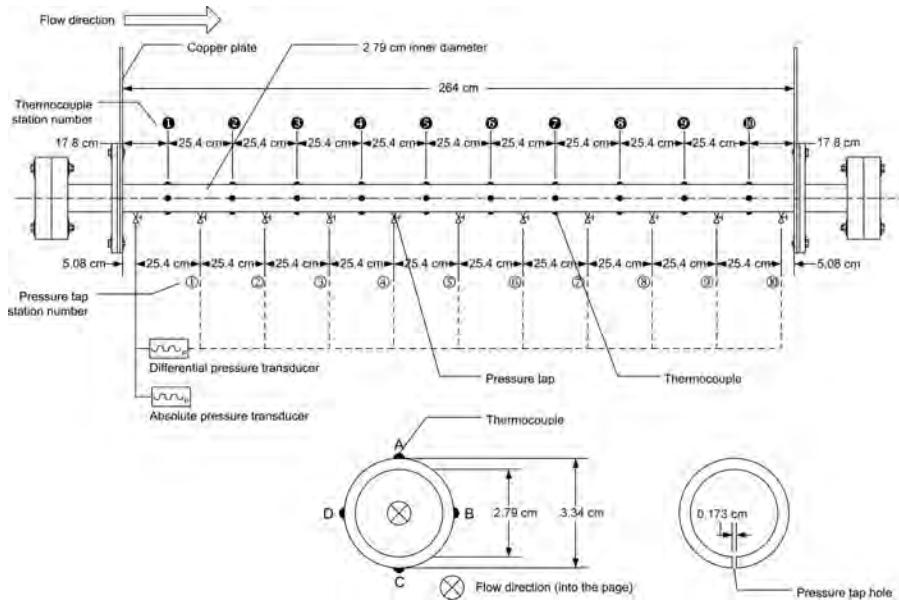


FIGURE 19.3 Schematic of heat transfer test section.

of the heat transfer section, 10 thermocouple stations, equally spaced with 25.4-cm interval, are placed along the test section. Each thermocouple station has four thermocouples attached circumferentially around the pipe surface. Figure 19.3 shows the circumferential locations of the thermocouples, with “A” at the top of the pipe, “B” at 90° from the top in the clockwise direction, “C” at the bottom of the pipe, and “D” at 90° from the bottom in the clockwise direction.

The thermocouples attached on the heat transfer section are Omega TT-T-30 copper-constantan insulated T-type thermocouples. Omega EXPP-T-20-TWSH extension wires are used for connecting the thermocouples to the data acquisition system. The thermocouples are cemented on the outside surface of the stainless steel test section with Omegabond 101 epoxy. The Omegabond 101 epoxy has a thermal conductivity of 1.04 W/mK and an electrical resistivity of $1 \times 10^{15} \Omega\text{m}$.

Two 12.7 cm \times 17.8 cm \times 0.64 cm copper plates are attached to one of each end of the stainless steel heat transfer test section. The two copper plates are attached to the ends of the test section by means of silver-soldering. Both copper plates are connected with 4 gauge insulated cables to a direct current arc welder, which provides power to heat the test section with uniform heat flux.

Uniform wall heat flux is supplied to the heat transfer test section by running high amperage direct current through the stainless steel test section from a direct current arc welder. For two-phase flow with superficial liquid Reynolds number above 2,000, the test section is heated by a LINCOLNWELD SA-750 arc welder. The minimum current supplied by the LINCOLNWELD SA-750 arc welder is 300 A, which is too high for flow with superficial liquid Reynolds number, $Re_{SL} < 2,000$, and may cause dry-out and local boiling. To avoid dry-out and local boiling, a Miller Maxtron 450 DC inverter arc welder is used to supply direct current through the stainless steel test section for flow with Re_{SL}

$< 2,000$. Using both arc welders provide a wide range of uniform wall heat flux, which makes measurement of heat transfer for two-phase flow at different gas and liquid flow rates possible, thus covering most major flow patterns.

Along the bottom of the heat transfer test section, there are 11 pressure taps equally spaced at 25.4-cm interval. The pressure tap holes, each with diameter of 1.73 mm, were drilled along the bottom of the test section (see Figure 19.3). The pressure taps are connected to standard self-tapping saddle valves with the tapping core removed. The system pressure is measured by Omega PX242-060G absolute pressure transducers, whereas the pressure drop is measured by a Validyne DP15 differential pressure transducer.

19.3 INSTRUMENTS FOR MEASUREMENT AND DATA ACQUISITION

The airflow rate is measured by Micro Motion (Model CMF100) Coriolis flow meter, and the water flow rate is measured by Micro Motion (Model CMF125) Coriolis flow meter. Both Coriolis flow meters are connected to Digital Field-Mount (Model RFT9739) transmitters. The signals from the Digital Field-Mount transmitters are relayed to the data acquisition system for data recording.

Surface temperatures of the test section are measured with Omega TT-T-30 T-type thermocouples cemented on the test section using Omegabond 101. Air–water mixture temperatures at the inlet and outlet of the heat transfer test section are measured with Omega TMQSS-125U-6 thermocouple probes. Signals from the thermocouples and the thermocouple probes are relayed to the data acquisition system for data recording. Isothermal measurements are conducted periodically, to ensure the accuracy and precision of the thermocouples and thermocouple probes.

Pressure drop between the first and the last pressure taps is measured using Validyne DP15 differential pressure transducer. The Validyne DP15 differential pressure transducer is connected to a Validyne CD15 carrier demodulator. Output signals from the differential pressure transducer are sent to the carrier demodulator to be demodulated, amplified, and filtered. Signals from the demodulator are then relayed to the data acquisition system for data recording.

The voltage drop across the heated test section is measured by a HP 3468B digital multimeter. The current flowing through the test section is determined by measuring the voltage drop across a shunt attached to the copper plate at the downstream of the test section. Knowing the voltage drop and the resistance across the shunt, the current flowing through the test section can be determined. The voltage drop across the shunt is recorded by the data acquisition system, and the corresponding current is determined. Knowing the voltage drop across the test section and the current through the test section, the heat flux on the test section can be determined.

A National Instruments data acquisition system is used for recording and storing the data measured during the experiment. An AC powered four-slot National Instruments SCXI 1000 Chassis houses the data acquisition system. The chassis provides a low noise environment for signal conditioning. There are three National Instruments SCXI control modules housed inside the chassis: two SCXI 1102/B/C modules and one SCXI 1125 module.

Input signals from 40 thermocouples, 2 thermocouple probes, volt and current meters, and flow meters are gathered by the 3 modules and sent to a computer to be recorded. A graphical user interface used for the data acquisition is a customized LabVIEW Virtual Instrument (VI) program developed specifically for this experimental setup. Figure 19.4

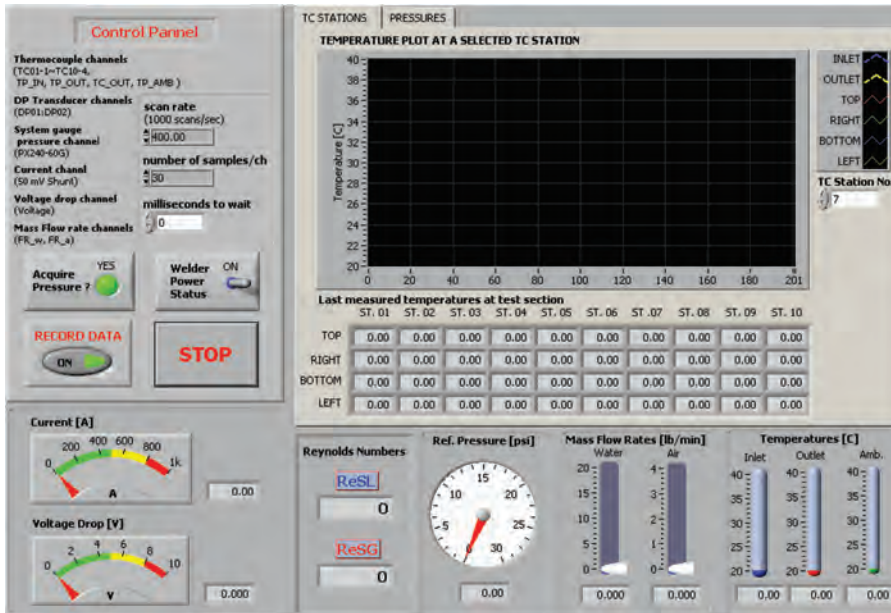


FIGURE 19.4 Graphical user interface of the LabVIEW VI program.

illustrates the graphical user interface of the LabVIEW VI program used for the data acquisition.

The LabVIEW VI graphical user interface provides users with the features to monitor and record data. Data such as inlet, outlet, surrounding, test section surface temperatures, pressure drop, system pressure, air and water mass flow rates, superficial gas and liquid Reynolds numbers, and current supplied by the DC arc welder to the test section are displayed on the graphical user interface. With these features on the graphical user interface, users can monitor the recorded data and readily identify any anomaly during the course of an experiment.

19.4 HEAT TRANSFER EXPERIMENT PROCEDURES

The process of acquiring accurate experimental data requires a consistent adherence to a set of defined experimental procedures. The purpose of adhering to the experimental procedures is to maintain the integrity and repeatability of the experimental data, and proper functioning of the equipments and experimental setup. The experimental procedures are categorized into start-up procedure and measurement procedure. The start-up procedure lists several steps to initiate the experimental setup and prepare it for conducting the measurements. The measurement procedure lists several steps to acquire quality and accurate experimental data.

The start-up procedure is a process that includes several steps to prepare the experimental setup for measurements. The main function of this start-up procedure is to ensure that the experimental setup is operating properly and safely before the process of acquiring experimental data.

The first step in the start-up procedure is turning on all electrical instruments: Digital Field-Mount RFT9739 transmitters for both Coriolis flow meters, Validyne CD15 carrier demodulator for the differential pressure transducer, and the National Instruments data acquisition system. Once the data acquisition system is turned on, the LabVIEW VI program is launched and readings of all the thermocouples are monitored to ensure they are the same with the surrounding temperature. This process of checking the thermocouples' readings with the surrounding temperature provides an initial and quick indication whether the thermocouples are working properly.

The second step in the start-up procedure is checking the Speedaire air filter and the Aqua-Pure water filter. This process of checking the air and water filters is to ensure their functionality, and to maintain periodic replacement of the old filters. Old, dirty, and worn filters not only are unable to function properly but also reduce the flow capabilities of the fluids to the test section.

The third step in the start-up procedure is turning on the tap water, which served as cooling fluid for the copper coil in the air line and the ITT Standard one shell and two-tube pass heat exchanger in the water line. The tap water cooled the air and distilled water flowing into the test section such that both air and distilled water have the same inlet temperature.

The fourth step in the start-up procedure is turning on the Ingersoll-Rand T30 air compressor and the Bell & Gosset centrifugal pump. With the air compressor and centrifugal pump turned on, air and water are supplied to the test section. The final step in the start-up procedure is checking for leakage in the flow loop, and verifying the test section inclination angle. Having completed the steps in the start-up procedure, the experimental setup is ready for the measurement procedure.

The measurement procedure is a process that includes several steps to successfully acquire experimental data. The purpose of the measurement procedure is to ensure the integrity and repeatability of the experimental data, while maintaining the functionality of the experimental setup. Figure 19.5 is a flow chart illustrating the individual steps of the measurement procedure.

Before beginning the measurement procedure, the welder cables connecting the DC arc welder with the copper plates attached to the stainless steel test section is properly checked. This is a precautionary step to ensure all connections are safe and ready for the experiment. Poor condition of cables and improper connections between the DC arc welder and the test section could result in short circuit, which could potentially cause equipment damage, over-heating, or fire hazard.

The first step in the measurement procedure is adjusting the air and water flow rates. The flow of air is regulated by a needle valve, whereas the water is regulated by a gate valve. Using the Digital Field-Mount transmitters, for the Coriolis flow meters, to monitor the air and water flow rates, the needle and gate valves are adjusted until the desired air and water flow rates are achieved.

Once the desired air and water flow rates are set, the second step in the measurement procedure is to turn on the DC arc welder. With the DC arc welder turned on, the current is adjusted to the desired amperage to be supplied to the stainless steel test section. By running DC current through the stainless steel pipe, heat flux is supplied to the test section.

The third step in the measurement procedure is to allow flow to achieve steady state condition. The flow is considered to have achieved steady state condition when each of the two thermocouple probes, which measure the inlet and outlet bulk temperatures, is indicating less than 0.5°C fluctuation for 5 min. The inlet and outlet bulk temperatures

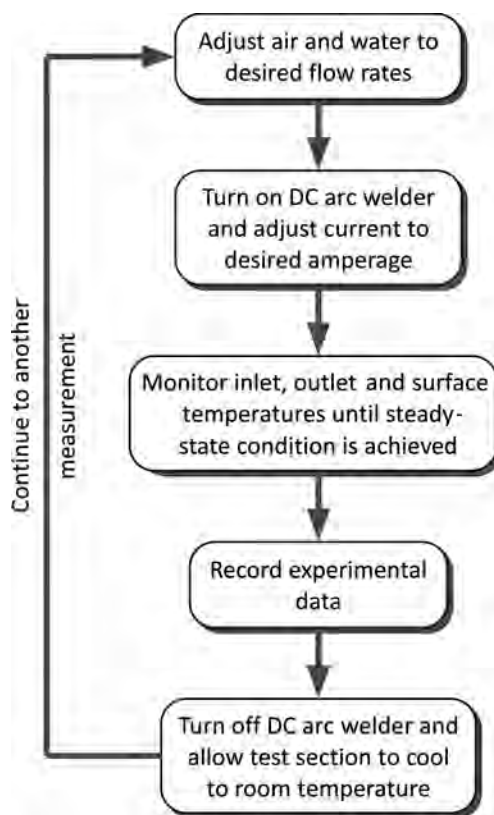


FIGURE 19.5 Measurement procedure for two-phase heat transfer experiments.

are monitored using the graphical user interface of the LabVIEW VI program illustrated in Figure 19.4. The thermocouples measuring the outer surface temperatures of the pipe are carefully monitored to avoid the temperature from rising beyond 60°C . This is to keep the flow strictly nonboiling and avoid the possibility of local boiling.

Once the steady state is achieved, the process of acquiring experimental data can begin, which is the fourth step in the measurement procedure. Experimental data collected are air and water mass flow rates, inlet and outlet bulk temperatures, surface temperatures of the test section, pressure drop, and uniform heat flux supplied to the test section through the current from the DC arc welder. The typical number of samples collected for each measurement run is 1000. When the experimental data are recorded, the following step in the measurement procedure is to turn off the DC arc welder and allow the test section to cool to room temperature. Once the test section is cooled to room temperature, another measurement run can begin.

For a uniform wall heat flux condition, the experiment involved the measurements of outside pipe wall surface temperatures at discrete locations (see Figure 19.3), as well as the inlet and outlet bulk temperatures. The circumferential heat transfer coefficient was calculated based on the knowledge of the heat flux and surface temperature at the inside wall of the pipe. Due to the difficulty of measuring the inside wall temperatures, they are instead calculated from the measured outside wall temperatures and the heat generation

within the pipe wall. The data reduction program developed by Ghajar and Kim (2006) was used for calculating the circumferential heat transfer coefficients and the inside wall temperatures from the outside wall temperatures measured at discrete locations along the uniformly heated pipe.

19.5 VERIFYING THE FUNCTIONALITY OF THE EXPERIMENTAL SETUP

The objective of conducting frictional pressure drop and heat transfer measurements for single-phase flow is to verify the functionality and reliability of the experimental setup. In this section, results of single-phase liquid flow frictional pressure drop and heat transfer measurements conducted with the experimental setup are discussed. The single-phase frictional pressure drop and heat transfer results discussed in this section is focused on horizontal pipe orientation.

19.5.1 Frictional Pressure Drop in Single-Phase Flow

The single-phase frictional pressure drop measurements were conducted using distilled water at the Reynolds number range from approximately 4,000 to 50,000. The measured frictional pressure drop was converted to Fanning friction factor using the following equation

$$c_f = \frac{2\tau_0}{\rho U^2} = \frac{\rho \pi^2 D^5}{32 \dot{m}^2} \frac{\Delta p}{L} \quad (19.1)$$

The single-phase friction factor (c_f) obtained experimentally is compared with the friction factor calculated from correlations available in the literature. Three different friction factor correlations are used in this comparison with the measured results. The first friction factor correlation is the well-cited correlation by Blasius (1913):

$$c_f = 0.079 Re^{-0.25} \quad 4,000 < Re < 10^5 \quad (19.2)$$

The second correlation used for comparison with the measured friction factor is the correlation by Petukhov (1970):

$$c_f = \frac{1}{4} (0.790 \ln Re - 1.64)^{-2} \quad 3,000 < Re < 5 \times 10^6 \quad (19.3)$$

The third friction factor correlation is proposed by Haaland (1983):

$$\frac{1}{2c_f^{1/2}} \approx -1.8 \log \left[\frac{6.9}{Re} + \left(\frac{\varepsilon/D}{3.7} \right)^{1.11} \right] \quad (19.4)$$

In using the friction factor correlation by Haaland (1983), the value for stainless steel pipe wall roughness height of 2 μm is used.

Among the three friction factor correlations selected for comparing with the measured friction factor, two of them are for turbulent flow in smooth pipe: Blasius (1913) correlation, Equation (19.2), and Petukhov (1970) correlation, Equation (19.3). The correlation

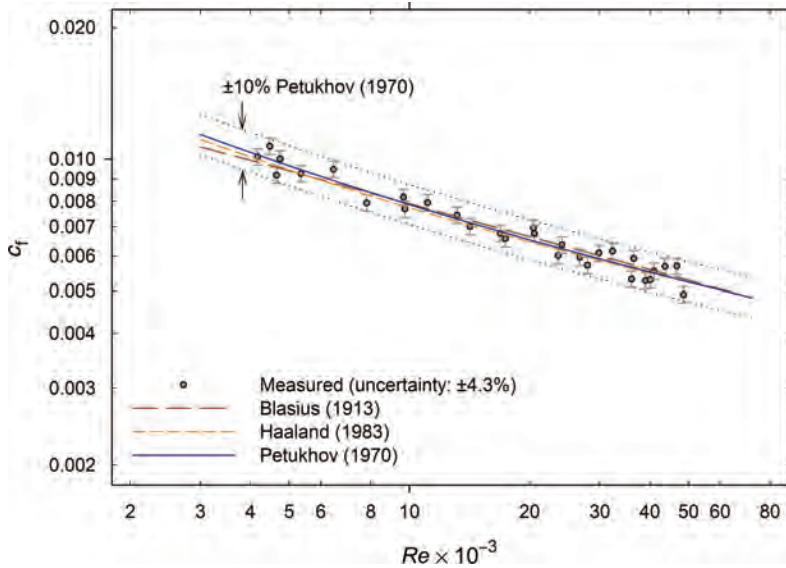


FIGURE 19.6 Comparison of measured single-phase friction factor data with correlations available in the literature.

proposed by Haaland (1983), Equation (19.4), takes into consideration the effect of wall roughness. The estimation by Haaland (1983) correlation varies by less than 2% from the Colebrook (1939) correlation (White, 1999).

The results of the comparison between the measured friction factor and the calculated friction factor are illustrated in Figure 19.6. The absolute mean deviation of the measured data from the calculated results using the Blasius (1913) correlation, Equation (19.2), is 4.7%, while the root mean square deviation is 5.5%. The comparison of the measured friction factor with the calculation using the Petukhov (1970) correlation, Equation (19.3), is shown in Figure 19.6. The measured data deviate from the calculated results using the Petukhov (1970) correlation with an absolute mean deviation and root mean square deviation of 3.8 and 4.3%, respectively. Using the Haaland (1983) correlation, Equation (19.4), to compare with the measured friction factor, the comparison also shows very satisfactory agreement. The absolute mean deviation of the measured data from the calculated results using the Haaland (1983) correlation is 4.1%, while the root mean square deviation is 4.8%. Figure 19.6 shows that most of the measured data points are well within $\pm 10\%$ agreement with the three correlations.

The satisfactory comparison of the measured friction factors with the three different friction factor correlations available in the literature verified the functionality of the experimental setup for pressure drop measurements. The highest uncertainty estimated for the measured friction factor is $\pm 4.3\%$. The uncertainties are estimated using the propagation of the uncertainty method described by Kline and McClintock (1953).

19.5.2 Heat Transfer in Single-Phase Flow

The single-phase heat transfer measurements were conducted using distilled water at the Reynolds number range from approximately 3,000 to 50,000. The single-phase heat

transfer coefficient obtained experimentally is compared with the heat transfer coefficient calculated from correlations available in the literature. Three different heat transfer correlations are used in this comparison with the measured results. The first heat transfer correlation is the Sieder and Tate (1936) correlation:

$$Nu = 0.027 Re^{4/5} Pr^{1/3} \left(\frac{\mu_b}{\mu_w} \right)^{0.14} \quad \begin{array}{l} 0.7 \leq Pr \leq 16,700 \\ Re > 10,000 \\ L/D > 10 \end{array} \quad (19.5)$$

The second correlation is the Gnielinski (1976) correlation:

$$Nu = \frac{(c_f/2)(Re - 1000)Pr}{1 + 12.7(c_f/2)^{1/2}(Pr^{2/3} - 1)} \quad \begin{array}{l} 0.5 < Pr < 2,000 \\ 3,000 < Re < 5 \times 10^6 \end{array} \quad (19.6)$$

The Fanning friction factor (c_f) in Equation (19.6) is calculated by the Petukhov (1970) correlation, Equation (19.3). The third heat transfer correlation is the Ghajar and Tam (1994) correlation:

$$Nu = 0.023 Re^{0.8} Pr^{0.385} \left(\frac{L}{D} \right)^{-0.0054} \left(\frac{\mu_b}{\mu_w} \right)^{0.14} \quad \begin{array}{l} 4 < Pr < 34 \\ 7,000 < Re < 49,000 \\ 1.1 < \mu_b/\mu_w < 1.7 \\ 3 < L/D < 192 \end{array} \quad (19.7)$$

The results of the comparison between the measured heat transfer data and the calculated heat transfer results are illustrated in Figure 19.7. When compared with the Sieder and

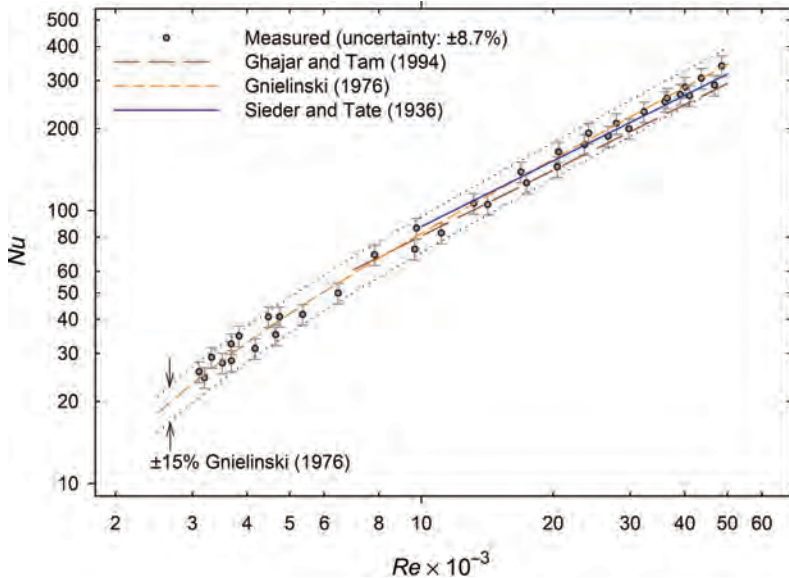


FIGURE 19.7 Comparison of measured single-phase heat transfer data with correlations available in the literature.

Tate (1936) correlation, Equation (19.5), the absolute mean deviation of the measured data from the calculated results using the correlation is 7.9%, while the root mean square deviation is 9.3%. Using the Gnielinski (1976) correlation, Equation (19.6), to compare with the measured heat transfer data, the comparison also shows very satisfactory agreement. The measured data deviate from the calculated results using the Gnielinski (1976) correlation with an absolute mean deviation and root mean square deviation of 6.5 and 6.9%, respectively. The comparison of the measured heat transfer data with the calculation using the Ghajar and Tam (1994) correlation, Equation (19.7), is shown in Figure 19.7. When compared with the Ghajar and Tam (1994) correlation, Equation (19.7), the absolute mean deviation of the measured data from the calculated results using the correlation is 6.8%, while the root mean square deviation is 8.3%. Most of the measured data points are well within $\pm 15\%$ agreement, when compared with the three heat transfer correlations, see Figure 19.7.

Among the three heat transfer correlations, the measured data agree best with the Gnielinski (1976) correlation, whereas the agreement with the Sieder and Tate (1936) and the Ghajar and Tam (1994) correlations is very satisfactory as well. Thus, verified the functionality of the experimental setup for heat transfer measurements. The highest uncertainty estimated for the measured heat transfer is $\pm 8.7\%$. The uncertainties are estimated using the propagation of the uncertainty method described by Kline and McClintock (1953).

19.6 EXPERIMENTAL RESULTS OF TWO-PHASE FLOW

19.6.1 Flow Patterns and Flow Maps

The different interpretations given to the various flow patterns by different investigators are subjected to the investigators' observations. So far, no uniform procedure exists for describing and classifying flow patterns. For the present study, the identification of flow pattern was based on the procedures suggested by Taitel and Dukler (1976), as well as Kim and Ghajar (2002). All observations of the flow pattern were conducted at the clear polycarbonate observation sections, located at the upstream and downstream of the heat transfer test section (see Figure 19.1). By systematically coordinating the flow rates of air and water, the flow patterns were observed. Flow pattern data were obtained at isothermal condition with the pipe at horizontal position and at 2° , 5° , and 7° inclined positions. Representative digital images of each flow pattern were taken using a Nikon D50 digital camera with Nikkor 50 mm f/1.8D lens. The flow map with the representative photographs of the various flow patterns for horizontal flow is shown in Figure 19.8. The various flow patterns for horizontal flow, depicted in Figure 19.8, show the capability of the current experimental setup to cover a multitude of flow patterns. The shaded lines represent the boundaries of the observed flow patterns.

The flow pattern transition boundaries for horizontal flow were found to be quite different from the flow pattern transition boundaries for inclined flow when slight inclinations of 2° , 5° , and 7° were introduced (see Figure 19.9). The changes in the flow pattern transition boundaries from horizontal to slightly inclined flow are the transition boundaries for stratified flow and slug/wavy flow. When the pipe was inclined from horizontal to slight inclination angles of 2° , 5° , and 7° , the stratified flow region was replaced by slug flow and slug/wavy flow for $Re_{SG} < 4,000$ and $4,000 < Re_{SG} < 10,000$, respectively.

Flow pattern boundaries for plug-to-slug transition and slug-to-slug/bubbly transition were observed to be shifted slightly to the upper left direction as inclination angles were

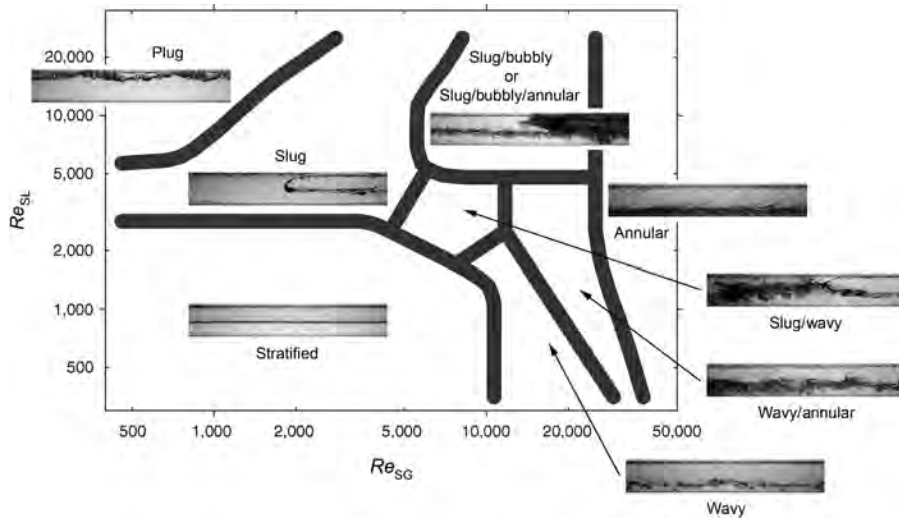


FIGURE 19.8 Flow map for horizontal pipe with photographs of representative flow patterns.

slightly increased from horizontal to 7° . For slightly inclined flow of 2° , 5° , and 7° , there were no drastic changes in the flow pattern transition boundaries. However, it should be mentioned that although the flow patterns were named the same for both horizontal and inclined flows; it does not mean that the flow patterns in the inclined positions have identical characteristics of the comparable flow patterns in the horizontal position. For

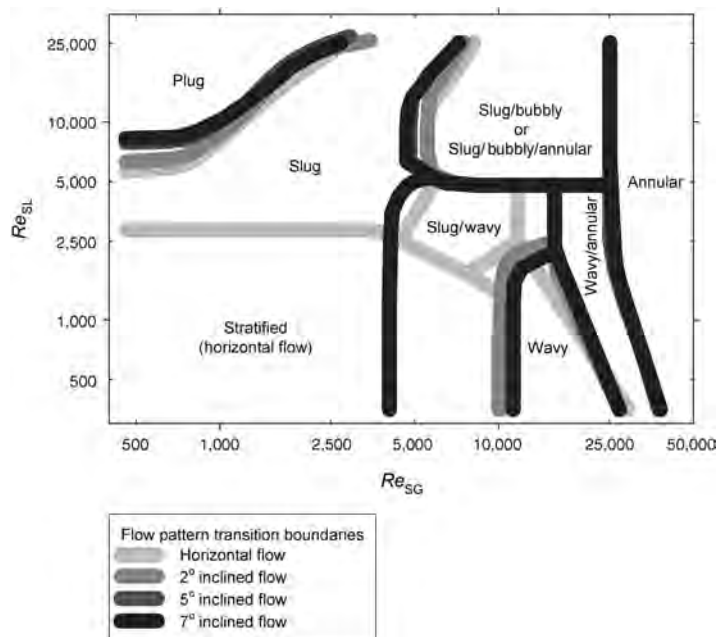


FIGURE 19.9 Change of flow pattern transition boundaries as pipe inclined upward from horizontal orientation.

example, it was observed that the slug flow patterns in the inclined positions of 5° and 7° have reverse flow between slugs due to the gravitational force, which can have a significant effect on the heat transfer.

The flow pattern maps for horizontal and 2° upward inclined pipe were verified with flow patterns data from Barnea et al. (1980). The comparison between the flow patterns data from Barnea et al. (1980) and the flow pattern maps for horizontal and 2° upward inclined flow showed very satisfactory agreement, especially among the distinctive major flow patterns such as annular, slug, and stratified (Ghajar and Tang, 2010). In a recent article, Vaze and Banerjee (2011) showed that their flow pattern data for horizontal pipe are in good agreement with the flow pattern map showed in Figure 19.8. A total of 203 flow pattern data points out of 225, observed by Vaze and Banerjee (2011), are found in the exact flow pattern regions illustrated in Figure 19.8.

19.6.2 Heat Transfer in Horizontal and Slightly Upward Inclined Pipe Flows

In this section, an overview of the different trends that have been observed in the heat transfer behavior of the two-phase air–water flow in horizontal and inclined pipes for various flow patterns is presented. Once the flow maps of the flow patterns for horizontal and inclined flows were established, the heat transfer data for different flow patterns were then collected. The nonboiling two-phase heat transfer data were obtained by systematically varying the air and water flow rates and the pipe inclination angle. Table 19.1 summarizes the two-phase heat transfer data measured for horizontal, 2° , 5° , and 7° inclined flows.

Flow pattern as well as flow rates of gas and liquid has significant influence on the two-phase heat transfer coefficient. Figure 19.10 provides an overview of the pronounced influence of the flow pattern, superficial liquid Reynolds number (water flow rate) and superficial gas Reynolds number (airflow rate) on the two-phase heat transfer coefficient in horizontal and inclined pipe flow. The results presented in Figure 19.10 clearly show that two-phase heat transfer coefficient ($h_{TP,exp}$) is strongly influenced by the superficial liquid Reynolds number (Re_{SL}) for horizontal and inclined flows.

In addition, at a constant superficial liquid Reynolds number, the two-phase heat transfer coefficient is influenced by the superficial gas Reynolds number (Re_{SG}), and each flow pattern shows its own distinguished heat transfer trend for horizontal and slightly inclined flow. Typically, heat transfer increases at low superficial gas Reynolds number (the regime of plug flow), and then slightly decreases at the mid range of Re_{SG} (the regime of slug and slug-type transitional flows), and increases again at the high Re_{SG} (the regime of annular flow).

TABLE 19.1 Summary of Experimental Conditions and Measured Two-Phase Heat Transfer Data for Horizontal and Slightly Upward Inclined Pipe

	Pipe Orientation			
	Horizontal	2° Inclined	5° Inclined	7° Inclined
No. of data points	180	184	184	187
Re_{SL} range	700–26,000	700–26,000	700–26,000	700–26,000
Re_{SG} range	700–48,000	700–48,000	700–48,000	700–48,000
Heat flux range (W/m^2)	1,860–10,800	2,820–10,800	2,900–10,800	2,600–10,900
$h_{TP,exp}$ range (W/m^2K)	101–5,457	242–5,140	286–5,507	364–5,701

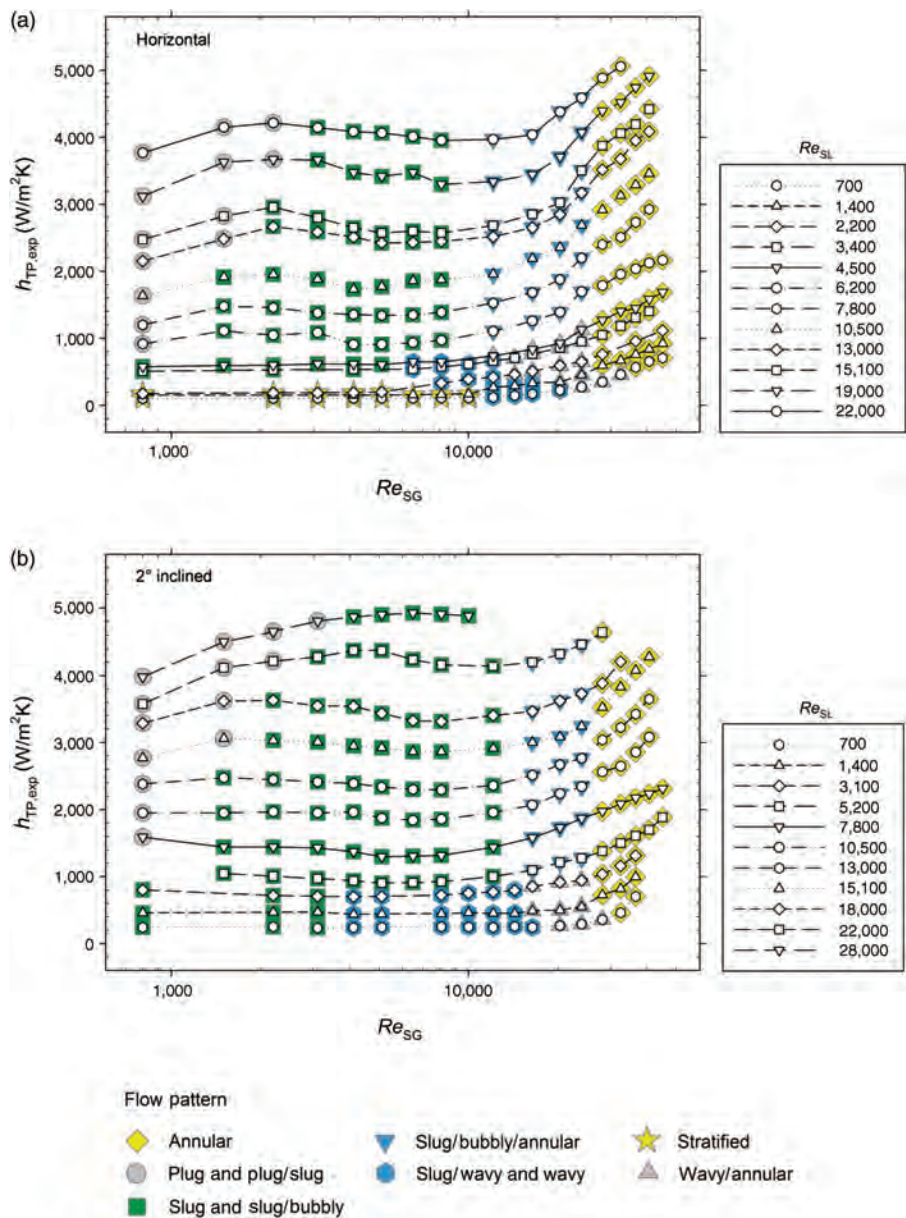


FIGURE 19.10 Variation of two-phase heat transfer coefficient with superficial gas and liquid Reynolds numbers.

Heat transfer in two-phase flow is also influenced by the orientation of the pipe. Heat transfer coefficients measured for slightly upward inclined flow showed some enhancement when compared with data measured for horizontal flow. In order to observe the behavior of heat transfer due to pipe inclination, it is necessary to categorize the experimental heat transfer data into comparable superficial gas and liquid Reynolds numbers.

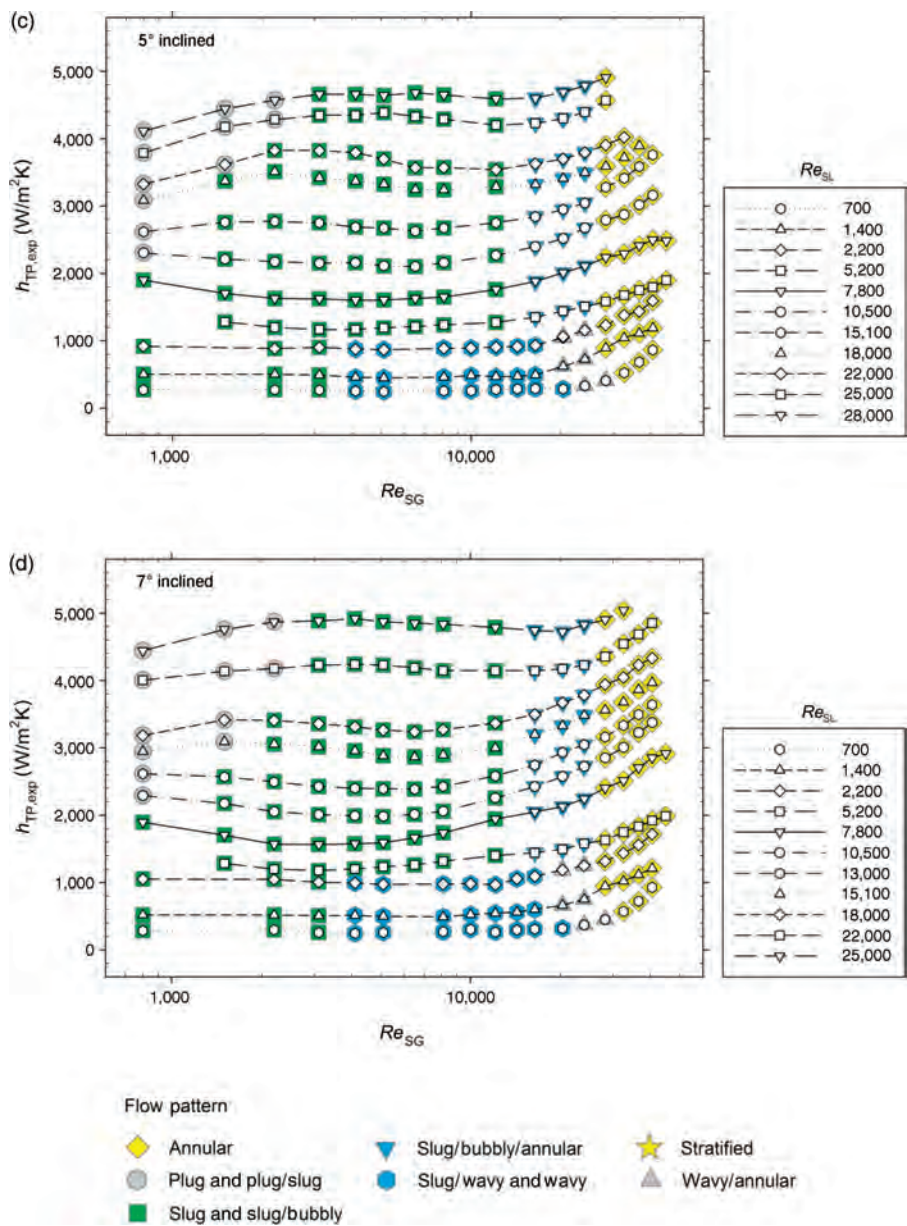


FIGURE 19.10 (Continued)

Figure 19.11 illustrates the influence of pipe inclination on two-phase heat transfer coefficient for varying superficial gas and liquid Reynolds numbers. In Figure 19.11, the ratios of two-phase heat transfer coefficient for inclined pipe to two-phase heat transfer coefficient for horizontal pipe ($h_{TP,\theta}/h_{TP,0}$) versus the superficial liquid Reynolds numbers (Re_{SL}) for various superficial gas Reynolds numbers (Re_{SG}) are plotted.

For low superficial liquid Reynolds number ($Re_{SL} = 1,400$), pipe inclination has a significant influence on the two-phase heat transfer coefficient. In the region of $Re_{SL} < 2,500$

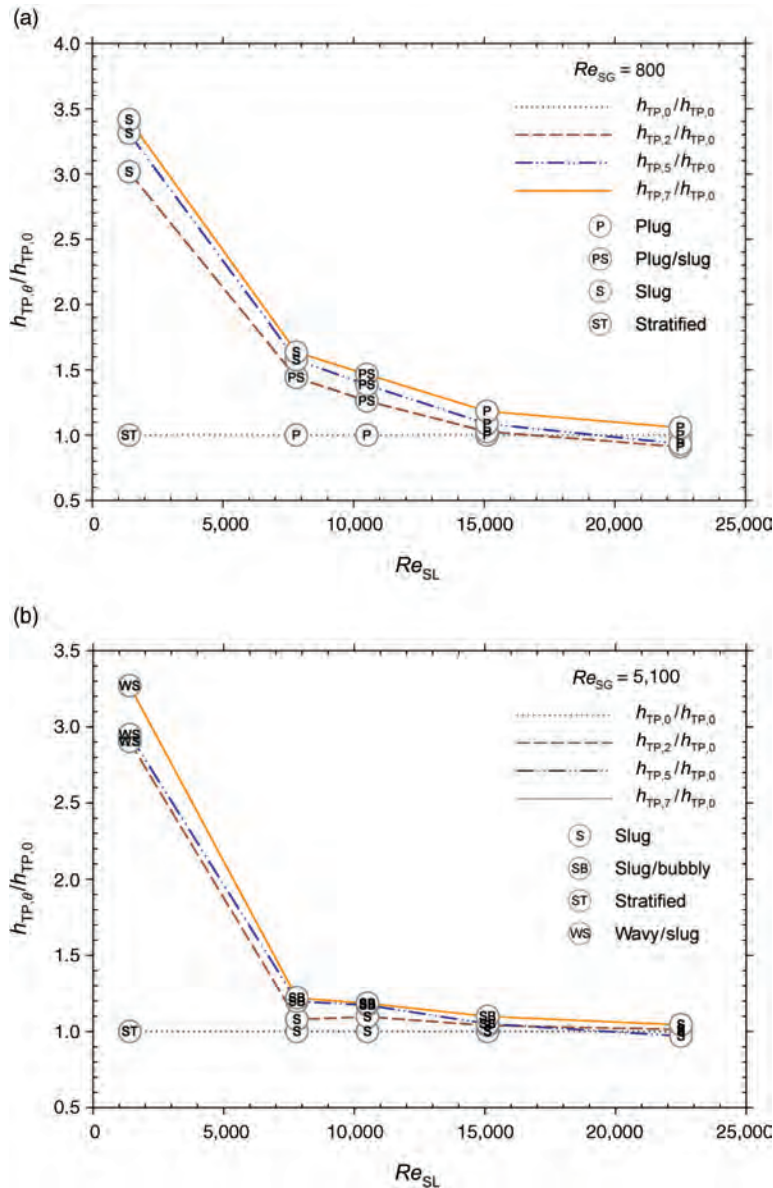


FIGURE 19.11 Influence of pipe inclination on two-phase heat transfer coefficient for varying superficial gas and liquid Reynolds numbers.

and $Re_{SG} < 10,000$, pipe inclined from horizontal position to 2° , 5° , and 7° can experience as much as 250% enhancement in two-phase heat transfer coefficient. The main reason that can be attributed to this enhancement in heat transfer coefficient is the change in flow pattern from stratified for horizontal flow to slug or wavy/slug for inclined flow.

For instance, at the respective superficial gas and liquid Reynolds numbers of 800 and 1,400, inclination of pipe from horizontal to 2° , 5° , and 7° resulted in enhancements of

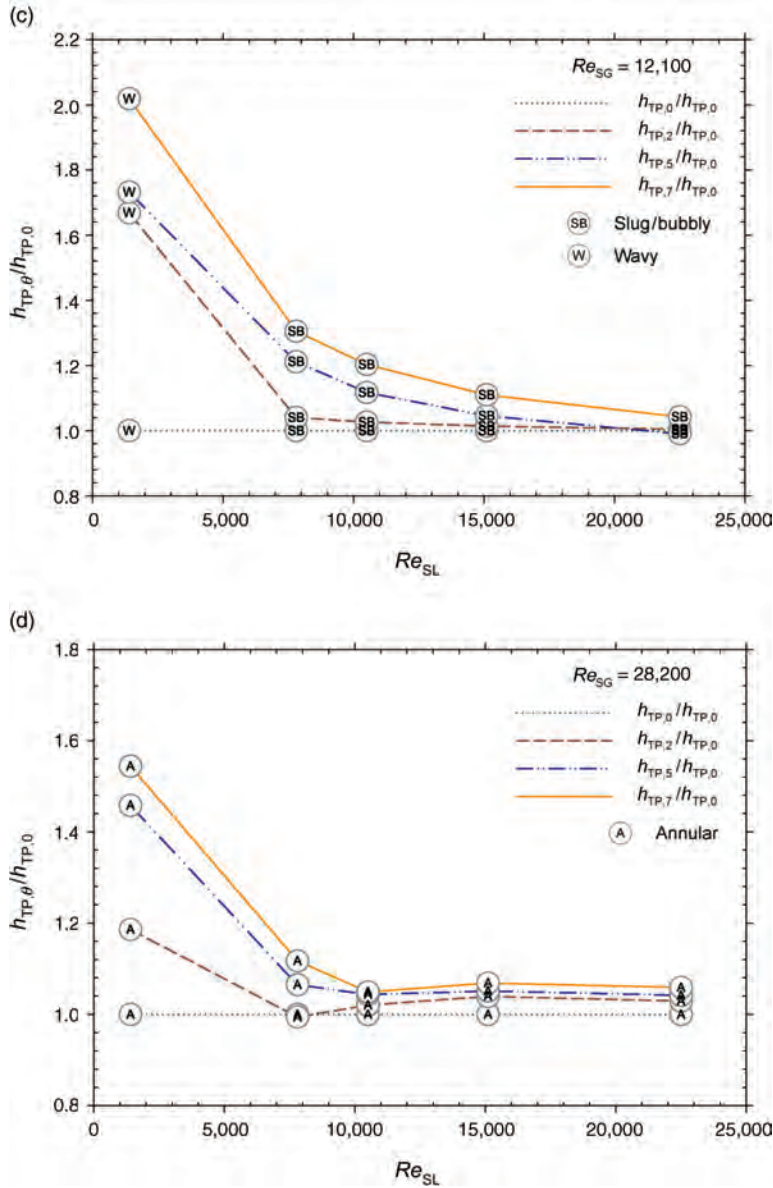


FIGURE 19.11 (Continued)

two-phase heat transfer coefficient by 200, 230, and 240%, respectively (see Figure 19.11a). In this particular case, the inclination of pipe from horizontal to 2°, 5°, and 7° caused a change in the flow pattern from stratified to slug. At the respective superficial gas and liquid Reynolds numbers of 5,100 and 1,400, inclination of pipe from horizontal to 2°, 5°, and 7° resulted in enhancements of two-phase heat transfer coefficient by 190, 195, and 230%, respectively (see Figure 19.11b). In this particular case, the inclination of pipe from horizontal to 2°, 5°, and 7° caused a change in the flow pattern from stratified to wavy/slug.

For wavy flow, an increase in inclination angle caused enhancement in the two-phase heat transfer coefficient. For instance, at the respective superficial gas and liquid Reynolds numbers of 12,100 and 1,400, Figure 19.11c shows that there are enhancements in two-phase heat transfer coefficient by 67, 73, and 102% for pipe inclinations from horizontal to 2°, 5°, and 7°, respectively. From flow pattern visual observations, increase in inclination angle for wavy flow caused the liquid backwash to become more significant, adding to the wave height and causing larger and more frequent splashes on the top part of the inner pipe surface, thus increasing the two-phase heat transfer coefficient.

For annular flow with superficial liquid Reynolds number below 7,000, pipe inclination brings enhancement to two-phase heat transfer coefficient. For instance, at the respective superficial gas and liquid Reynolds numbers of 28,200 and 1,400, Figure 19.11d shows that there are enhancements in two-phase heat transfer coefficient by 19, 46, and 54% for pipe inclinations from horizontal to 2°, 5°, and 7°, respectively. Visual observations revealed that the increase in inclination angle for annular flow caused the liquid film of the annular flow to thicken due to the effects of gravity. The thicker annular liquid film enhanced heat transfer capability, thus increasing the two-phase heat transfer coefficient.

Another perspective in the observation of heat transfer enhancement due to pipe inclination in nonboiling two-phase flow is by focusing on the two-phase heat transfer coefficient measured at the top part of the pipe (see Figure 19.3 for circumferential location “A”). Figure 19.12 shows the variation of the two-phase heat transfer coefficient at the top part of the pipe ($h_{TP,top}$) with superficial gas Reynolds number (Re_{SG}) for different pipe inclinations. The heat transfer coefficients at the top part of the pipe for an inclined pipe are consistently higher than those for a horizontal pipe. Figure 19.12 also shows that subsequent increase in pipe inclination from 2° to 7° caused increase in heat transfer coefficient.

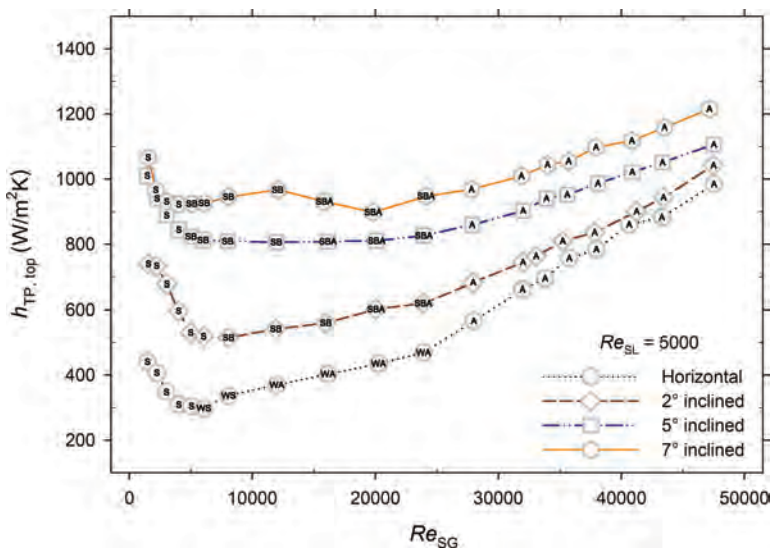


FIGURE 19.12 Variation of the two-phase heat transfer coefficient ($h_{TP,top}$) with the superficial gas Reynolds number for $Re_{SL} = 5,000$ (A = annular, S = slug, SB = slug/bubbly, SBA = slug/bubbly/annular, WA = wavy/annular, WS = wavy/slug).

For inclined pipes, the higher heat transfer coefficients at the top part of the pipe contributed to the overall heat transfer enhancement. The higher $h_{TP,top}$ suggests that the top part of the pipe experienced higher level of contact with the liquid phase for an inclined pipe than for a horizontal pipe. Depending on the flow pattern, the increase in the level of liquid contact with the top part of the pipe can be in terms of liquid film thickness or turbulence and speed of traversing slug.

Figure 19.13 shows the heat transfer coefficient of two-phase flow at different circumferential locations of the pipe. For inclined pipes, Figure 19.13 shows that the heat transfer coefficients are higher than those for horizontal pipe at all circumferential locations.

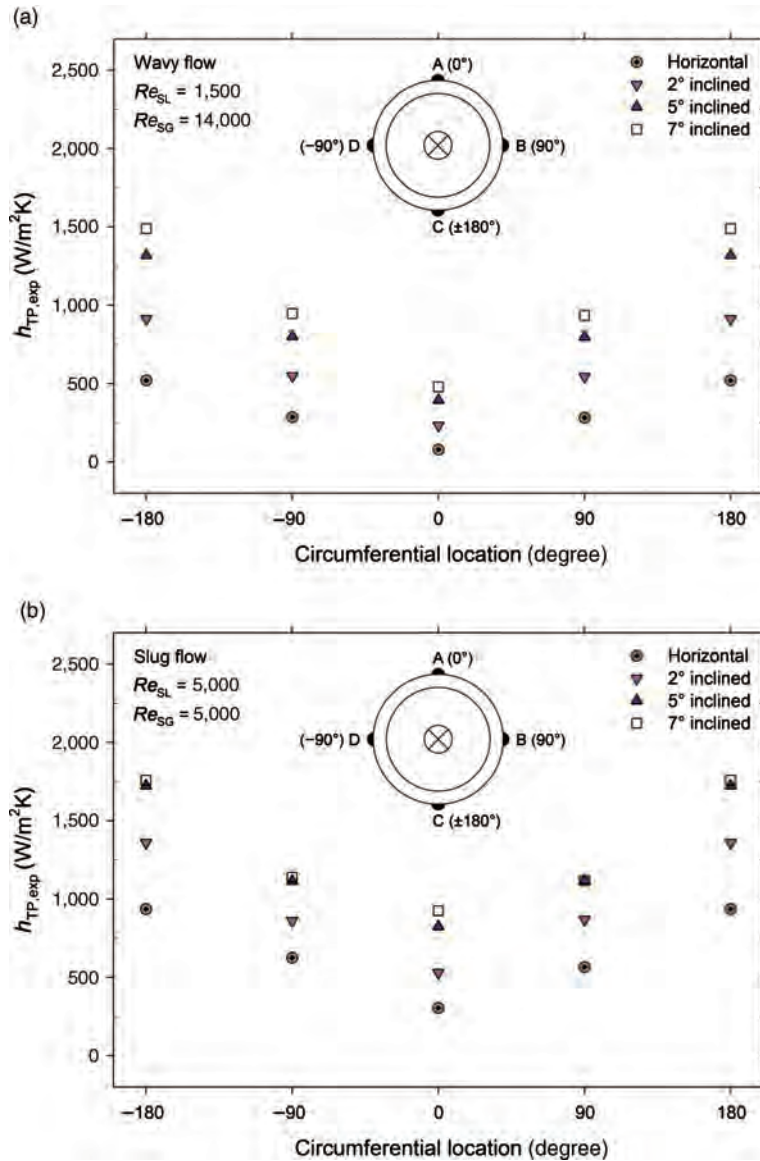


FIGURE 19.13 Two-phase heat transfer coefficient at circumferential locations of the pipe.

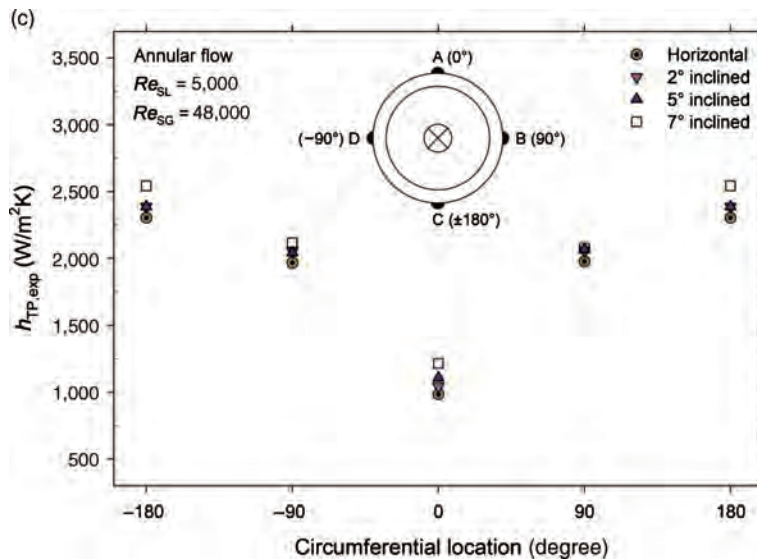


FIGURE 19.13 (Continued)

For annular flow (Figure 19.13c), the differences in the circumferential heat transfer coefficient for various pipe inclinations are less significant than those for wavy and slug flows (Figure 19.13a and b). The inertia force in annular flow is higher than that in wavy and slug flows. With inertia force dominant over buoyancy force, the influence of pipe inclination on heat transfer coefficient diminishes.

19.7 CONCLUDING REMARKS

The characterization and measurement of heat transfer in two-phase flow in pipe with different inclinations require a functional experimental setup as well as proper experimental procedures. Heat transfer in single-phase pipe flow generally produces uniform heat transfer coefficient circumferentially at an axial location. Unlike single-phase flow, two-phase flow in pipe is influenced by forces such as buoyancy, inertia, viscosity, and surface tension, which result in various flow patterns. The existence of flow patterns in two-phase flow leads to the occurrence of nonuniform heat transfer coefficient circumferentially at an axial location. The positions of the thermocouples on the heat transfer section, with four thermocouples positioned at equal interval circumferentially for each station, are employed to properly measure heat transfer coefficient of two-phase pipe flow.

The results of the present study had prompted the need of further investigation for nonboiling two-phase flow in pipe inclinations not addressed presently. A more comprehensive experimental study of two-phase flow heat transfer requires a new robust experimental setup that is equipped for measuring heat transfer, pressure drop, void fraction, and also conducting flow visualization of nonboiling two-phase flow for all major flow patterns and inclination angles from 0° (horizontal) to $\pm 90^\circ$ (upward and downward vertical). Details of the design and construction of this experimental setup as well as the validation of its functionality were documented by Cook (2008). With the new

experimental setup, the influences of void fraction, pressure drop, and flow patterns on nonboiling heat transfer can be evaluated. Void fraction plays an important role in the modeling of two-phase pressure drop, flow pattern transition, and heat transfer. Since void fraction is a very important parameter in characterizing two-phase flow, systematic measurements of void fraction for all pipe inclinations are important for establishing the fundamental understanding of two-phase flow. At the time of this writing, experiments on measuring void fraction for air–water flow in horizontal, upward and downward vertical pipes have been conducted. The results were documented by Ghajar and Tang (2010), Godbole et al. (2011), and Bhagwat and Ghajar (2012).

The experimental results that can be generated from the new experimental setup would provide further understanding of nonboiling two-phase flow and heat transfer in pipes. Experimental results for flow patterns, heat transfer, pressure drop, and void fraction, which cover the entire range of pipe inclinations, will be a significant milestone in the road map of establishing a fundamental understanding for two-phase flow.

NOMENCLATURE

c_f	Fanning friction factor, $c_f = 2\tau_0/(\rho U^2)$, dimensionless
c_p	Specific heat, J/kgK
D	Pipe inner diameter, m
h	Heat transfer coefficient, W/m ² K
$h_{TP,exp}$	Measured two-phase heat transfer coefficient, W/m ² K
$h_{TP,top}$	Two-phase heat transfer coefficient at the top part of the pipe, W/m ² K
$h_{TP,0}$	Two-phase heat transfer coefficient for horizontal pipe, 0°, W/m ² K
$h_{TP,\theta}$	Two-phase heat transfer coefficient for a specific inclination angle, θ , W/m ² K
k	Thermal conductivity, W/mK
L	Length, m
\dot{m}	Mass flow rate, kg/s
Nu	Nusselt number, $Nu = hD/k$, dimensionless
Pr	Prandtl number, $Pr = c_p\mu/k$, dimensionless
Δp	Pressure difference, N/m ²
Re	Reynolds number, $Re = \rho UD/\mu$, dimensionless
Re_{SG}	Superficial gas Reynolds number, $Re_{SG} = \rho_G U_{SG} D/\mu_G$, dimensionless
Re_{SL}	Superficial liquid Reynolds number, $Re_{SL} = \rho_L U_{SL} D/\mu_L$, dimensionless
U	Average velocity, m/s
U_{SG}	Superficial gas velocity, m/s
U_{SL}	Superficial liquid velocity, m/s

Greek Symbols

ε	Pipe wall roughness height, m
θ	Pipe inclination angle, degree
μ	Dynamic viscosity, Ns/m ²

μ_b	Dynamic viscosity evaluated at bulk temperature, Ns/m^2
μ_G	Gas phase dynamic viscosity, Ns/m^2
μ_L	Liquid phase dynamic viscosity, Ns/m^2
μ_w	Dynamic viscosity evaluated at wall temperature, Ns/m^2
ρ	Density, kg/m^3
ρ_G	Gas phase density, kg/m^3
ρ_L	Liquid phase density, kg/m^3
τ_0	Wall shear stress, N/m^2

REFERENCES

- Barnea D, Shoham O, Taitel Y, Dukler AE. Flow pattern transition for gas-liquid flow in horizontal and inclined pipes: comparison of experimental data with theory. *International Journal of Multiphase Flow* 1980;6:217–225.
- Bhagwat SM, Ghajar AJ. Similarities and differences in the flow patterns and void fraction in vertical upward and downward two phase flow. *Experimental Thermal and Fluid Science* 2012;39:213–227.
- Blasius H. Das Ähnlichkeitsgesetz bei Reibungsvorgängen in Flüssigkeiten. *Forschungsarbeiten des VDI* 1913; no. 134.
- Celata GP, Chiaradia A, Cumo M, D'Annibale F. Heat transfer enhancement by air injection in upward heated mixed-convection flow of water. *International Journal of Multiphase Flow* 1999;25:1033–1052.
- Colebrook CF. Turbulent flow in pipes, with particular reference to the transition region between the smooth and rough pipe laws. *Journal of the Institution of Civil Engineers (London)* 1939; 11(4):133–156.
- Cook WL. *An experimental apparatus for measurement of pressure drop, void fraction and non-boiling two-phase heat transfer and flow visualization in pipes for all inclinations [M.S. thesis]*. Stillwater (OK): Oklahoma State University; 2008.
- Ewing ME, Weinandy JJ, Christensen RN. Observations of two-phase flow patterns in a horizontal circular channel. *Heat Transfer Engineering* 1999;20(1):9–14.
- Furuholt EM. Multiphase technology: is it of interest for future field developments? Society of Petroleum Engineers European Petroleum Conference. Paper No. 18361. October 17–19; London, England; 1988.
- Ghajar AJ, Kim J. Calculation of local inside-wall convective heat transfer parameters from measurements of local outside-wall temperatures along an electrically heated circular tube. In: Kutz M, editors. *Heat Transfer Calculations*. New York: McGraw-Hill; 2006. p. 23.3–23.27.
- Ghajar AJ, Tam LM. Heat transfer measurements and correlations in the transition region for a circular tube with three different inlet configurations. *Experimental and Thermal Fluid Science* 1994;8(1):79–90.
- Ghajar AJ, Tang CC. Importance of non-boiling two-phase flow heat transfer in pipes for industrial applications. *Heat Transfer Engineering* 2010;31(9):711–732.
- Godbole PV, Tang CC, Ghajar AJ. Comparison of void fraction correlations for different flow patterns in upward vertical two-phase flow. *Heat Transfer Engineering* 2011;32(10):843–860.
- Gnielinski V. New equations for heat and mass transfer in turbulent pipe and channel flow. *International Chemical Engineering* 1976;16:359–368.
- Haaland SE. Simple and explicit formulas for the friction factor in turbulent pipe flow. *Journal of Fluids Engineering, Transactions of the ASME* 1983;105(1):89–90.

- Kim D. *An experimental and empirical investigation of convective heat transfer for gas-liquid two-phase flow in vertical and horizontal pipes [Ph.D. thesis]*. Stillwater (OK): Oklahoma State University; 2000.
- Kim D, Ghajar AJ. Heat transfer measurements and correlations for air-water flow of different flow patterns in a horizontal pipe. *Experimental Thermal and Fluid Science* 2002;25:659–676.
- Kline SJ, McClintock FA. Describing uncertainties in single-sample experiments. *Mechanical Engineering* 1953;1:3–8.
- McClafflin GG, Whitfill DL. Control of paraffin deposition in production operations. *Journal of Petroleum Technology* 1984;36(12):1965–1970.
- Petukhov BS. Heat transfer and friction in turbulent pipe flow with variable physical properties. *Advances in Heat Transfer* 1970;6:503–564.
- Shiu KC, Beggs HD. Predicting temperatures in flowing oil wells. *Journal of Energy Resources Technology* 1980;102(1):2–11.
- Sieder EN, Tate GE. Heat transfer and pressure drop of liquids in tubes. *Industrial & Engineering Chemistry* 1936;28(12):1429–1435.
- Singh P, Venkatesan R, Fogler HS, Nagarajan N. Formation and aging of incipient thin film wax-oil gels. *AIChE Journal* 2000;46(5):1059–1074.
- Taitel Y, Dukler AE. A model for predicting flow regime transitions in horizontal and near horizontal gas-liquid flow. *AIChE Journal* 1976;22(1):47–55.
- Trevisan OV, Franca FA, Lisboa AC, Trevisan OV, Franca FA, Lisboa AC. Oil production in offshore fields: an overview of the Brazilian technology development program. Proceedings of the 1st World Heavy Oil Conference. Beijing, China, Paper No. 2006-437, November 13–15, 2006.
- Vaze MJ, Banerjee J. Experimental visualization of two-phase flow patterns and transition from stratified to slug flow. *Proceedings of the Institution of Mechanical Engineers, Part C: Journal of Mechanical Engineering Science* 2011;225(2):382–389.
- White FM. *Fluid Mechanics*, 4th ed. New York: McGraw-Hill; 1999.

20

SOLAR ENERGY MEASUREMENTS

TARIQ MUNEEER and YIENG WEI THAM

- 20.1 Introduction
 - 20.1.1 Radiation measurement history
 - 20.1.2 Solid angle
 - 20.1.3 Intensity and flux
 - 20.1.4 Radiation units
 - 20.1.5 Radiation laws
 - 20.1.6 Thermal detector
 - 20.1.7 Photo detectors
 - 20.1.8 Classification of pyrheliometer
 - 20.1.9 Working standard
- 20.2 Measurement equipment
 - 20.2.1 Pyranometer
 - 20.2.2 Pyranometer with shading device
 - 20.2.3 Pyrheliometer
 - 20.2.4 Pyrgeometer
 - 20.2.5 Albedometer
 - 20.2.6 Sunshine recorder
- 20.3 Equipment error and uncertainty
- 20.4 Operational errors
- 20.5 Diffuse radiation data measurement errors
 - 20.5.1 Description of models
- 20.6 Types of sensors and their accuracy
- 20.7 Modern developments
- 20.8 Data quality assessment
 - 20.8.1 United state national renewable energy laboratory (NREL)
 - 20.8.2 Commision internationale de l'éclairage (CIE) automatic quality control
 - 20.8.3 Page model
 - 20.8.4 Muneer and Fairouz quality control procedure
- 20.9 Statistical evaluation of models
 - 20.9.1 Slope of the best-fit line, s
 - 20.9.2 Coefficient of determination, r^2
 - 20.9.3 Coefficient of correlation, r
 - 20.9.4 Student's t -distribution

- 20.9.5 Root mean squared error, RMSE
- 20.9.6 Mean bias error, MBE
- 20.9.7 Mean of absolute deviations, MAD
- 20.9.8 Nondimensional MBE, MAD, and RMSE
- 20.9.9 Figure of merit, ψ
- 20.10 Outlier analysis
- Acknowledgments
- References

20.1 INTRODUCTION

20.1.1 Radiation Measurement History

The Basra born Physicist Abu Ali al-Hasan ibn al-Hasan ibn al-Haytham was one of the first scientists who made a systematic attempt at understanding the transmission of solar radiation through terrestrial atmosphere. Al-Haytham was born in the year 965 and in his treatise “Balance of Wisdom,” he discusses the density of the atmosphere and the phenomenon of atmospheric refraction. He also attempted to measure the height of the atmosphere, which has now been confirmed to be 99% accurate (Rozenberg, 1966).

The invention of telescope by Galileo in the 16th century might have been the time when the sun was investigated with any reported significance. Since then, great many discoveries have been made and physical laws formulated and explained such as radiation electromagnetic theory, color of sunlit sky, solar radiation absorption by water vapor, and wave theory of light.

The electrical compensation pyrheliometer invented by Knut Ångström in 1893 was among the earliest radiation measurement instruments. This instrument is still used in many countries as the standard for absolute radiant energy determination (Coulson, 1975).

The advancement of measurement instrument development started after WWII when technology was furthered. The replacement of conventional resistance strips in Ångström-type pyrheliometer with thermopile is one of the many examples. Furthermore, the use of computerized data logging and data acquisition has replaced the conventional manual, time-consuming methods.

The World Meteorological Organization (WMO) has recommended a standard terminology for various radiation fluxes and the classification of the associated instruments. The units have been chosen on the basis of standard use and ease of physical interpretation. Some of the relevant basic concepts are explained in the following sections.

20.1.2 Solid Angle

Refer to Figure 20.1 wherein an elemental solid angle, $d\omega$, is shown. The concept of solid angle can be illustrated as a straight line through point 0 moving in space and intersecting an arbitrary surface located at some distance s from point 0. If the locus of the point of intersection forms a closed path on the surface but does not intersect itself, then a unique area is defined on the surface. Assume the area is an elemental area, da , the surface normal of which makes an angle with the direction to point 0.

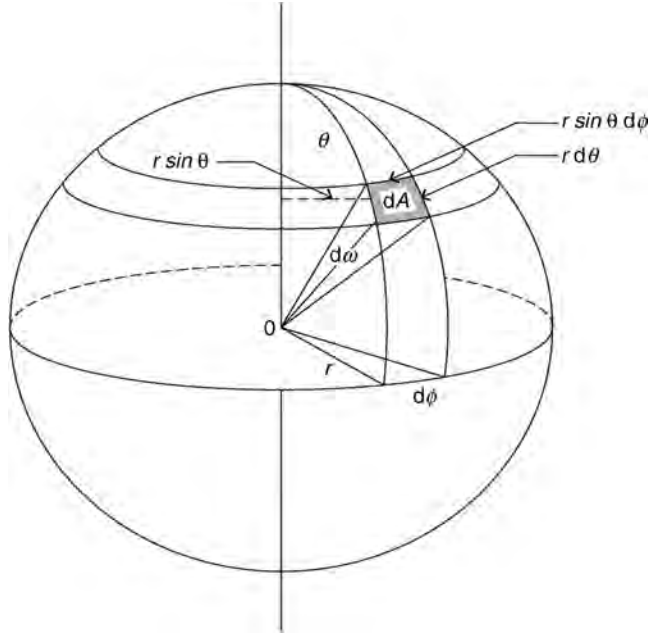


FIGURE 20.1 Illustration of solid angle and its interpretation in spherical coordinates.

For purpose of illustration, assume it is the surface of a sphere of radius, r , as shown in Figure 20.1. For this case, the solid angle subtended at the center of the sphere by area dA on its surface is $d\omega = dA/r^2$. For the special case of a unit sphere ($r = 1$), $d\omega$ and dA have the same numerical value if $d\omega$ is expressed in steradians (sr) and dA and r are expressed in the same system of units. Because the area of the surface of a sphere is $4\pi r^2$, the total solid angle subtended at a point by the entire surrounding sphere is $4\pi r^2/r^2 = 4\pi$ sr.

A hemispheric solid angle is 2π sr. As can be seen from Figure 20.1, an elemental solid angle is conveniently expressed in a spherical coordinate system as

$$d\omega = \frac{(r d\theta)(r \sin\theta d\phi)}{r^2} \quad (20.1)$$

20.1.3 Intensity and Flux

Consider a pencil of radiation crossing the elemental area $d\sigma$ of Figure 20.2 and confined to the elemental angle $d\omega$, which is oriented at some angle θ to the normal of $d\sigma$. The energy dE_v contained in the frequency interval $d\nu$, which crosses $d\sigma$ in the time increment dt is given by

$$dE_v = I_v d\nu dt d\omega d\sigma \cos\theta \quad (20.2)$$

This relation defines the monochromatic specific intensity in the most general way as

$$I_v = \frac{dE_v}{d\nu dt d\omega d\sigma \cos\theta} \quad (20.3)$$

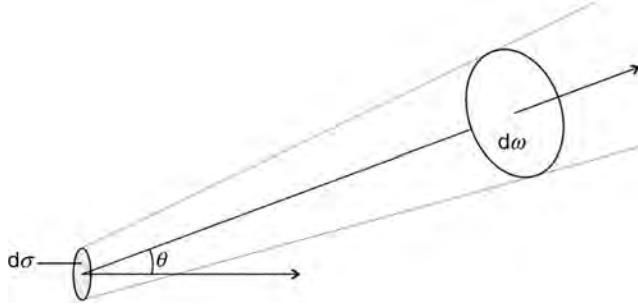


FIGURE 20.2 Beam of radiation passing through an elemental area.

Thus, the definition of specific intensity implies directionality. The term flux, however, is flow of energy, and it may or may not have an implied direction. For instance, the monochromatic flux of energy across $d\sigma$ is given by the integration of the normal component of I_v , over the entire spherical solid angle Ω . Thus,

$$F_v = dv dt d\sigma \int_{\Omega} I_v(\omega) \cos \theta d\omega \quad (20.4)$$

or, in terms of spherical coordinates,

$$F_v = dv dt d\sigma \int_0^{2\pi} \int_0^{\pi} I_v(\theta, \phi) \sin \theta \cos \theta d\theta d\phi \quad (20.5)$$

20.1.3.1 Terminology in Radiation Instruments Considerable confusion in the terminology applied to various types of solar radiation instruments has built up over the years. For instance, an instrument measuring the total hemispherical flux of solar radiation has been variously called pyrheliometer, pyranometer, solarimeter, actinograph, and sunshine recorder.

To standardize terminology, the WMO, through its Commission for Instruments and Method of Observation, has recommended the following classification of radiation instruments:

Pyrheliometer: An instrument for measuring “the intensity of direct solar radiation at normal incidence.”

Pyranometer: An instrument for measuring “the solar radiation received from the whole hemisphere. It is suitable for the measurement of the global or sky radiation.”

A pyranometer for measuring the radiation on a spherical surface is a “spherical pyranometer.”

Pyrgeometer: An instrument for measuring “the net atmospheric radiation on a horizontal upward-facing black surface at ambient air temperature.”

Pyrradiometer: An instrument for measuring “both the solar and terrestrial radiation (total radiation).”

Net Pyrradiometer: An instrument for measuring “the net flux downward and upward total (solar, terrestrial, surface, and atmospheric) radiation through a horizontal surface.” A net radiometer is sometimes termed a balance pyrradiometer or radiation balance meter.

20.1.4 Radiation Units

The principal units of radiation are given in Table 20.1.

20.1.5 Radiation Laws

1. *Planck's Law:* Max Planck showed that the spectral density emitted by a blackbody at temperature T is given by the Planck function to the following, the first being expressed on the basis of frequency ν and the second on the basis of wavelength λ of the radiation:

$$B_{\nu}(T) = \frac{2h\nu^3}{c^2(e^{h\nu/kT} - 1)} \quad (20.6)$$

$$B_{\lambda}(T) = \frac{C_1\lambda^{-5}}{(e^{C_2/\lambda T} - 1)} \quad (20.7)$$

The values of the constants are

h = Planck's constant = 6.6256×10^{-27} erg s

k = Boltzmann's constants = 1.3805×10^{-16} erg deg $^{-1}$

c = speed of light *in vacuo* = 2.998×10^{10} cm s $^{-1}$

C_1 = first radiation constant = $2hc^2 = 3.74150 \times 10^{-5}$ erg cm 2 s $^{-1}$

C_2 = second radiation constant = $hc/k = 1.43879$ cm deg.

2. *Kirchhoff's Law:* Gustav Kirchhoff stated that for a medium in thermodynamic equilibrium, the ratio between the mass emission coefficient and mass absorption coefficient has a value, which is independent of the nature of the material and is dependent only on wavelength of the radiation and temperature of the medium.

A note on the limits of applicability of Kirchhoff's law to the terrestrial radiation field of the atmosphere is appropriate. The law requires thermodynamic

TABLE 20.1 The Principal Units of Radiation

Quantities	Units
Wavelength	Micrometer (μm) or angstrom (\AA)
Frequency (ν)	s $^{-1}$
Wave number	cm $^{-1}$
Specific intensity: spectral	W/cm 2 /sr/cm
Specific intensity: total	W/cm 2 /sr
Radian flux: spectral	W/cm 2 / μm
Radian flux: total	W/cm 2

equilibrium, one characteristic of which is that the radiation field be isotropic. Obviously, the radiation field for the atmosphere as a whole is not isotropic. On the other hand, the field in the localized volume of the troposphere or stratosphere is approximately isotropic, and it is in the context of this local thermodynamic equilibrium that Kirchhoff's law is applicable to the atmosphere. A second characteristic of local thermodynamic equilibrium is that the populations of atomic and molecular states be those of their equilibrium distribution. In such a case, the energy transitions are controlled by molecular collisions and not by interaction of particles with the radiation field itself. In the atmosphere, molecular collisions dominate the energy transitions at all altitudes below 60–70 km, indicating that local thermodynamic equilibrium is a good approximation through more than 90% of the mass of the atmosphere.

3. *Stefan–Boltzmann Law*: By integrating the monochromatic blackbody function of Equation (20.7) over the entire wavelength range from 0 to ∞ , an expression for the total rate of energy emission E by a unit area of blackbody surface in terms of its absolute temperature T may be derived as,

$$E = \sigma T^4 \quad (20.8)$$

where σ is the Stefan–Boltzmann constant which is $5.6704 \times 10^{-8} \text{ W/m}^2/\text{K}^4$.

4. *Wien Displacement Law*: Wien's displacement law is obtained by differentiating the Planck function with respect to wavelength, setting the result equal to 0 and thereby determining the value λ_{\max} at which $B_{\lambda}(T)$ is a maximum. Hence,

$$\lambda_{\max} T = \text{const} \quad (20.9)$$

The value of the constant is 0.2897 cm deg if λ is in centimeter.

Important data related to solar radiation measurement are provided in Table 20.2.

20.1.6 Thermal Detector

The mode of operation of thermal detector starts when the radiant energy transfers into heat energy, with consequent rise of temperature of some material. They respond only to total energy absorbed, and thus, at least theoretically, nonselective as to the spectral distribution of the energy. Because of the limitation of the absorbing materials, this non-selective feature is difficult to achieve completely in operation, but it is more closely achieved in thermal detectors than in any other type. The main types of thermal detectors are calorimeters, thermocouples or thermopiles, and bolometer.

20.1.7 Photo Detectors

The advantage of photo detectors is that the sensor is activated by discrete events of photons striking the material and not by a change in temperature because of the absorption of radiation, as in the thermal detectors. The advantage of sensing discrete events is that much faster responses and higher sensitivities can be achieved than with thermal devices. There are three principal types of photo detectors, namely, photovoltaic, photoconductive, and photo-emissive cells.

TABLE 20.2 Highlights of Solar Radiation Instruments Development

Year	Events
1825	Herschel's pyrheliometer (actinometer) developed
1837	Invention of Pouillet's pyrheliometer
1840	Invention of first photographic sunshine recorder by Jordan
1879	Campbell–Stokes sunshine recorder started in used
1885	Invention of photographic sunshine recorder by Jordan
1885	Invention of new form polarimeter and of pole-star recorder by Pickering
1886	Invention of new type of radiometer by Ångström
1891	Concept of Maring–Marvin sunshine recorder by Maring
1893	Invention of electric compensation pyrheliometer by Ångström
1898	Invention of Callender pyranometer
1902	Secondary standard silver-disk pyrheliometer developed by Abbot
1905	Ångström electrical compensation pyrheliometer adopted as a WMO standard
1908	Invention of Michelson pyrheliometer
1910	Invention of Marvin pyrheliometer
1915	First bolometer constructed
1916	Pyranometer for measuring global radiation devised.
1917	Ultrasensitive vacuum bolometer constructed
1919	Honeycomb (melikeron) pyranometer constructed
1922	Dorno's pyrheliograph invented
1923	Invention of Kimball–Hobbs pyranometer (forerunner of first Eppley pyranometers)
1923	Invention of Moll thermopile
1924	Moll thermopile used by Gorczynski, later known as Kipp solarimeter
1927	Double water-flow pyrheliometer developed by Shulgin
1932	First standard design of Robitzsch pyranometer developed
1948	Menzel developed sensitive sky photometer
1953	Invention of sunshine switch by Foster and Foskett of U.S Weather Bureau
1957	International Pyrheliometric Scale 1956' put into effect; original Ångström scale increased by 1.5% and 1913 Smithsonian scale decreased by 2.0%
1962	Adoption of the Campbell–Stokes sunshine recorder as an 'Interim Reference Sunshine Recorder' by the Commission on Instruments and Methods of Observation, WMO
1965	Introduction of Eppley precision pyranometer
1965	Development of automatic control of Ångström pyrheliometer by Marsh
1969	Introduction of Eppley black and white (star-type) pyranometer
1990	Wide-scale use of programmable data loggers for recording irradiance
2001	Invention of BF3 sensor by Delta-T company of Cambridge, England. The instrument integrates sunshine duration recorder with global and diffuse irradiance measurement

20.1.8 Classification of Pyrheliometer

The Commission for Instruments and Method of Observation of the WMO (1965), pyrheliometers are classified as standard, first, or second class, in accordance with the criteria given in Table 20.3. On the basis of these criteria, the commercially available pyrheliometer are classified as in Table 20.4.

20.1.9 Working Standard

Working standard pyrheliometers of Ångström type (referenced to Davos) are maintained at the regional and national centers designated by the WMO for the purpose of

TABLE 20.3 Classification of Pyrheliometers by WMO

	Standard	First Class	Second Class
Sensitivity (mW/cm^2)	± 0.2	± 0.4	± 0.5
Stability (% change in year)	± 0.2	± 1	± 2
Temperature (maximum error due to changes of ambient temperature, %)	± 0.2	± 1	± 2
Selectivity (maximum error due to departure from assumed spectral response, %)	± 1	± 1	± 2
Linearity (maximum error due to nonlinearity not accounted for, %)	± 0.5	± 1	± 2
Time constant (maximum)	25 s	25 s	1 min

TABLE 20.4 Classification of Commercial Pyrheliometers

Class	Type
Standard	Ångstrom electrical compensation pyrheliometer Abbot silver-disk pyrheliometer
First class	Michelson bimetallic pyrheliometer Linke–Feussner iron-clad pyrheliometer New Eppley pyrheliometer (temperature compensated) Yanishevsky thermoelectric pyrheliometer
Second class	Moll-Gorczynski pyrheliometer Old Eppley pyrheliometer

reproducing the International Pyrheliometric Scale 1956 (IPS). To a limited extent, the Abbot silver-disk pyrheliometer is still used at such centers. The secondary transfer instruments in current use are the other Ångstrom pyrheliometer and instruments of Eppley, Linke–Feussner, and Michelson types.

20.2 MEASUREMENT EQUIPMENT

According to European Solar Radiation Atlas (ESRA), solar radiation measurements can be broadly classified as ground-based measurements derived from geostationary satellites, which measures the energy reflected by the system (earth/atmosphere) in different wavelength bands (ESRA, 2000).

20.2.1 Pyranometer

This instrument measures global solar radiation. Figure 20.3 shows the structure of the CM11 Kipp and Zonen pyranometer. The pyranometer has the spectral response of between 335 and 2200 nm. A thermal detector in the sensing element responds to the total power absorbed from the solar radiation at any spectral distribution. The absorption of radiation on the black disk generates heat, and the heat energy flows to the heat sink through a thermal resistance. The temperature difference across the thermal resistance of the disk is converted into a small voltage, which can be detected by the logging system or computer. To avoid temperature fluctuation and reduce thermal radiation losses to the

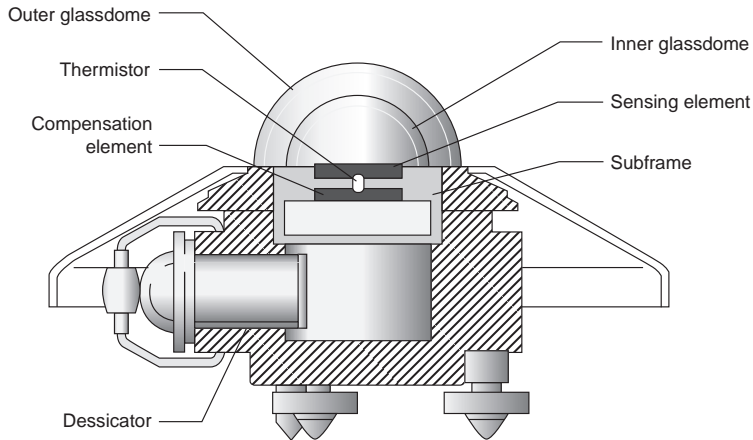


FIGURE 20.3 Kipp and Zonen CM11 pyranometer (Anon, 2010b).

atmosphere, the pyranometer is built with a double glass envelope. Furthermore, silica gel crystal is used to prevent moisture within the pyranometer. Periodic cleaning of the glass dome is recommended as debris may be collected over time. The working principle of the pyranometer given here is as discussed by Muneer (2004).

20.2.2 Pyranometer with Shading Device

This type of pyranometer measures the diffuse solar radiation. The shadow ring or disk shades the sun's direct beam from the pyranometer. Figure 20.4 shows a Kipp



FIGURE 20.4 Kipp and Zonen CM11 pyranometer with shading device (Clima, 2010).

and Zonen CM11 pyranometer with shadow ring. The shadow ring needs to be adjusted according to the sun's declination angle. A more expensive approach has been designed to track the sun's declination, where the disk will move accordingly or synchronous with the sun's movement. Hence, it produces a more accurate estimation of diffuse radiation.

The WMO classified pyranometer according to the features like stability, sensitivity, and so on. Classifications of pyranometer are divided into three classes, that is, first, second, and third class. The same features of classification apply to pyrheliometer where they are divided only into two classes, first and second class as mentioned before. The WMO characteristics of operational pyranometers are given in Table 20.5. Figure 20.5 shows the shade ring correction factors for the measurement of diffuse sky radiation.

TABLE 20.5 WMO Characteristic of Operational Pyranometer (Instruments, E. 2004)

Characteristic	High Quality ^a	Good Quality ^b	Moderate Quality ^c
Response time (95% response)	<15 s	<30 s	<60 s
Zero offset:			
(a) Response to 200 W/m ² net thermal radiation (ventilated)	7 W/m ²	15 W/m ²	30 W/m ²
(b) Response to 5 K/h change in ambient temperature	2 W/m ²	4 W/m ²	8 W/m ²
Resolution (smallest detectable change)	1 W/m ²	5 W/m ²	10 W/m ²
Stability (change per year, percentage of full scale)	0.8	1.5	3
Directional response for beam radiation (the range of errors caused by assuming that the normal incidence responsivity is valid for all directions when measuring, from any direction, a beam radiation whose normal incidence irradiance is 1000 W/m ²)	10 W/m ²	20 W/m ²	30 W/m ²
Temperature response (percentage maximum error due to any change of ambient temperature within an interval of 50 K)	2	4	8
Nonlinearity (percentage deviation from the responsivity at 500 W/m ² due to any change of irradiance within the range 100–1000 W/m ²)	0.5	1	3
Spectral sensitivity (percentage deviation of the product of spectral absorptance and spectral transmittance from the corresponding mean within the range 0.3–3 μ m)	2	5	10
Tilt response (percentage deviation from the responsivity at 0° tilt (horizontal) due to change in tilt from 0° to 90° at 1000 W/m ² irradiance)	0.5	2	5
Achievable uncertainty, 95% confidence level:			
Hourly totals	3%	8%	20%
Daily totals	2%	5%	10%

^aNear state-of-the-art, suitable for use as a working standard; maintainable only at stations with special facilities and staff.

^bAcceptable for network operations.

^cSuitable for low-cost networks where moderate to low performance is acceptable.

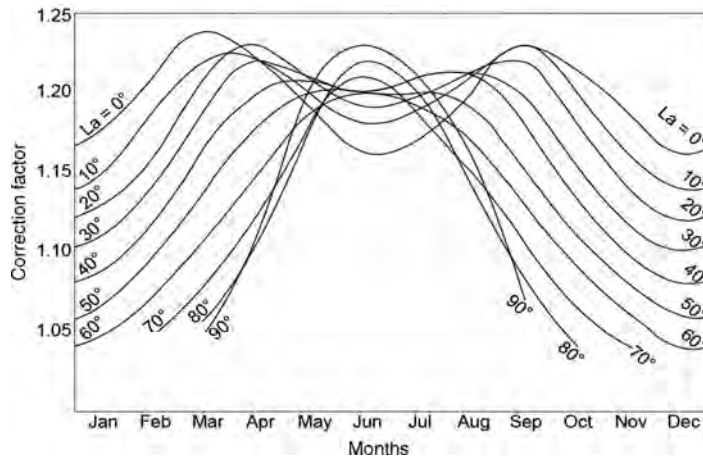


FIGURE 20.5 Shade ring correction factors for measured sky diffuse radiation.

20.2.3 Pyrheliometer

A pyrheliometer is used to measure beam (direct) radiation at normal incidence. Figure 20.6 shows a DN5 pyrheliometer used in an active tracking system. The long barrel of the pyrheliometer may be seen below the glass dome of the pyranometer in this picture. This equipment is equipped with a sun tracking system to enable it to



FIGURE 20.6 Middleton Solar-DN5 pyrheliometer (in active solar tracking system) (Anon, 2010a).



FIGURE 20.7 Eppley normal incidence pyrheliometer (EPLAP, 2010).

measure direct radiation as the sun moves. Inside the pyrheliometer there is a collimator with an optimum aperture of 6° , which can completely include the sun's disk. A multi-junction thermopile, which converts the heat from the sun's radiation to an electrical signal, is used so that the data can be read and recorded by data logger. To obtain the equivalent radiant energy flux (W/m^2) a calibration factor is applied. Other types of pyrheliometer are shown in Figures 20.7 and 20.8. Table 20.6 gives the characteristics of operational pyrheliometers.

The direct equipment cost of a pyrheliometer is approximately six times than a shaded pyranometer (Muneer, 2004). Because the measurement cost of direct normal radiation



FIGURE 20.8 The EKO-instrument's Sky scanner MS-321LR (Instruments, E. 2004).

TABLE 20.6 WMO Characteristic of Operational Pyrheliometers (Instruments, E. 2004)

Characteristic	High Quality ^a	Good Quality ^b
Response time (95% response)	<15 s	<30 s
Zero offset (response to 5 K/h change in ambient temperature)	2 W/m ²	4 W/m ²
Resolution (smallest detectable change in W/m ²)	0.5	1
Stability (percentage of full scale, change/year)	1	0.5
Temperature response (percentage maximum error due to change of ambient temperature within an interval of 50 K)	1	2
Nonlinearity (percentage deviation from the responsivity at 500 W/m ² due to the change of irradiance within 100 W/m ² to 1100 W/m ²)	0.2	0.5
Spectral sensitivity (percentage deviation of the product of spectral absorptance and spectral transmittance from the mean within the range 0.3–3 μ m) corresponding	0.5	1
Tilt response (percentage deviation from the responsivity at 0° tilt (horizontal) due to change in tilt from 0° to 90° at 1000 W/m ² irradiance)	0.2	0.5
Achievable uncertainty, 95% confidence level (see above)		
1 min totals		
Percent	0.9	1.8
kJ/m ²	0.56	1
1 h totals		
Percent	0.7	1.5
kJ/m ²	21	54
Daily totals		
Percent	0.5	1
kJ/m ²	200	400

^aNear state-of-the-art, suitable for use as a working standard; maintainable only at stations with special facilities and staff.

^bAcceptable for network operations.

is high, a simple relationship relating horizontal of global (I_G), diffuse (I_D), and beam (I_B) radiation may be used to estimate the latter component. The relation can be given as

$$I_G = I_D + I_B \sin \text{SOLALT} \quad (20.10)$$

where SOLALT is the solar altitude.

According to Perez et al. (1990) and Gueymard (2003), the present state of solar radiation and daylight model is such that they are approaching the accuracy limits set out by the measuring equipment. Hence, modeling procedures are now used to crosscheck the measurements.

20.2.4 Pyrgeometer

Pyrgeometer is used to measure long wave radiation, which falls within the infrared radiation spectrum (4.5–100 μ m). It consists of a thermopile with a 200–100 μ m radiation band sensitivity, a dome mirror with solar blind filter coating, which eliminates short wave radiation, a temperature sensor which measures the body temperature of the



FIGURE 20.9 EKO M-S202 pyrheliometer (Instruments, E., 2005).

instrument and a sun shield to reduce heat from radiation. Figure 20.9 shows an EKO MS-202 pyrheliometer.

20.2.5 Albedometer

The term “ground albedo” or simply “albedo” is often used interchangeably with “ground reflectance.” On the other hand, as Monteith (1959) has pointed out, the term “albedo” or “whiteness” refers to the reflection coefficient in the visible range of the spectrum, whereas “reflectance” denotes the reflected fraction of short wave energy. In this book, the term “albedo” has been used synonymously with reflectance, applying to the total short wave energy.

The importance of knowing the albedo for the determination of radiation balance of macro- and microclimates is well known. A good estimate of albedo of the surrounding terrain is a prerequisite for representative calculations related to the energy balance of vegetation, amount of potential transpiration, energy interception of walls, windows, roofs, and solar energy collectors. Therefore, the small- and large-scale variation of albedo is of interest. The variation in albedo is spatial and temporal owing to the changing landscapes of the earth and due to the seasonal presence of snow and to some extent moisture deposition.

As the name of the instrument suggest, it measures the reflected radiation as well as the global radiation. Figure 20.10 shows an albedometer. Generally, albedometer is constructed with two pyranometers. One is facing upward to measure incident global radiation and the other facing downward to measure the ground-reflected radiation. Both pyranometers provide output individually. Albedo can be calculated from the output data of the two pyranometers. To obtain the albedo value, the following equation may be used:

$$\text{Albedo} = \frac{\text{reflected radiation}}{\text{global radiation}} \quad (20.11)$$



FIGURE 20.10 Kipp and Zonen CMA 6 Albedometer (Envco., 2009).

20.2.6 Sunshine Recorder

According to the WMO, definition of sunshine duration is “*sunshine duration during a given period is defined as the sum of that sub-period for which the direct solar irradiance exceeds 120 W/m^2* ” (WMO, 2003).

In many countries, diurnal duration of bright sunshine is measured at a wide number of places. For over a century, these data have been measured using the well-known Campbell–Stokes sunshine recorder as shown in Figure 20.11, which uses a solid glass spherical lens to burn a trace of the sun on a treated paper, the trace being produced whenever the beam irradiation is supposedly above the aforementioned critical level. Although the

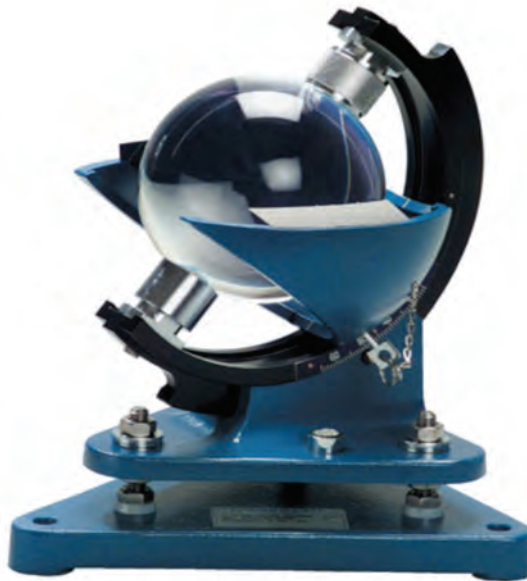


FIGURE 20.11 240-1070-L Campbell–Stokes Pattern Sunshine Recorder (Corporation, N., 2010).

critical threshold varies loosely with the prevailing ambient conditions, the sunshine recorder is an economic and robust device and hence used widely.

The limitations of the Campbell–Stokes sunshine recorder are well known and have been discussed in *Observers' Handbook* (1969), Painter (1981), and Rawlins (1984). Some of the associated limitations with this device are that the recorder does not register a burn on the card below a certain level of incident radiation (about 150–300 W/m²). On a clear day with a cloudless sky, the burn does not start until 15–30 min after sunrise and usually ceases about the same period before sunset. This period varies with the season. On the other hand, under periods of intermittent bright sunshine the burn spreads. The diameter of sun's image formed by the spherical lens is only about 0.7 mm. However, a few seconds' exposure to bright sunshine may produce a burnt width of about 2 mm. As such, intermittent sunshine may be indistinguishable from a longer period of continuous sunshine. In the past, a more sophisticated photoelectric sunshine recorder called the Foster sunshine switch (Foster and Foskett, 1953) has been used by the US Weather Service. This device incorporates two photovoltaic cells, one shaded and the other exposed to solar beam. Incident beam irradiation above a given threshold produces a differential output from the above two cells, the diurnal duration of which determines the hours of bright sunshine.

Suehrcke (2000) has presented an elegant and simple relationship that enables estimation of \bar{G} from the monthly averaged daily bright sunshine fraction and clear sky clearness index, $\bar{G}_{\text{clear}}/\bar{E}$. Thus:

$$\bar{G}/\bar{E} = (n/N)(\bar{G}_{\text{clear}}/\bar{E}) \quad (20.12)$$

where \bar{G} and \bar{E} are the monthly averaged daily terrestrial and extraterrestrial irradiation, \bar{G}_{clear} the monthly averaged daily terrestrial irradiation on clear day, n the average daily hours of bright sunshine, and N the day length.

Suehrcke has argued that $\bar{G}_{\text{clear}}/\bar{E}$ varies narrowly between 0.65 and 0.75. Given this small range of variation, in the absence of specific site information, an average

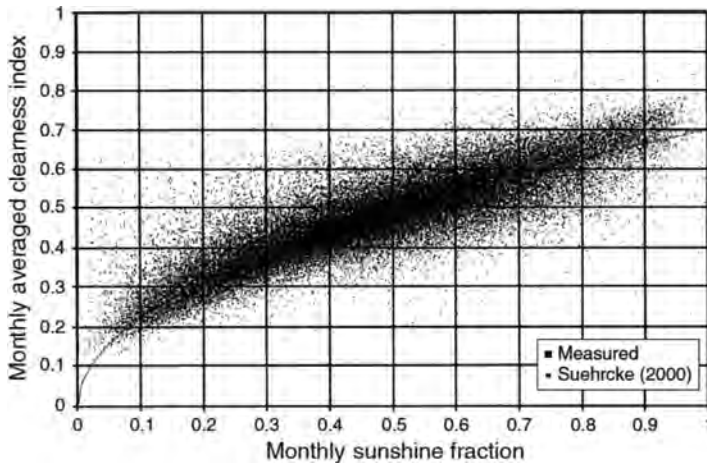


FIGURE 20.12 Driesse and Thevenard's (2002) evaluation of Suehrcke's "universal" relationship, see Equation (20.12).

value of 0.7 may be assumed for $\bar{G}_{\text{clear}}/\bar{E}$, thus giving Equation 20.12 the potential for worldwide application.

Driesse and Thevenard (2002) have evaluated the above claim. They have used measurements from 700 sites, compiled by the World Radiation Data Center with over 70 000 sunshine and radiation data. The relevant plot is shown in Figure 20.12. It may be noted that there is a considerable dispersion in the values of solar radiation. This variation (root mean square errors, RMSEs) was found to be around 12%. The latter team has fitted the above data to find regressed values of globally applicable Ångström a and b coefficients. These values are 0.2336 and 0.4987, respectively. Using the latter “universal” coefficients the performance of Suehrcke and Ångström procedures for estimating \bar{G} was found to be on par.

20.3 EQUIPMENT ERROR AND UNCERTAINTY

With any measurement there exist errors, some of which are systematic and others inherent of the equipment used. Angus (1995) has provided an account of the measurement errors associated with solar irradiance and illuminance measurements. These are summarized herein. The most common sources of error arise from the sensors and their construction. These are broken down into the most general types of error as follows:

- (a) cosine response,
- (b) azimuth response,
- (c) temperature response,
- (d) spectral selectivity,
- (e) stability,
- (f) nonlinearity,
- (g) thermal instability,
- (h) zero offset due to nocturnal radiative cooling.

To be classed as a secondary standard instrument (such as the CM11) pyranometers have to meet the specifications set out by WMO.

Of all the aforementioned errors, the cosine effect is the most apparent and widely recognized. This is the sensor’s response to the angle at which radiation strikes the sensing area. The more acute the angle of the sun, that is, at sunrise and sunset, the greater the error is (at altitude angles of sun below 6°). Cosine error is typically dealt with through the exclusion of the recorded data at sunrise and sunset times.

The azimuth error is a result of imperfections of the glass domes and in the case of solarimeters, the angular reflection properties of the black paint. This is an inherent manufacturing error, which yields a similar percentage error as the cosine effect. Like the azimuth error, the temperature response of the sensor is an individual fault for each cell. The photometers are thermostatically controlled, and hence the percentage errors due to fluctuations in the sensor’s temperature are reduced. However, the CM11 pyranometers have a much less elaborate temperature control system. The pyranometers rely on the two glass domes to prevent large temperature swings. Ventilation of the instrument is an additional recommended option.

The spectral selectivity of the CM11 is dependent on the spectral absorptance of the black paint and the spectral transmission of the glass. The overall effect contributes only

a small percentage error to the measurements. Each sensor possesses a high level of stability with the deterioration of the cells resulting in approximately $\pm 1\%$ change in the full-scale measurement per year. Finally, the nonlinearity of the sensors is a concern especially with photometers. It is a function of illuminance or irradiance levels. It, however, tends to contribute only a small percentage error toward the measured values. In addition to the above sources of equipment-related errors, care must be taken to avoid operational errors such as incorrect sensor leveling and orientation of the vertical sensors, as well as improper screening of the vertical sensors from ground-reflected radiation.

20.4 OPERATIONAL ERRORS

In Section 20.3, the likely errors resulting from equipment were introduced. Understanding and an assessment of operational errors are also of equal importance, and these are addressed below:

- (a) complete or partial shade-ring misalignment,
- (b) dust, snow, dew, water droplets, bird droppings, and so on,
- (c) incorrect sensor leveling,
- (d) shading caused by building structures,
- (e) electric fields in the vicinity of cables,
- (f) mechanical loading on cables,
- (g) orientation and/or improper screening of the vertical sensors from ground-reflected radiation,
- (h) station shutdown,
- (i) improper application of diffuse shade-ring correction factor,
- (j) inaccurate programming of calibration constants.

The sources of operation-related errors itemized above are self-explanatory. It is a good practice to protect cables from strong electric fields such as elevator shafts. Another source of error that may arise is from cables under mechanical load (piezoelectric effects). The piezoelectric effect is the production of electrical polarization in a material by the application of mechanical stress. Failure to protect cables from the above sources may produce “spikes” in the data, and these are shown as unusually high values of irradiance. Figure 20.13 demonstrates the sources of error categorized under items (a) and (b) discussed above. Such errors are best highlighted via cross plotting the diffuse ratio (the ratio of horizontal sky diffuse and the total or global irradiance) against clearness index (the ratio of horizontal global to extraterrestrial irradiance). Any serious departure of data from the normally expected envelope is thus identified. Figure 20.14 highlights error categorized under item (f).

20.5 DIFFUSE RADIATION DATA MEASUREMENT ERRORS

Historically, meteorological offices worldwide have used shade-ring correction procedure that is based on the assumption of an isotropic sky. However, during the past 15 years, a number of alternate, more precise methods that are based on a realistic, anisotropic sky

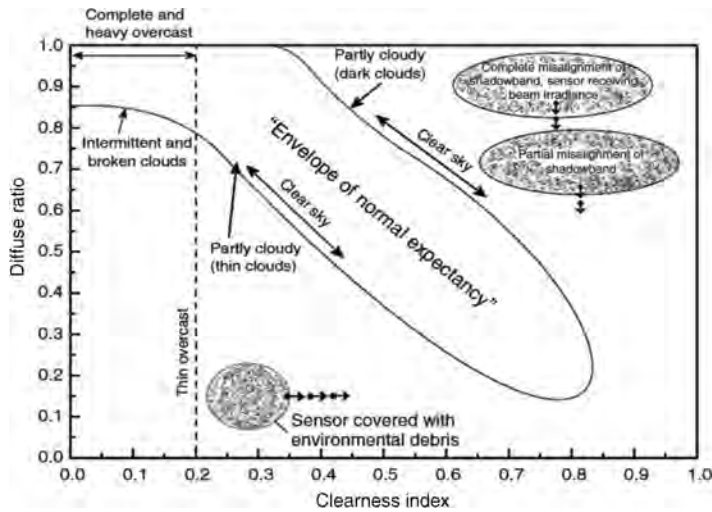


FIGURE 20.13 Demonstration of the sources of measurement errors.

have been established. The older, isotropic-sky corrected diffuse irradiation records are slightly higher for overcast conditions and lower by up to 10% for clear sky conditions. It is imperative that due care is taken in using a precise and validated shade-ring correction procedure since any errors in horizontal diffuse irradiance records will be multiplied by a large factor when horizontal beam irradiance and subsequently the total slope energy computations are undertaken.

The diffuse irradiance can be calculated from measurements of global parameter and the beam normal irradiance by using the following equation:

$$I_d^{\text{true}} = I_g - I_n \sin \alpha \quad (20.13)$$

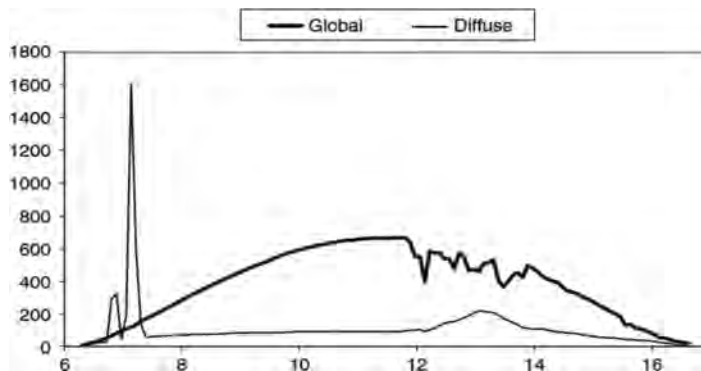


FIGURE 20.14 Demonstration of problems associated with mechanical loading on cables connecting data logger to irradiance sensor. Note: 5 min averaged data for Bahrain for December 12, 2001 (x-axis: the time of the day, y-axis: irradiance, W/m²).

where I_d^{true} is the diffuse horizontal irradiance which is referred hereafter as “true” diffuse, I_g the global irradiance on a horizontal surface, I_n the beam normal irradiance, and α the solar altitude. Measurements of beam normal irradiance are made using a pyrheliometer. Unfortunately, the collection of pyrheliometric data can be very expensive; the direct equipment cost alone is almost six times the expense of alternate collection methods, for example, prices quoted for Kipp and Zonen, premium grade equipment (>ISO 9000 standard) in February of 2003, were (USD) \$18909 for a pyrheliometric system and \$3245 for a pyranometer with a shadow band (Omni Instruments, 2012). The overall expense is most certainly directly related to the equipment cost but has a significant indirect cost, which is linked to the high level of daily maintenance required while using the system. An alternate and relatively inexpensive approach that is often used is to use a pyranometer with an occulting device to block the beam direct component from reaching the sensor. The most commonly used occulting devices are shadow-bands or shade rings. All shading devices are usually coated with a low albedo black paint to reduce any reflected radiation from the shadow band. The shadow bands are aligned and fixed parallel to the earth’s polar axis matching the sun’s track in the sky. This method produces a constant shadow on the pyranometer’s sensor, effectively blocking the beam direct irradiance and theoretically leaving only the diffuse irradiation. Shadow-band use is economical and effectively shades direct beam irradiation. Regrettably, shadow bands also shade a portion of the sky and the corresponding slice of the diffuse irradiance that is obscured by the band itself. This shading introduces errors that can markedly influence the accuracy of measurements. Because of this error, a correction factor needs to be introduced to the diffuse irradiance readings. The correction factor can be derived by accounting for the physical amount of sky that is blocked by the band and the amount of diffuse irradiance that is representative of that blocked portion. Various authors (Drummond, 1956; Stanhill, 2000; Pollard and Langevine, 1988; Dehne, 1984; LeBaron et al., 1990; Batlles et al., 1995; Muneer and Zhang, 2002; Painter, 1981; Kasten et al., 1983; Kudish and Ianetz, 1993; Batlles et al., 1998) have proposed many models, isotropic and anisotropic in nature, to correct the diffuse irradiance component when deploying a shadow band.

Historically, Drummond (1956) suggested a theoretical, isotropic model based on solar geometric calculations for shadow-band correction. The model could be applied anywhere in the world, but Drummond noted that as much as 7% additional correction was needed under cloudless sky to account for anisotropic conditions. The additional correction required for overcast or cloudy skies was 3%. Stanhill (2000) found that additional corrections for anisotropic conditions actually ranged between 14 and 30% to the isotropic correction. Other models have been investigated using various parameters such as solar declination, clearness index, cloud cover, and atmospheric turbidity (Pollard and Langevine, 1988). Dehne (1984) reported that attempts to use these models or duplicate the results proved disappointing for sites other than the site that was used to develop the models. LeBaron et al. (1990) suggested a model using four parameters covering both isotropic and anisotropic conditions. More recently, Batlles et al. (1998) and Muneer and Zhang (2002) introduced correction models describing an improved methodology accounting for both the isotropic and anisotropic correction contribution to “true” diffuse irradiance. The objective of this work is to evaluate the overall performance of the shadow-band correction models proposed by LeBaron et al. (1990), Batlles et al. (1998), and Muneer and Zhang (2002), against the benchmark Drummond’s model (Drummond, 1956), using data collected in disparate locations to those where models were originally developed.

20.5.1 Description of Models

All four of the diffuse shadow-band correction models considered herein (Drummond, 1956; LeBaron et al., 1990; Muneer and Zhang, 2002; Batlles et al., 1998) are well established and widely used by metrological offices worldwide. They have been adequately described in sufficient detail in the previously cited references. For the sake of brevity, these models are described in the following sections of this article. Hereafter, we refer to each model by the name of the first author.

1. *Drummond (1956)*: Drummond's model assumes an isotropic sky condition. The portion of the horizontal irradiance intercepted for the shadow band X is calculated according to the following equation:

$$X = \frac{2b}{\pi r} \cos^3 \delta \left[\left(\frac{\pi}{180} \varpi_0 \right) \sin \phi \sin \delta + \cos \phi \cos \delta \sin \varpi_0 \right] \quad (20.14)$$

where r is the radius of the shadow band, b the width of the shadow band, ϖ_0 the sunset/sunrise hour angle, ϕ the latitude, and δ the declination. The Drummond correction factor C_D is then expressed as

$$C_D = \frac{1}{1 - X} \quad (20.15)$$

By using Drummond's model, the corrected diffuse irradiance I_{dD} may be obtained by applying the following formula:

$$I_{dD} = C_D I_d \quad (20.16)$$

2. *LeBaron et al. (1990)*: This model uses four parameters that take into account varying (isotropic and anisotropic) sky conditions. Three of the parameters used within this method are solar altitude and two dimensionless indices introduced by Perez et al. (Pollard and Langevine, 1988), termed, respectively, as the clearness of the sky, ε , and the brightness of the sky, Δ . The forth parameter is the Drummond's correction factor C_D describe earlier. Using these four parameters, LeBaron et al. (Gueymard, 2003) clustered the data according to 256 categories. For each one of these categories, they determined an appropriate value of the correct factor.
3. *Batlles et al. (1995)*: Batlles et al. (WMO, 2003) developed a piecewise multiple linear model using the same four parameters used in LeBaron's approach. The correction factor C_B is then expressed as an analytical function of C_D , α , and Δ , which is parameterized against ε . The set of equations are

$$\varepsilon \leq 3.5 \quad C_B = 1.178C_D + 0.207 \log \Delta + 0.122e^{-1/\sin \alpha} \quad (20.17a)$$

$$3.5 \leq \varepsilon \leq 8 \quad C_B = 1.454C_D + 0.665 \log \Delta + 0.4756e^{-1/\sin \alpha} \quad (20.17b)$$

$$8 \leq \varepsilon \leq 11 \quad C_B = 1.486C_D + 0.495 \log \Delta \quad (20.17c)$$

$$\varepsilon > 11 \quad C_B = 1.384C_D + 0.363 \log \Delta \quad (20.17d)$$

Measurements of diffuse irradiance are then corrected for isotropic and anisotropic conditions using:

$$I_{dB} = C_B I_d \quad (20.18)$$

4. *Muneer and Zhang (2002)*: This model is based on the sky patch radiance distribution work of Moon and Spencer (Dehne, 1984). The model incorporates both isotropic and anisotropic elements by means of the following equation:

$$D = \left(\frac{\pi L_Z}{6} \right) \left[\frac{3 + 2b_1}{1 + b_1} + \frac{3 + 2b_2}{1 + b_2} \right] \quad (20.19)$$

where D is the diffuse irradiance calculated in terms of the radiance indices and L_Z . Parameters b_1 and b_2 are the radiance distribution indices for the two sky quadrants containing sun and opposed to sun, given as
for $k > 0.2$,

$$b_1 = \frac{3.6 - 10.462k_t}{-0.4 + 6.974k_t} \quad (20.20)$$

$$b_2 = \frac{1.565 - 0.990k_t}{0.957 - 0.660k_t} \quad (20.21)$$

for $k \leq 0.2$, $b_1 = b_2 = 1.68$. The correction factor C_M may be calculated as follows:

$$I_1 = \cos \phi \cos \delta \sin \omega_0 + \omega_0 \sin \phi \sin \delta \quad (20.22)$$

$$I_2 = \omega_0 \sin^2 \phi \sin^2 \delta + 2 \sin \omega_0 \sin \phi \cos \phi \sin \delta \cos \delta + \cos^2 \phi \cos^2 \delta \left[\frac{\omega_0}{2} + \frac{\sin 2\omega_0}{4} \right] \quad (20.23)$$

$$F = \frac{2b}{r} L_Z \cos^3 \delta \frac{I_1 + b_1 I_2}{1 + b_1} \quad (20.24)$$

$$C_M = \frac{1}{1 - \frac{F}{D}} \quad (20.25)$$

Tables 20.7–20.9 give the statistical results of the models discussed above for different types of sky condition. Figure 20.15 shows corrected diffuse irradiance versus true diffuse irradiance.

The performance of the models considered are analyzed by the residual differences (calculated as corrected diffuse irradiance minus true diffuse irradiance values) against clearness index and solar altitude. Figures 20.16a,b and 20.17a,b show the residuals

TABLE 20.7 Statistical Results for All-Sky Conditions (Mean Value of the True Diffuse Irradiance is 153.8 W/m²)

Correction Model	R^2	RMSE (%)	MBE (%)
Uncorrected	0.96	26.8	−20.9
Drummond	0.96	16.6	−10.2
LeBaron	0.96	15.8	−8.7
Batiles	0.96	13.6	−2.1
Muneer	0.96	12.9	−4.9

TABLE 20.8 Statistical Results for Cloudy to Part-Cloudy Sky Conditions (Mean Value of the True Diffuse Irradiance is 207.4 W/m²)

Correction Model	R^2	RMSE (%)	MBE (%)
Uncorrected	0.97	23.5	-18.0
Drummond	0.98	12.4	-7.2
LeBaron	0.98	8.9	-3.1
Batiles	0.98	9.7	2.2
Muneer	0.98	9.0	-3.6

TABLE 20.9 Statistical Results for Cloudless Sky Condition (Mean Value of the True Diffuse Irradiance is 124.1 W/m²)

Correction Model	R^2	RMSE (%)	MBE (%)
Uncorrected	0.89	29.4	-23.5
Drummond	0.89	20.5	-13.0
LeBaron	0.87	21.7	-13.9
Batiles	0.88	17.2	-6.0
Muneer	0.89	16.4	-6.1

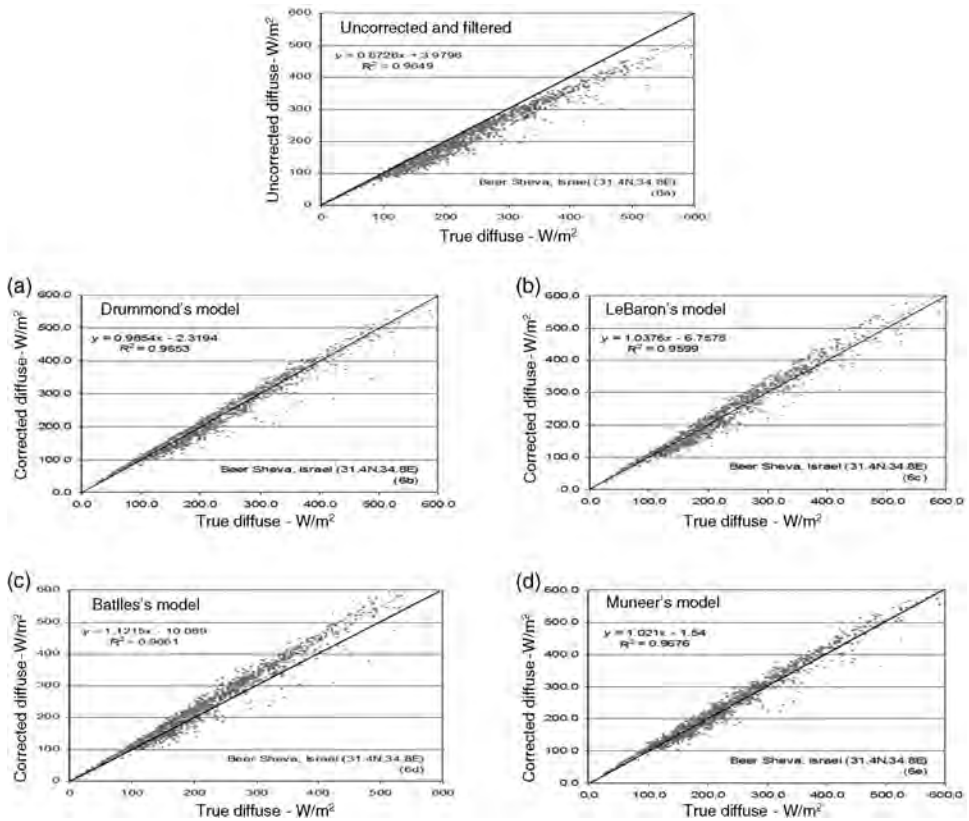


FIGURE 20.15 Shadow band diffuse irradiance corrected using (a) Drummond's, (b) LeBaron's, (c) Batiles's, and (d) Muneer's models against true diffuse irradiance.

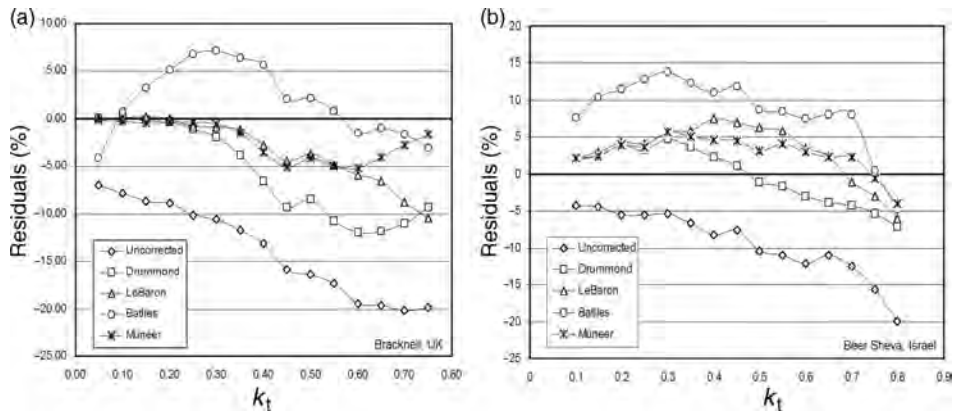


FIGURE 20.16 Residual differences against clearness index. (a) Bracknell, UK, and (b) Beer Sheva, Israel.

expressed as percentage of the mean value of true diffuse irradiance, respectively, for each interval of k_t and solar altitude. Uncorrected measurements of diffuse irradiance present estimates ranging from -5% for low k_t values to -20% for k_t values around 0.8. Residuals from Drummond's isotropic model present almost no deviation for $k_t < 0.2$, whereas an accentuated underestimation is noted for $k_t > 0.6$ with residuals lower than -8 to -10% . This result shows that under cloudy conditions, diffuse irradiance can be considered isotropic, and thus, a correction factor accounting only for isotropic conditions, such as the one proposed by Drummond, leads to fairly accurate corrections. Under cloudless conditions, sky anisotropy plays a major role and an isotropic correction factor is not enough. It is interesting to note that residuals from LeBaron's and Muneer's model present a similar profile, with the worst performers being the Drummond and Battles models.

Four models for correcting diffuse irradiance measurements using shadow band have been analyzed. The Drummond's model treats the sky as an isotropic entity, while the

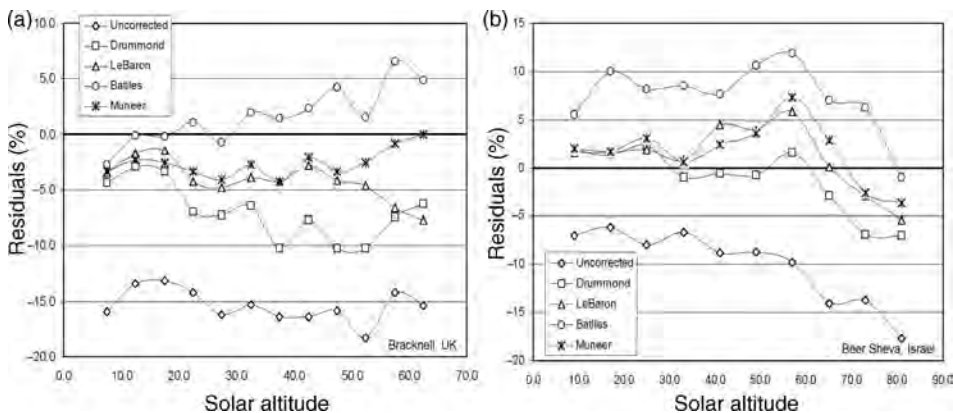


FIGURE 20.17 Residual differences against solar altitude. (a) Bracknell, UK, (b) Beer Sheva, Israel.

other three newer models are of an anisotropic nature. As shown above, using an isotropic correction factor leads to an overall underestimation of the actual diffuse irradiation of 4–11%. The use of anisotropic correction schemes will considerably reduce the above underestimation depending on the model that is chosen. The results showed that the improvement in accuracy is even further improved under a cloudless sky and that the LeBaron and Muneer's models were better in performance than the Drummond and Batlles procedures. Finally, owing to the large errors it generates, particularly in the higher clearness index range where the capture of solar energy is most crucial, it is recommended that Drummond's model is dropped as the *de facto* procedure currently used by numerous meteorological organizations.

20.6 TYPES OF SENSORS AND THEIR ACCURACY

A survey of radiation instruments undertaken by Lof et al. (1965) showed that of the 219 sensors in use across Europe, 65 were of the CM11 type pyranometers while 107 sensors were the simpler and less expensive Robitzch actinographs with a bimetallic temperature element. The latter instrument is also quite popular in the developing Asian (89 such sensors were reported to be in use), African (20.20.16 sensors), and South American (20.47 sensors) countries where maintenance is often the key factor. The lead author (Muneer) has in the past visited a solar radiation measurement station in the middle of the Sahara desert and seen the Robitzch actinograph faithfully recording a regular trace of irradiation. The weekly changeover of the recording chart makes this instrument an ideal choice for remote locations. Although not in use with the North American meteorological network, it is known to be of use over there in biological and agriculture-related work (Coulson, 1975). Drummond (1965) estimates that accuracies of 2–3% are attainable for daily summations of radiation for pyranometers of first-class classification. Individual hourly summations even with carefully calibrated equipment may be in excess of 5%. Coulson (1975) infers that the errors associated with routine observations may be well in excess of 10%. Isolated cases of poorly maintained equipment but those that are in the regular network may exhibit monthly averaged errors of 10% or more. The Robitzch actinograph, even with all the modifications to improve its accuracy, is suitable only for daily summations. At this interval, it provides an accuracy of around 10%. However, not all designs of the latter sensor can claim even this level of accuracy. These figures must be borne in mind when evaluating the accuracy of the relevant computational models.

20.7 MODERN DEVELOPMENTS

Delta-T device of Cambridge, England, has produced a relatively new instrument that measures the horizontal global and diffuse irradiance as well as sunshine duration from a single, stationary sensor. Unlike the Campbell–Stokes sunshine recorder, the Delta-T device neither requires any cards or shade ring for the measurement of sky-diffuse irradiance. The BF3 model, shown in Figure 20.18, enables simultaneous recording of the abovementioned three sets of data in a sensor that used no moving parts and required no specific polar alignment or routine adjustment. The electronic outputs are compatible with electronic data loggers and work at any latitude.



FIGURE 20.18 The BF3 sensor (D-T Devices, 2010). (Photo courtesy of Delta-T, Cambridge, England).

The device uses a system of photodiodes and a shading pattern such that wherever the sun is in the sky at least one photodiode is always exposed to the full solar beam and at least another one diode is always completely shaded. All photodiodes receive an equal sampling of diffuse light from the rest of the sky hemisphere. A special layout of seven photodiodes on a hexagonal grid, covered by a hemispherical shading pattern, ensures the satisfaction of the above constraints. A sketch of the shading pattern of BF3 is shown in Figure 20.19, plotted in a 180° fisheye lens view.

The sensor has a digital output of sunshine presence. As pointed out above, bright sunshine is defined by the WMO as the duration with irradiance greater than 120 W/m^2 in the



FIGURE 20.19 Hemispherical shading pattern for Delta-T BF3 irradiance sensor.

TABLE 20.10 Statistical Summary of BF3 Hourly Average with Respect to Kipp CM11 Readings, Feb 22–Jul 3, 2000

	Calibration Error (%)	R^2	Standard Error W/m ²
Global	4.7	0.994	16.5
Diffuse	1.4	0.980	13.4

direct beam measured perpendicular to the beam. The microprocessor algorithm within the BF3 sensor uses the measured global and diffuse irradiance to provide an estimate of sunshine duration. An account of BF3's performance, evaluated against Kipp and Zonen irradiance sensors and Campbell–Stokes sunshine recorders, has been presented by Wood et al. (2003).

Table 20.10 gives the results of regression analysis of the BF3 sensor with respect to the Kipp CM11 using the following indicators:

1. calibration error, the deviation of the slope of the line of best fit from unity (expressed in percentage terms),
2. coefficient of determination (R^2),
3. standard error, the root mean square deviation of the measured BF3 values from the line of best fit.

These results show a good match between the global and diffuse outputs of the BF3, and those measured using the Kipps and shade ring.

For evaluation against Campbell–Stokes recorders, two recorders were placed adjacent to each other, namely, CS1 and CS2. These are compared to each other, and also to the WMO reference, for the days when all the data are available. Cards from the CS1 were independently analyzed by two different people, giving results CS1 (SY) and CS1 (JW). Table 20.11 gives a summary of these different regressions. Figure 20.20 shows the BF3 and CS1 values plotted against the WMO reference.

These results show that the Campbell–Stokes recorder is a relatively poor performer when judge against the WMO sunshine definition. It shows a typical error of nearly an hour, some four times greater than the BF3. Although the two adjacent C–S recorders gave fairly consistent results when interpreted by the same person, the two independent operators gave very different interpretations of the same set of record cards, despite working from the same set of guidelines. The variability in interpretation was nearly half as

TABLE 20.11 Sunshine Hour Regressions

Regression	Calibration Error (%)	R^2	Standard Error Hr
BF3 v WMO	−0.2	0.993	0.23
CS1(JW) v WMO	1.3	0.902	0.86
CS1(SY) v WMO	7.5	0.893	0.91
CS2(SY) v WMO	6.3	0.893	0.90
CS1(JW) v CS1(SY)	6.1	0.980	0.38
CS1(SY) v CS2(SY)	1.1	0.999	0.09

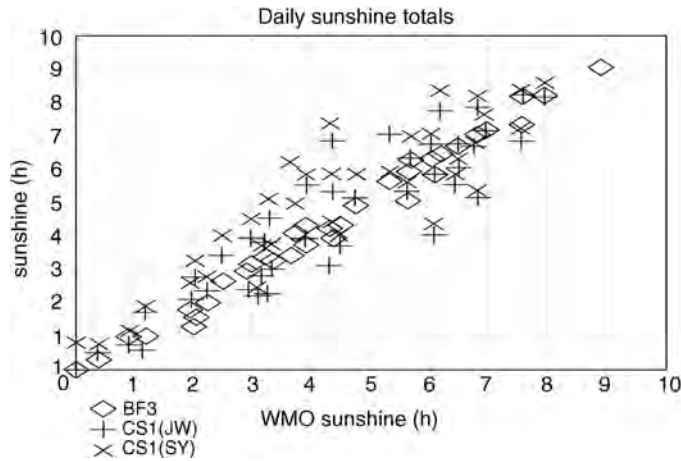


FIGURE 20.20 BF3 and Campbell–Stokes recorders compared to WMO reference. Note: BF3 = Delta-T BF3 irradiance sensor, CS1 (JW) = Campbell–Stokes recorder 1, CS1 (SY) = Campbell–Stokes recorder 2. Names of observers: John Wood (JW) and Serge Younes (SY).

much as the total error relative to the WMO standard, though neither operator was in fact consistently more accurate than the other.

20.8 DATA QUALITY ASSESSMENT

Data quality assessment is a process or a procedure to avoid spurious data to be included in the data set. Gueymard and Kembezidis (2004) pointed out that even though a data set has passed the quality assessment process or procedure, the data must be examined for its uncertainty that has been transferred from a measuring sensor to the actual measurement. Here, in this section, a summary of a few assessment methods will be discussed.

20.8.1 United State National Renewable Energy Laboratory (NREL), 1993 (NREL, 1993)

The US NREL developed a quality assessment procedure named SERI QC. This procedure will assess the three radiation data elements, namely, global horizontal, diffuse horizontal, and direct normal. Summary of the key features are discussed as follows:

- At the beginning, SERI QC will perform one-element test by defining a range of acceptable values for K_t , K_d , or K_n between minimum and maximum, which depends on the element that is being tested, based on air mass regimes and month of the year.
- If the zenith angle is less than or equal to 80° and all three elements are present, then SERI QC will perform a three-element test. A range of acceptable values will be defined so that the equation $K_t = K_d + K_n$ is satisfied within the arbitrary error limit of ± 0.03 , which accounts for the measurement uncertainties.
- If the data pass the three-element test or two elements pass the one-element test, SERI QC will perform a two-element test by defining a range of acceptable values within the boundaries empirically to determine three different air mass regimes for each month using data collected from the site.

- Flags are assigned to the data after the test. The flagging system of SERI QC permits the assignment of uncertainties that depend on the nature of the test performed (one, two, or three elements) and the distance by which the data point exceed the expected limit.

The values of K_t , K_d , K_n can be defined as follows:

K_t = clearness index or global horizontal transmittance or global horizontal radiation/extraterrestrial horizontal radiation,

K_d = diffuse horizontal transmittance or diffuse horizontal radiation/extraterrestrial horizontal radiation,

K_n = direct normal transmittance or direct normal radiation/extraterrestrial direct normal radiation.

For more in-depth information regarding the procedure, readers are referred to the following website: http://rredc.nrel.gov/solar/pubs/seri_qc/

20.8.2 Commission Internationale de l'éclairage (CIE) Automatic Quality Control (Kendrick, 1994)

Brief summary of the five tests in CIE quality control for radiation and illuminance is presented below:

1. A rough boundary limits for global and diffuse irradiance and direct irradiance is set to be less than the extraterrestrial irradiance.
2. To ensure consistency, it utilize the redundancy among three solar radiation components or the diffuse component to be less than the global component plus a 10% allowance for shade ring correction where beam component is not measured.
3. Each aspect of the global irradiance and illuminance is tested, that is, north, east, south, and west.
4. Intercomparisons tests is done between irradiance and illuminance.
5. Tests to compare the zenith luminance with either diffuse irradiance or illuminance were also included.

Bear in mind that CIE noted that automatic testing should not be performed when global irradiance is below 20 W/m^2 and solar elevation is less than 4° .

20.8.3 Page Model

The Page model is based on the work undertaken for the production of the European Solar Radiation Atlas (ESRA) and the Chartered Institution of Building Services Engineers (CIBSE) Guide on weather and solar data (ESRA, 2000; Page, 1997). Page sets out the following steps to control all daily totals of solar radiation data:

- Values for global solar radiation have to be less than the extraterrestrial radiation, and sunshine values have to be less than or equal to corresponding astronomical values.
- Solar radiation values have to lie within the range of the expected clear sky extreme values by considering the influence of the atmospheric layer.

- Basic relationship between different radiation components should be fulfilled.
- Values of solar radiation parameters have to be in specific range compared to nearby station's value with allowance for spatial variability.
- The variation of relative terms (G/E) of the Ångström regression should lie within a defined range.

Note that further details and software for the Page model are provided by Muneer (2004).

20.8.4 Muneer and Fairooz Quality Control Procedure (Muneer and Fairooz, 2002)

This quality control procedure consists of four levels of tests, which emphasize on the global and diffuse radiation. The procedure was developed based on CIE recommendation for first-level test and Page irradiance model for fourth-level test. Those levels of tests are summarized as follows:

1. Adopted from CIE quality control;

$$0 < G < 1.2 E_n,$$

$$0 < D < 0.8 E_n,$$

where E_n is the normal incidence extraterrestrial irradiance.

2. Consistency test that compares between diffuse and global irradiation and between global and horizontal extraterrestrial irradiation.
3. Test based on an expected diffuse ratio-clearness index envelop. This check is to make sure that the diffuse irradiation data is conformed to the limit that set out by the envelope of acceptance.
4. Check on the quality of diffuse irradiance is performed by comparing its value with the diffuse irradiance under two extreme conditions as define by Page.

A further test is carried out on the diffuse and global irradiance by investigating the Linke turbidity values, for example, when the Linke turbidity value is less than 2.5 or greater than 12, a close inspection of the corresponding data is required.

Refer to the graphical procedure as mentioned in level three tests above, Younes et. al. (2005) proposed a new standard deviation procedure to produce an envelope of acceptance. This procedure basically categorize diffuse ratio-clearness index in the band of k_t . For any given band of k_t , outliers are identified as data points lying outside the envelope, which is defined by $\bar{k} \pm 2\sigma_k$ boundaries. For more in-depth discussion, readers are referred to the reference article.

20.9 STATISTICAL EVALUATION OF MODELS

As pointed out in Section 20.1, the accuracy of mathematical models for estimating solar radiation is approaching the measurement accuracy of instruments. A number of routines, therefore, now use mathematical models for assessing the quality of measured data. Thus, in this section, an account of crosschecking the two set of procedures, that is, relationship or trend between the two sets of quantities, that is, models and measured data is presented.

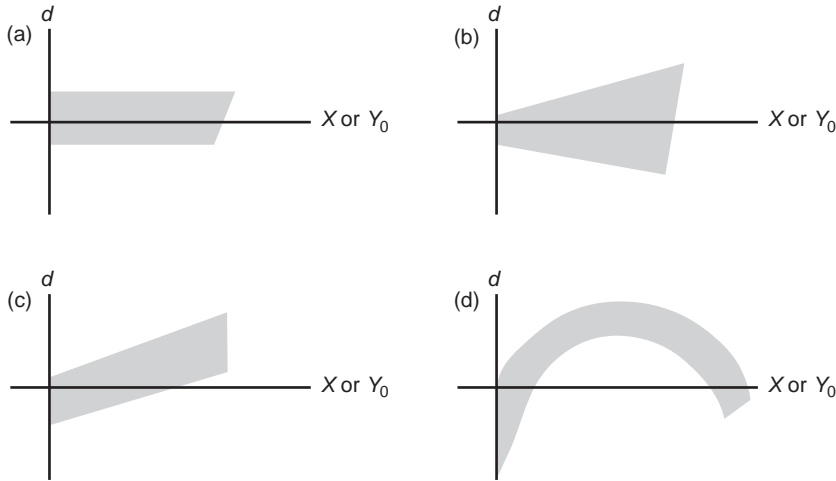


FIGURE 20.21 Plot of residuals for evaluating the adequacy of the model: (a) adequacy, (b) Y_o needs transformation, (c) missing linear independent variable, and (d) missing linear or quadratic independent variable.

Checking on the adequacy of the mathematical model describing any physical process such as a given solar radiation model is important not only in the final stages of the work program, but more particularly in the initial phase. An examination of residuals is recommended. The procedure is to produce a graph of the residuals d (the difference between observed Y_o and calculated Y_c values of the dependent variable) plotted against the independent variable X or the observed value Y_o .

If the residuals fall in a horizontal band as shown in Figure 20.21a, the model may be judged as adequate. If the band widens as X or Y_o increases, displayed in Figure 20.21b, it indicates a lack of constant variance of the residuals. The corrective measure in this case is a transformation of the Y variable. A plot of the residuals such as Figure 20.21c indicates the absence of an independent variable in the model under examination. If, however, a plot such as Figure 20.21d is obtained, a linear or quadratic term would have to be added.

Often correlation between two quantities is also to be examined. In solar energy literature, it has become a common practice to refer regression models as “correlation equations.” Strictly speaking, this is wrong usage of statistician’s language. Correlation is the degree of relationship between variables, and one seeks to determine how well a linear or other model describes the relationship. On the other hand, regression is a technique of fitting linear or nonlinear models between a dependent and a set of independent variables. Thus, fitting an equation of the form:

$$Y = a_0 + a_1X \quad (20.26)$$

for n pairs of (X, Y) is an example of linear regression. On the other hand, fitting:

$$Y = a_0 \exp(a_1X) + b_0 \sin(b_1X) \quad (20.27)$$

is an example of a nonlinear model. A number of low-priced software packages are available, which adequately cover the requirements of fitting linear and nonlinear models. The popular spreadsheet packages such as Lotus 1-2-3 (a product of Lotus Corporation) and

Excel (a product of Microsoft Corporation) as well as more specialist statistical packages such as SOLO and BMDP (products of BMDP Statistical Software Inc., California) are a few examples. For handling very large data arrays, one has to resort to FORTRAN and C environments. The text *Numerical Recipes* by Press et al. (Press, 1992), with its companion electronic suite of programs, offers solutions at this end. All of the above packages use robust and efficient routines, which obviate any particular need for developing optimization programs from scratch. In the following paragraphs, a brief discussion on the statistical examination of models is provided.

20.9.1 Slope of the Best-fit Line, s

The slope of the best-fit line, given by Equation 20.28, between the computed and measured variable is desired to be as close as possible to unity. Slope values exceeding one indicate overestimation, while slope values under one indicate underestimation of the computed variable,

$$s = \frac{\sum(Y_m - \bar{Y}_m)(Y_c - \bar{Y}_c)}{\sum(Y_m - \bar{Y}_m)^2} \quad (20.28)$$

Note that Y_c is the calculated value of the dependent variable, Y_m the measured or observed value, and \bar{Y}_m the mean value of the measured variable.

20.9.2 Coefficient of Determination, r^2

The coefficient of determination (r^2) is the ratio of explained variation to the total variation. r^2 lies between zero and one,

$$r^2 = \left[\frac{\sum(Y_m - \bar{Y}_m)(Y_c - \bar{Y}_c)}{\sqrt{\sum(Y_m - \bar{Y}_m)^2 \sum(Y_c - \bar{Y}_c)^2}} \right]^2 \quad (20.29)$$

A high value of r^2 , thus indicating a lower unexplained variation, is desirable. r^2 is often used to judge the adequacy of a regression model, but it should not be the sole criterion for choosing a particular model. In the present context, r^2 provides an indication of the order of scatter between Y_c and Y_m . Further information may be obtained in Montgomery and Peck (1992) and Draper and Smith (1998).

20.9.3 Coefficient of Correlation, r

The square root of the coefficient of determination is defined as the coefficient of correlation r . It is a measure of the relationship between the variables based on a scale ranging between +1 and -1. Whether r is positive or negative depends on the inter-relationship between x and y , that is, whether they are directly proportional (y increases and x increases) or vice versa. Once r has been estimated for any fitted model, its numerical value may be interpreted as follows. Let us assume that for a given regression model $r = 0.9$. This means $r^2 = 0.81$. It may be concluded that 81% of the variation in Y has been explained (removed) by the model under discussion, leaving 19% to be explained by other factors.

20.9.4 Student's t -Distribution

Often the modeler is faced with the question as to what quantitative measure is to be used to evaluate the value of r^2 obtained for any given model (Owen and Jones, 1990). Clearly, r^2 would depend on the size of the data population. For example, a lower value of r^2 obtained for a model fitted against a large database may or may not be better than another model that used a smaller population. In such situations, the Student's t -test may be used for comparing the above two models. The following example demonstrates the use of this test of significance for r^2 .

EXAMPLE 20.1: *For a given location, a regression model between average clearness index (\bar{K}_T) and monthly averaged sunshine fraction (n/N) gives $r^2 = 0.64$ for 12 pairs of data points. Using Student's t -test investigates the significance of r .*

The test statistic $t = (n - 2)^{0.5} [r / \sqrt{(1 - r^2)}]$, where n is the number of data points and $(n - 2)$ is the degrees of freedom (d.f.).

Thus,

$$\text{Test statistic } t = (12 - 2)^{0.5} [0.8 / \sqrt{(1 - 0.64)}] = 4.216.$$

In this example there are 10° d.f. Thus from Table 20.12, the value of $r = 0.8$ is significant at 99.8% but not at 99.9% (note that for d.f. = 10 and $t = 4.216$ lies between 4.144 and 4.587 corresponding to columns for 0.998 and 0.999, respectively). In layman terms, this means that using the above regression model, \bar{K}_T may be estimated with a 99.8% confidence.

TABLE 20.12 Percentile Values for Student's t -Distribution (EPLAP, 2010)

d.f	$P = 0.95$	0.98	0.99	0.998	0.999
1	12.706	31.821	63.657	318.310	636.620
2	4.303	6.965	9.925	22.327	31.598
3	3.182	4.541	5.841	10.214	12.924
4	2.776	3.747	4.604	7.173	8.610
5	2.571	3.365	4.032	5.893	6.869
6	2.447	3.143	3.707	5.208	5.959
7	2.365	2.998	3.499	4.785	5.408
8	2.306	2.896	3.355	4.501	5.041
9	2.262	2.821	3.250	4.297	4.781
10	2.228	2.764	3.169	4.144	4.587
15	2.131	2.602	2.947	3.733	4.073
20	2.086	2.528	2.845	3.552	3.850
25	2.060	2.485	2.787	3.450	3.725
30	2.042	2.457	2.750	3.385	3.646
40	2.021	2.423	2.704	3.307	3.551
60	2.000	2.390	2.660	3.232	3.460
120	1.980	2.358	2.617	3.160	3.373
200	1.972	2.345	2.601	3.131	3.340
500	1.965	2.334	2.586	3.107	3.310
1000	1.962	2.330	2.581	3.098	3.300
∞	1.960	2.326	2.576	3.090	3.291

20.9.5 Root Mean Squared Error, RMSE

The root mean squared error (RMSE) gives a value of the level of scatter that the model produces. This is an important statistical test as it highlights the readability and repeatability of the model. It provides a term-by-term comparison of the actual deviation between the predicted and the measured values. Because it is a measure of the absolute deviation, RMSE is always positive. A lower absolute value of RMSE indicates a better model. Mathematically, it is given by the following equation:

$$\text{RMSE} = \sqrt{\sum \left[\frac{(Y_c - Y_m)^2}{n} \right]} \quad (20.30)$$

20.9.6 Mean Bias Error, MBE

The mean bias error (MBE) provides an indication of the trend of the model, whether it has a tendency to underpredict or overpredict the modeled values. MBE can be expressed either as a percentage or as an absolute value. Nevertheless, within a data set, an overestimation of one observation can cancel an underestimation of another. A MBE nearest to zero is desired. It is given by the following equation:

$$\text{MBE} = \frac{\sum (Y_c - Y_m)}{n} \quad (20.31)$$

20.9.7 Mean of Absolute Deviations, MAD

Another metric that is often used in such analysis is the mean of absolute deviations, (MAD), and is given by,

$$\text{MAD} = \frac{\sum |Y_c - Y_m|}{n} \quad (20.32)$$

Unlike MBE, the MAD metric provides an insight into the scatter between Y_c and Y_m . Note that the MAD is similar to RMSE and provides a measure of absolute deviations.

20.9.8 Nondimensional MBE, MAD, and RMSE

The above formulas provide MBE, MAD, and RMSE, which have the same physical units as the dependent variable, Y . In some instances, nondimensional MBE (NDBE), MAD (NDMAD), and RMSE (NDRMSE) are required. These are obtained as follows:

$$\text{NDBE} = \frac{\sum \left[\frac{(Y_c - Y_m)}{Y_m} \right]}{n} \quad (20.33)$$

$$\text{NDMAD} = \frac{\sum \left| \left[\frac{(Y_c - Y_m)}{Y_m} \right] \right|}{n} \quad (20.34)$$

$$\text{NDRMSE} = \sqrt{\frac{\sum \left[\frac{(Y_c - Y_m)}{Y_m} \right]^2}{n}} \quad (20.35)$$

20.9.9 Figure of Merit, ψ

One of the important steps in the evaluation of different functions is the interpretation of different statistical parameters, namely, slope, r^2 , MBE, MAD, and RMSE. Often when two or more models are intercompared for their relative strengths and weaknesses, there may be a tie between the above-mentioned metrics. For example, a model may have a lower MBE but a higher RMSE. Therefore, an overall accuracy score is highly desirable to facilitate a discrete comparison between different models. In this article, a novel statistical tool combines the five metrics mentioned above to produce an overall score. With the view to demonstrate this point, Figure 20.22a shows a slope that has a large deviation from the ideally sought value of 1 but a high value of r^2 , whereas in Figure 20.22b the slope is very close to the ideal value, but a low value of r^2 is realized due to large data scatter. Therefore, case in Figure 20.22b would be preferable over case in Figure 20.22a. Similarly, Figure 20.22c presents a smaller but systematic trend of deviation, notice the negative deviations in the middle range, with positive outcomes at the lower and higher ends. In the case of Figure 20.22d, an almost equal spread of positive and negative but larger deviations is noticed. Although case in Figure 20.22d would provide a much higher value of MBE, case in Figure 20.22d would, however, be preferable over case in Figure 20.22c. Overall, it can be concluded that the slope parameter provides a much more important indication of the validity of any given model. The r^2 of the line fitted between computed and observed data, MBE, MAD, and RMSE for the given model's deviation provide second order information as higher values of r^2 or lower values of MBE, MAD, and RMSE do not warrant a better model. Ideally, the latter four parameters ought to be examined in conjunction with the value of slope. The following overall accuracy score is

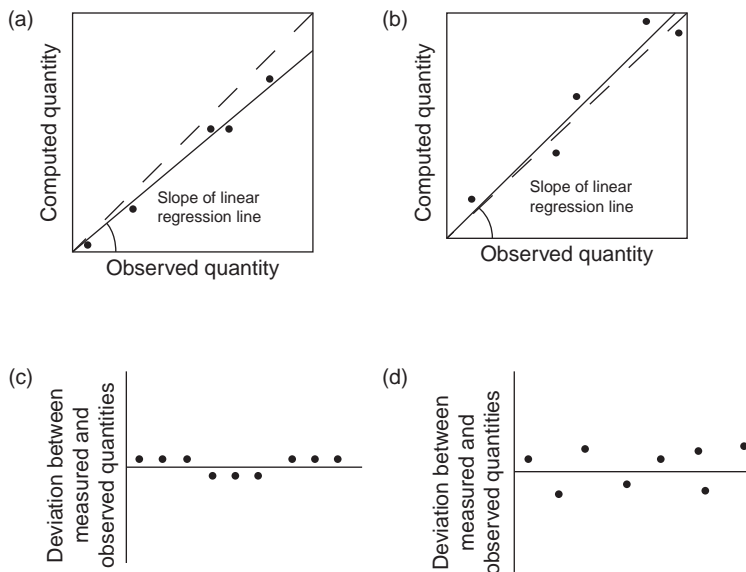


FIGURE 20.22 Basic concepts for the statistical parameters used. (a) Slope has a large deviation but with a reduced data scatter about the fitted line. (b) Slope is very close to ideal value but with an enhanced data scatter. (c) Smaller but a systematic trend of deviations. (d) An almost equal spread of positive and negative but larger deviations.

proposed with varying weighing factors of 3, 1, 1, 1, and 1 for s , r , RMSE, MBE, and MAD, respectively,

$$\Psi = 3[s] + [r] + \left[1 - \frac{\text{RMSE}}{\text{RMSE}_{\max}}\right] + \left[1 - \frac{|\text{MBE}|}{|\text{MBE}|_{\max}}\right] + \left[1 - \frac{\text{MAD}}{\text{MAD}_{\max}}\right] \quad (20.36)$$

Note that s and r are dimensionless, unlike RMSE, MBE, and MAD, and therefore, the latter three are, respectively, divided by the values of RMSE_{\max} , $|\text{MBE}|_{\max}$, and MAD_{\max} . Ψ is a convenient figure of merit, by means of which it is possible to compare the performance of any suite of models. Therefore, for a perfect fit, the overall accuracy score Ψ will be 7.

20.10 OUTLIER ANALYSIS

Often in solar radiation studies, we encounter data that lie unusually far removed from the bulk of the data population. Such data are called “outliers.” One definition of an outlier is that it lies three or four standard deviations or more from the mean of the data population. The outlier indicates peculiarity and suggests that the datum is not typical of the rest of the data. As a rule, an outlier should be subjected to particularly careful examination to see whether any logical explanation may be provided for its peculiar behavior. Automatic rejection of outliers is not always very wise. Sometimes, an outlier may provide information that arises from unusual conditions. Outliers may however be rejected if the associated errors may be traced to erroneous observations due to any one or a combination of factors described in previous sections of statistical analysis.

Statistically, a “near outlier” is an observation that lies outside 1.5 times the interquartile range. The interquartile is the interval from the 1st quartile to the 3rd quartile.

The *near outlier* limits are mathematically defined by:

$$\text{Lower outlier limit : 1st quartile} - 1.5 (\text{3rd quartile} - \text{1st quartile}) \quad (20.37)$$

$$\text{Upper outlier limit : 3rd quartile} + 1.5 (\text{3rd quartile} - \text{1st quartile}) \quad (20.38)$$

Likewise, *far outliers* are defined as the data whose limits are defined below:

$$\text{Lowerlimit : 1st quartile} - 3 (\text{3rd quartile} - \text{1st quartile}) \quad (20.39)$$

$$\text{Upper limit : 3rd quartile} + 3 (\text{3rd quartile} - \text{1st quartile}) \quad (20.40)$$

A high number of outliers in a given data set signify that the observations have a high degree of variability or a large set of suspect data indicating poor station operation. For a more rigorous discussion on outlier analysis, readers are referred to the reference made to Draper and Smith (1998) and Montgomery and Peck (1992).

ACKNOWLEDGMENTS

This contribution has relied heavily on the work of a number of contributions, all of which have been presently referenced. However, the three sources that were of

particular and significant use were the following: *Solar and Terrestrial Radiation* by Coulson (1975), *Solar Radiation and Daylight Models* by Muneer (2004), and photographs of sensors and equipment that were obtained from Delta-T of England, EKO Instruments of USA, Eppley of USA, Kipp and Zonen of Netherlands, and Middleton Solar of Australia.

REFERENCES

- Angus RC. *Illuminance Models for the United Kingdom*. Edinburgh: Edinburgh Napier University; 1995.
- Anon. *Active Solar Tracking System*. [cited November 25, 2010]; Available from: <http://www.middletonsolar.com/products/product12.htm>; 2010a.
- Anon. *Measurement Equipment*. [cited November 25, 2010]; Available from: http://users.du.se/~ffi/SERC/Hybrid/technical_data.htm; 2010b.
- Batlles FJ, A.-A. L, Olmo FJ. On shadowband correction methods for diffuse irradiance measurements. *Solar Energy* 1995;54:105–114.
- Batlles FJ, Barbero J, López, G, Pérez, M, Rodrigo F, Rubio MA. *Fundamentos de la radiación solar y aspectos climatológicos de Almería, 1990–1996*. (in Spanish): Servicio de Publicaciones de la Universidad de Almería; 1998.
- Clima T. *Shadow Ring CM 121 Typ B*. [cited November 25, 2010]; Available from: http://www.thiesclima.com/global_radiation.html; 2010.
- Corporation, N. *240-1070-L Campbell–Stokes Pattern Sunshine Recorder*. 2010 [cited November 25, 2010]; Available from: <http://www.novalynx.com/240-1070.html>.
- Coulson KL. *Solar and Terrestrial Radiation*. New York: Academic Press; 1975.
- Dehne K. *Diffuse Solar Radiation Measured by Shade Ring Method Improved by a Correction Formula*. World Meteorological Organization; 1984.
- D.-T Devices. *Sunshine Sensor—BF3*. 2010 [cited November 25, 2010]; Available from: <http://www.delta-t.co.uk/products.html?product2005092016583>.
- Draper N, Smith H. *Applied Regression Analysis*. New York: Wiley; 1998.
- Driesse A. and Thevenard, D. A test of Suehrcke's sunshine-radiation relationship using a global data set. *Solar Energy* 2002;72:167.
- Drummond AJ. Techniques for the measurement of solar and terrestrial radiation fluxes in plant biological research: a review with special reference to arid zones, in Montpiller Symposium 1965: UNESCO.
- Drummond AJ. On the measurement of sky radiation. *Archives for Meteorology, Geophysics, and Bioclimatology* 1956.B7:413–436.
- Envco. *Albedometer—Global and Reflected Radiation*. 2009 [cited November 25, 2010]; Available from: <http://www.envcoglobal.com/catalog/product/analog-solar-radiation-sensors/albedometer-global-and-reflected-radiation.html>.
- EPLAP. *Normal Incidence Pyrheliometer*. 2010 [cited December 6, 2010]; Available from: <http://www.eppleylab.com/>; 2010.
- ESRA, *European Solar Radiation Atlas*. Paris: Ecole des Mines; 2000.
- ESRA. *European Solar Radiation Atlas 2000. Vol. 1: Fundamentals and Maps*. 4th ed. Ecole des mines de, Paris: Les Presse de l'Ecole. Chapter 3 and 4, 2000. p. 17–40.
- Foster NB, Foskett LW. A photoelectric sunshine recorder. *Bulletin of the American Meteorological Society* 1953;(20.34):212.
- Geuymard C, Kambezidis HD. Solar spectral radiation. In: Muneer T, editor. *Solar Radiation and Daylight Models*. Oxford: Elsevier Ltd; 2004. p. 221–302.

- Gueymard CA. Direct solar transmittance and irradiance predictions with broadband models. Part 1: detailed theoretical performance assessment. *Solar Energy* 2003;(74):355–379.
- Instruments, E. *Pyrgometer: MS-202/MS-202F*. 2005 [cited November 25, 2010]; Available from: http://www.eko-usa.com/products/am/MS-202_MF-11/MS-202/MS-202.html.
- Instruments, E. *Sky Scanner MS-321LR*. 2004 [cited December 8, 2010]; Available from: <http://www.eko-usa.com/products/am/MS-321LR/MS-321LR.html>.
- Kasten F, Dehne K, Brettschneider W. Improvement of measurement of diffuse solar radiation, in *Solar Radiation Data Series F*. 1983. p. 221.
- Kendrick DEA. *Guide to Recommended Practice of Daylight Measurement*. Wein, Austria: International Commission on Illumination (CIE); 1994.
- Kudish AI, Ianetz A. Analysis of diffuse radiation data for Beer Sheva: measured (shadow ring) versus calculated (global-horizontal beam) values. *Solar Energy* 1993;51:495–503.
- LeBaron BA, Michalsky J. J., Perez R. A simple procedure for correcting shadowband data for all sky conditions. *Solar Energy* 1990; 44(20.5):249–256.
- Lof GOG, Duffie JA, Smith CO. World distribution of solar radiation, in *Engineering Experiment Station Report* Madison, USA: University of Wisconsin; 1965.
- Monteith JL. The reflection of short-wave radiation by vegetation. *Quarterly Journal of the Royal Meteorological Society* 1959;85:392.
- Montgomery D, Peck E. *Introduction to Linear Regression Analysis*. New York: Wiley; 1992.
- Muneer T, Zhang X. A new method for correcting shadowband diffuse irradiance data. *ASME* 2002;124:34–43.
- Muneer T, Fairouz F. Quality control of solar radiation and sunshine measurements – lessons learnt from processing worldwide databases. *Building Services Engineering Research* 2002;23(20.3): 151–166.
- NREL. Quality Assessment with SERI_QC. 1993 [cited November 25, 2010]; Available from: http://rredc.nrel.gov/solar/pubs/seri_qc/.
- Observers Handbook*. London: HSMO; 1969.
- Omni Instruments, *Kipp and Zonen retail prize*: Dundee, Scotland.
- Owen F, Jones R. *Statistics*. London: Pitman Publishing; 1990.
- Page JK. *Proposed Quality Control Procedures for the Meteorological Office Data Tapes Relating to Global Solar Radiation, Diffuse Solar Radiation, Sunshine and Cloud in UK*, the London, UK: Chartered Institution of Building Services Engineers; 1997.
- Painter H.E. The Performance of a Campbell–Stokes sunshine recorder compared with a simultaneous record of the normal incidence irradiance. *Met. Mag.* 1981;110:102–87.
- Painter HE. The shade ring correction for diffuse irradiation measurements. *Solar Energy* 1981;26:361.
- Perez R, Ineichen P, Seals R. Modelling daylight availability and irradiance components from direct and global irradiance. *Solar Energy* 1990;44:271.
- Pollard DG, Langevine LP. An anisotropic correction for diffuse irradiance measurements in Guyana. in *Proceedings of the 1988 Annual Meeting*. 1988: ASES Cambridge.
- Rawlins F. The accuracy of estimates of daily global irradiation from sunshine records for the United Kingdom. *Meteorological Magazine* 1984;(20.20.113):187.
- Rozenberg GV. *Twilight – A Study in Atmospheric Optics*. New York: Plenum Press; 1966.
- Stanhill G. Observations of the shade ring corrections for diffuse sky radiation measurements at the dead sea. *Quarterly Journal of the Royal Meteorological Society* 1985;111: 1125–1130.
- Suchrcke H. On the relationship between duration of sunshine and solar radiation on the earth's surface: Ångström's equation revisited. *Solar Energy* 2000;68:417.

- Press WH et al. *Numerical Recipes in FORTRAN: The Art of Scientific Computing*. Cambridge: Cambridge University Press; 1992.
- Wood J, Muneer T, Kubie J. Evaluation of a new photodiode sensor for measuring global and diffuse irradiance, and sunshine duration. *Journal of Solar Energy Engineering* 2003;125 (20.20.1):43–48.
- World Meteorological Organization. *Manual on the Global Observing System*. Geneva; 2003.
- World Meteorological Organization. *Guide to Meteorological Instruments and Methods of Observation*. Preliminary 7th ed. Geneva, Switzerland. 2006.
- Younes S, Claywell R, Muneer T. Quality control of solar radiation data: present status and proposed new approaches. *Energy* 2005;30:1533–1549.

21

WIND ENERGY MEASUREMENTS

PETER GREGG

- 21.1 Introduction
 - 21.1.1 Terms and definitions
- 21.2 Concepts
 - 21.2.1 Power in the wind
 - 21.2.2 Power curve
 - 21.2.3 Micrositing
- 21.3 Measurements
 - 21.3.1 Basic setup
 - 21.3.2 Meteorological towers
 - 21.3.3 Remote sensing devices
 - 21.3.4 Wind speed
 - 21.3.5 Wind direction
 - 21.3.6 Air density
 - 21.3.7 Barometric pressure
 - 21.3.8 Temperature
 - 21.3.9 Humidity
 - 21.3.10 Precipitation
 - 21.3.11 Wind turbine measurements
- 21.4 Evaluation
 - 21.4.1 Data cleaning
 - 21.4.2 Data calculations, correction, and processing
 - 21.4.3 Wind resource assessment
 - 21.4.4 Site suitability, wind turbine performance characteristics, and turbine selection
 - 21.4.5 Estimation of annual energy production
 - 21.4.6 Power curve measurement

References

21.1 INTRODUCTION

Understanding the wind is critical to any wind energy application. The starting point in any wind energy project is the wind resource assessment (WRA) that quantifies the wind resource available at a given location, as well as establishes several other important parameters. The wind resource assessment feeds into the site suitability analysis, which determines the suitability of a particular wind turbine to the local wind conditions. As part of the site suitability process, a mechanical loads analysis (MLA) may be performed to assess turbine hardware constraints against the loading and fatigue expected at the site given the output of the wind resource assessment, thus establishing the suitability of the proposed turbine hardware to the proposed site. The wind is the fuel for a wind turbine power plant and so wind is a critical input in the energy estimation.

21.1.1 Terms and Definitions

- a. *Anemometer*: a device used to measure wind speed.
- b. *Hub Height Wind Speed*: for a horizontal axis turbine, the wind speed at the turbine rotor centerline.
- c. *Mechanical Loads Analysis (MLA)*: a detailed engineering study to calculate the mechanical stresses for a specific wind turbine configuration which factors in the specific atmospheric conditions at a given location.
- d. *Rated Wind Speed*: the wind speed at which a turbine achieves rated power.
- e. *SCADA*: system control and data acquisition.
- f. *Shear*: the gradient of wind speed with height above the ground.
- g. *Stable Atmosphere*: an atmospheric condition where the atmosphere forms thermal layers, characterized by a high temperature gradient, a high wind speed gradient (shear), and low turbulence.
- h. *Turbulence*: departure from smooth, laminar flow or fluctuations in the speed and direction of the flow.
- i. *Turbulence Intensity*: the standard deviation of wind speed divided by the mean wind speed over a 10-min period, which is used as a numerical quantification of turbulence.
- j. V_{Ref} (*Reference Wind Speed*): 50-year extreme 10-min average wind speed.
- k. V_{50} : the 50-year maximum 3 s gust wind speed.

21.2 CONCEPTS

21.2.1 Power in the Wind

The power in the wind available to a wind turbine is determined from the momentum of the mass flowing through the turbine. The mass flow rate is determined by the wind speed, air density, and rotor area. Of course not all of the energy in the wind is converted into electrical energy as that would result in a velocity of zero downwind of the turbine, so a

term is added for the efficiency of the energy capture called the power coefficient, C_p . Thus, the power in the wind is given by the equation:

$$P = 0.5 \times \rho \times V^3 \times A \times C_p \quad (21.1)$$

where P = power, W; ρ = air density, kg/m³; V = wind speed, m/s; A = rotor area, m²; and C_p = power coefficient.

From this equation it can be seen that power is proportional to air density and more important to this discussion, power is proportional to the cube of wind speed. Thus, the sensitivity of the wind speed measurement to our parameter of interest, power, is very high. The accuracy of the wind speed measurement is, therefore, critical for most wind energy applications, and in particular for wind energy assessment and for evaluation of wind turbine performance.

The maximum theoretical efficiency for a wind turbine is the Betz limit, which is 59.3% or 0.593.

21.2.2 Power Curve

The power output of a wind turbine is characterized by its power curve, which is simply the electric power output of the wind turbine plotted against wind speed. Power curves vary by turbine type and technologies employed, but power curves typically have several characteristics:

- *Cutin Wind Speed*: the minimum wind speed for the turbine to produce power, typically over 3 m/s.
- *Cutout Wind Speed*: the storm shut down wind speed above which the turbine does not operate for safety reasons.
- *Rated Wind Speed*: the wind speed at which the turbine achieves rated power.
- *Rated Power*: this is typically the maximum output of the wind turbine. Power is proportional to the cube of wind speed; however, a wind turbine limits its maximum power output due to limits of the turbine hardware.

There are two strategies commonly employed to limit the power output of Horizontal Axis wind turbines so that it does not exceed the turbine's rated power. These are active pitch control and stall control.

Active pitch-controlled turbines pitch the blades to reduce the aerodynamic efficiency once rated power is reached. The blades can be pitched to maintain a constant power output once rated power is reached. Figure 21.1 shows an example of a power curve for an active pitch-controlled turbine.

Stall-controlled turbines typically have fixed blades that do not pitch. Instead the blades are designed such that above a certain wind speed, flow separation occurs sending the blades into aerodynamic stall, thus reducing the efficiency of the turbine. The power output of stall-controlled turbines usually decreases once rated power is reached because the stall becomes more severe with increasing wind speed. Figure 21.2 shows an example power curve for a stall-controlled turbine.

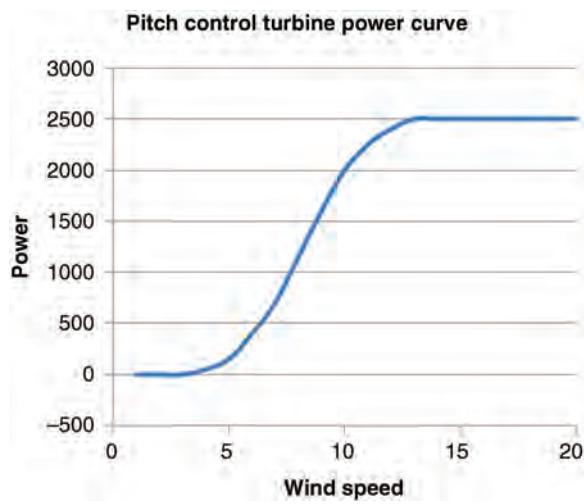


FIGURE 21.1 Example pitch control turbine power curve.

21.2.3 Micrositing

Micrositing is the process of locating and spacing wind turbines to optimize the power production, maintain loads within the turbine design limits, and balance the number of turbines with land constraints. Micrositing takes several factors into account, including

- *Terrain Effects:* ideally turbines will be located in positions that maximize power production. For example, it is best to locate a wind turbine at the top of a hill or ridge line so that it is exposed to the most wind and takes advantage of the natural speedup effect as the wind moves over the ridge.

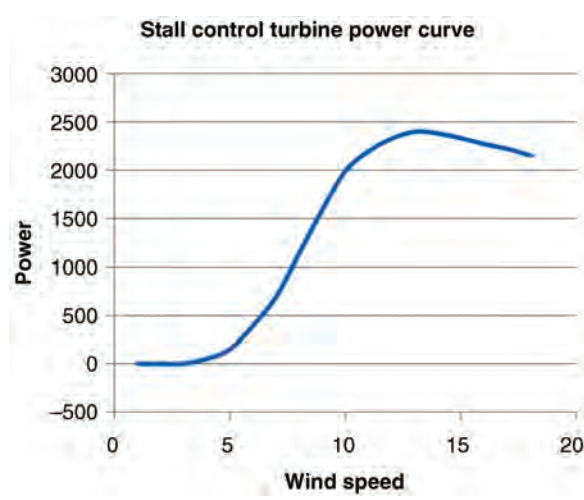


FIGURE 21.2 Example stall control turbine power curve.

- *Layout and Turbine Spacing*: upwind turbines remove energy from the wind for downwind turbines to capture, called waking. Also turbine wakes increase turbulence, which can impact the mechanical loads on the downwind turbines. Ideally turbines should be laid out in rows that are perpendicular to the wind and spaced far apart to allow for more energy capture and minimal wake effects.
- *Turbine hub height*: for horizontal axis turbines, the hub height may be optimized to capture the most energy. In general, taller turbines capture more energy; however, there are rare situations where lower hub height turbines are more desirable, such as along a ridge line. Also there is more cost associated higher hub heights and there may be loads limitations.
- *Setbacks to roads and residences*: this mainly driven by noise and safety concerns such as ice shedding.

21.3 MEASUREMENTS

21.3.1 Basic Setup

A typical measurement involves a meteorological tower with sensors at different heights. At a minimum, wind speed and wind direction are measured near the top of the tower. These should be measured at the hub height of the turbine. If this is not the case, then shear must be measured to allow the wind speed to be extrapolated to the turbine hub height. In order to measure wind shear and veer, wind speed and wind direction measurements should be made at lower elevations. Temperature and pressure are measured to record the site air density, and also the temperature measurement is useful for sensor icing detection. A precipitation sensor may also be included.

21.3.2 Meteorological Towers

There are several types of meteorological towers commonly used. Each has its advantages and disadvantages. The most common types are discussed below.

21.3.2.1 Tilt-Up Towers Tilt-up towers are low cost and relatively easy to transport and install. Typically, a tilt-up tower is a tubular design supported by guywires. Also a tilt-up tower may be mounted on a base plate staked into the ground, not requiring a concrete foundation. The segments are bolted together with the tower lying horizontally on the ground. Booms are added and the sensors are mounted while the tower is still on the ground. The tower then is winched up to a vertical position and the guywires are tightened to secure it. Tilt-up towers are popular for wind resource assessment because they are low cost, easy to transport and set up in remote locations, and do not require a crane to install. The main drawbacks to tilt-up towers are that they are limited in height, typically to not more than 60 m, and if a sensor fails the tower must be lowered to replace it.

21.3.2.2 Guyed Lattice Towers Guyed lattice towers are very sturdy and may be erected to very high heights. Very tall towers used for research (up to 200 m) are guyed lattice. They are erected vertically in place using either a crane or a gin-pole. Typically, the



FIGURE 21.3 Guyed lattice tower.

tower and guywire anchors require concrete foundations. If a sensor fails, a guyed lattice tower may be climbed to replace it. Figure 21.3 shows an example of a guyed lattice tower.

21.3.2.3 Freestanding Towers The main advantage of freestanding tower is that it has a smaller footprint than a guyed lattice tower while it is still able to reach the hub heights of modern large wind turbines (typically 60–100 m). This makes it ideal for use in farmland where the guywires and guywire anchors would otherwise be an obstacle to farm equipment, making the land around the tower base unusable. The main drawback is that the smaller footprint comes at a higher cost relative to guyed lattice towers. Figure 21.4 shows an example of a freestanding meteorological tower.

21.3.3 Remote Sensing Devices

Remote sensing devices are becoming more common with recent advances in remote sensing technology. The most common remote sensing devices are LIDAR and SODAR. Both technologies operate on the same principles

Beams of energy (light in the case of LIDAR and sound in the case of SODAR) are shot up into the atmosphere at various angles and directions forming a cone. The energy



FIGURE 21.4 Freestanding meteorological towers.

is reflected back to the device either by particles in the air or from atmospheric turbulence. The device then uses Doppler effect and determines the wind speed, direction and flow angle from the measurements at different points in the cone at several predetermined heights. See Figure 21.5 for an illustration.

The main advantage of this technology over met tower measurements is that the devices are portable, very easy and fast to set up, and may not require any permitting. Note that meteorological towers require extensive permitting for installation, which can be costly and time consuming.

The advantages and disadvantages will vary by device and manufacturer. In general, LIDAR provides data up to very high heights (200 m) with very low data loss rates; however, LIDAR also tends to have a very high power consumption and may require an external power source such as a generator or connection to the power grid. SODAR will also provide measurements up to similar heights; however, SODAR tends to have higher data loss rates at the higher heights. However, the data quality at the typical range of hub heights for large wind turbines, 60–100 m, is generally very good. SODAR typically has lower power consumption than LIDAR, and there are SODAR units commercially available that are self-powered via solar panel.

21.3.4 Wind Speed

The most important measurement and also the most sensitive is wind speed. Power is proportional to the cube of wind speed, so relatively small errors in wind speed translate to large errors in power/energy.

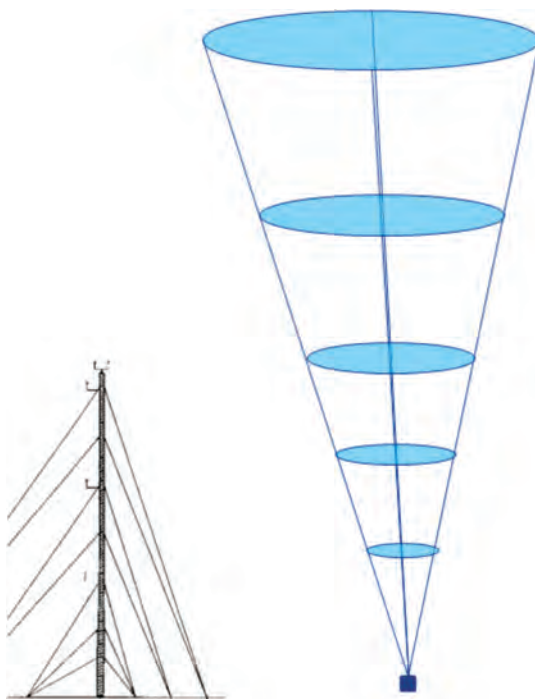


FIGURE 21.5 Remote sensing device versus a meteorological tower.

Devices that measure wind speed are called anemometers. Considerations when selecting an anemometer include the device accuracy (including operational response to turbulence, flow inclination, and temperature effects), durability for long-term field use, and resistance to icing and cost. The most common types will be discussed in the following sections.

21.3.4.1 Cup Anemometers Cup anemometers are the most common anemometer type in use. A cup anemometer consists of three cups mounted on a horizontal rotor. When the wind blows, the cups catch the wind, spinning the rotor. The rotor shaft is connected to an electronic sensor that typically registers a certain number of pulses each time the shaft rotates. The wind speed is determined by the number of pulses counted over a time interval, typically 1 s.

Cup anemometers typically have very low power consumption, making them ideal for remote applications where power supply is limited. Also, cup anemometers are usually designed to measure the horizontal wind speed, not a vector component, so the device measures the same regardless of the direction that the wind is coming from.

Cup Anemometer Operational Characteristics Cup anemometers are mechanical systems, and their accuracy will change in different wind conditions depending on the anemometer aerodynamics, rotor inertia, and other operational characteristics.

The IEC 61400-12-1 contains a classification scheme to determine the accuracy of cup anemometers in different environmental conditions. Three factors in particular stand out

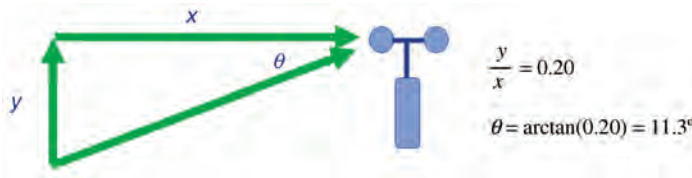


FIGURE 21.6 In-flow angle.

Upflow: Wind turbine performance is defined in terms of the horizontal wind speed component, as that is what the turbine converts to energy. Most cup anemometers are designed to measure the horizontal wind speed component; however, their ability to do this at increasing flow angles depends on the anemometer aerodynamics and differs by make and model. See Figures 21.6 and 21.7 for an example of how anemometer accuracy varies with inflow angle. At the example inflow angle of 11° , this anemometer would have an error of about 0.5% versus the ideal cosine response. This may not seem significant, but recall that power is proportional to the cube of wind speed.

Turbulence: A good illustration of the impact of turbulence on a wind sensor is the pinwheel, which is essentially a propeller on a stick. When you blow into it the propeller will spin but it does not stop spinning right away when you stop blowing. When talking about an anemometer, this is called over-speeding and causes the average wind speed recorded to be higher than the actual wind speed over time. The response time of an anemometer depends on the rotor inertia and aerodynamics, and so it will vary for different makes and models.

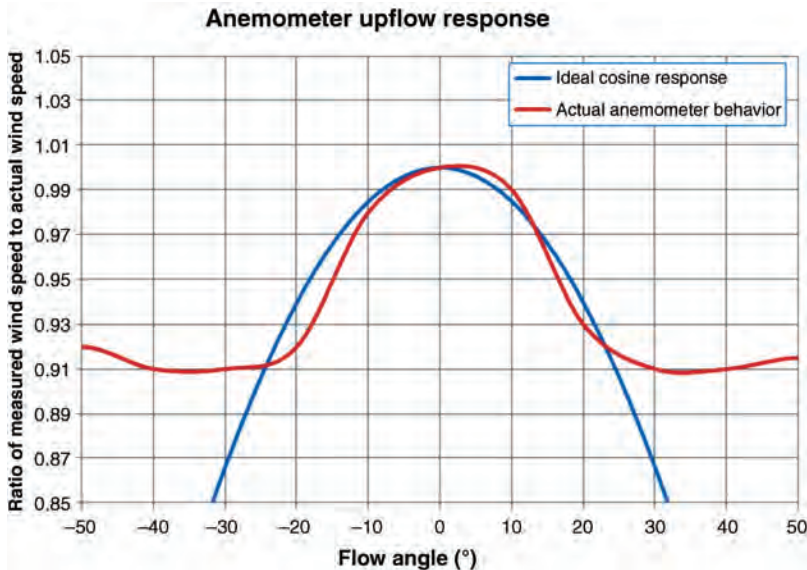


FIGURE 21.7 Anemometer upflow response.

Temperature: The friction characteristics of the anemometer bearings will change with temperature. The degree to which they change will impact the anemometer accuracy at those temperatures.

Note that this is different from icing of the anemometer cups, which will have a significant impact on the anemometer. Data where the device is believed to be iced are normally filtered out. Some anemometers have heated cups to prevent icing.

Also note that the sensor accuracy may change over time as the sensor bearings wear or if the cups become damaged.

The IEC 61400-12-1 provides a procedure to classify anemometer operational characteristics through controlled wind tunnel and laboratory tests. The anemometer classification report can be used to select a device with sufficient accuracy at the expected local conditions.

21.3.4.2 Ultrasonic Anemometers Ultrasonic anemometers measure the wind speed with sound waves and using the Doppler effect. Unlike SODAR that measures the wind at a distance, ultrasonic anemometers measure the wind as it passes through a gap or open area in the device.

Ultrasonic anemometers are either two-dimensional, measuring the two horizontal wind speed components, or three-dimensional, measuring the two horizontal components and the vertical component of the wind speed.

Since Ultrasonic anemometers measure the wind speed vectors, they also measure wind direction.

Ultrasonic anemometers do not have moving parts that may wear out over time; however, they do have higher power consumption than cup anemometers. Also the geometry of the device may cause flow distortion, which is different depending on the wind direction, for example, if the wind is along the axis of the device that is obstructed by the sensor prongs. The device may or may not compensate for this.

21.3.4.3 Propeller Anemometers There are some propeller anemometer devices commercially available; however, these are generally not considered to have sufficient accuracy for wind power measurements. If the propellers are fixed (do not move with wind direction), then the off-axis measurements will not have sufficient accuracy. If the propeller is mounted on a wind vane so as to keep the propeller pointed into the wind, then the flutter of the wind vane will interfere with the wind speed measurement.

21.3.4.4 Pitot Tube Devices Pitot tube devices are rarely used in wind energy applications. Typically, they are not very accurate at low wind speeds, they have the same issues as propeller anemometers with respect to orientation to the wind, and the narrow tubes are not well suited for outdoor, long-term measurement applications where they are exposed to dust, insects, and precipitation, which may clog the tubes.

21.3.4.5 Remote Sensing Devices Remote sensing devices previously discussed offer an attractive alternative to meteorological tower mounted sensors in some applications. It should be noted that remote sensing devices such as SODAR and LIDAR do not measure the same thing as meteorological tower mounted sensors. Meteorological tower mounted sensors measure the wind speed and direction at a point location, while a remote sensing device measures the wind speed and direction over an area. Although the results are

similar, this makes it difficult to compare them directly. This is especially evident in the turbulence intensity measured by a remote sensing device compared with that measured by a met tower. A remote sensing device may not be as well suited to long-term measurement than a met tower. Also remote sensing devices may not work in certain conditions. This limits the ability to determine the extreme wind conditions with these devices. At the time of print of this chapter, the industry consensus is for most applications remote sensing works best to compliment meteorological tower data rather than replace it. The intended use of the data must be considered when deciding whether to employ a remote sensing device for replacing met tower data.

21.3.5 Wind Direction

Wind direction measurements are typically made with a wind vane. Most wind direction sensors consist of a wind vane mounted on a shaft and connected to a potentiometer mounted in the sensor body. Thus, the resistance of the potentiometer is a function of the orientation of the wind vane. Note that there may be a small dead-band, typically at zero degrees, where the device passes through a gap in the potentiometer.

Wind direction may also be measured by the wind speed sensor, as discussed for ultrasonic sensors and remote sensing devices.

Two items worth discussing for wind direction measurement are the accuracy of the vane orientation and north jump.

21.3.5.1 Wind Direction Sensor Orientation Critical to the accuracy of the wind direction measurement is to know which way true north is relative to the sensor north. All wind direction sensors should have a zero degree mark or some indication of what the device considers zero degrees to be.

For a boom-mounted device, a best practice is to mount the device such that its zero degree mark is in line with the boom. Once installed, the direction of the boom can be determined from the ground using binoculars and a GPS or compass. The direction of the boom relative to north is the offset that must be applied to the wind direction signal.

Note that when using a compass, the measured boom direction must be corrected for magnetic declination; that is corrected from magnetic north to true north. The correction for magnetic declination can be as high as 15° or more for many locations within the continental United States. The National Oceanic and Atmospheric Administration (NOAA) has information on magnetic declination on its website, www.noaa.gov.

21.3.5.2 North Jump North jump is an error introduced by averaging. If over a 10-min period, the wind vane spends on average 5 min at 355° and 5 min at 5° , then the wind is clearly from the north, but the average of 355° and 5° is 180° , which is from the south. North jump is typically corrected for in the data logger program by using a vector average. It is difficult to correct by postprocessing the data.

North jump is easily detected by plotting wind direction standard deviation versus wind direction. If the data have not been corrected properly for north jump, a rainbow is observed with its peak at 180° . North jump may also be detected by plotting one wind vane against another (see Figure 21.8).

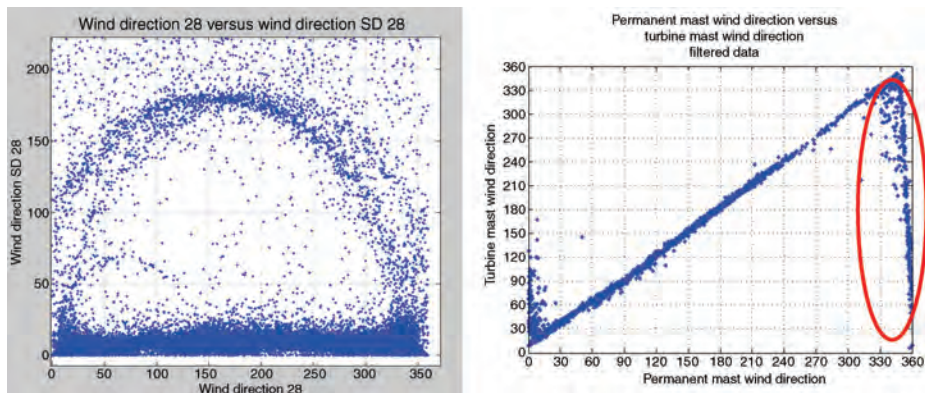


FIGURE 21.8 Detecting north jump.

21.3.6 Air Density

Power in the wind is proportional to air density. Air density varies with altitude, with weather patterns and seasonally. Over the course of a year, the air density at a site may vary by 8% or more from the average value. Air density is also an input to loads calculations. Therefore, it should be included in the measurement campaign to characterize the site and to allow for correction of the data to a reference air density where applicable.

Air density is not measured directly but is calculated from temperature and barometric pressure. Note that humidity is also a factor in the calculation of air density. The impact is small for low temperatures but can be significant at high temperatures, especially in humid climates.

21.3.7 Barometric Pressure

The barometric pressure measurement should be made at hub height if possible; however, if it is measured at a different elevation, the measured value may be corrected for elevation using the method in ISO 2533 (ISO standard atmosphere).

21.3.8 Temperature

Temperature may be measured with any number of temperature sensing devices. Resistance temperature devices (RTDs) are most common.

Since the measurements are made outdoors and exposed to the elements, the temperature sensor should be mounted inside a radiation shield to prevent the sun from heating up the device directly, which will bias the measurement. Also the temperature sensor should be mounted away from any walls, vents, enclosures, or surfaces that will absorb energy from the sun, raising the temperature in the immediate vicinity of the surface (including the ground).

Unlike pressure, which has a consistent gradient with altitude, the temperature gradient varies significantly with atmospheric stability. Therefore, the temperature measurement should be made at the top of the tower or as close to hub height as possible. Placing additional temperature sensors at lower altitudes on the meteorological tower will help to capture information about atmospheric stability.

21.3.9 Humidity

Humidity has a small influence on air density, but the impact can become significant, especially in hot and humid climates. It is therefore recommended that humidity be measured in hot and humid climates; however, in many applications it may be ignored or an assumed reference value used with minimal error to the air density calculation.

21.3.10 Precipitation

In some cases, precipitation may be a desired measurement. Although not used in any calculations, it may be a useful reference parameter. The type of sensor should be selected based on the intended use of the data.

Simple Boolean type sensors provide a signal indicating whether or not precipitation is detected; however, they typically do not provide information about the type or quantity of precipitation. These sensors will typically work for all types of precipitation and may be useful for detection of sensor icing.

A tip-bucket sensor captures information about the quantity and intensity of precipitation; however, these devices may not work for snow and ice.

21.3.11 Wind Turbine Measurements

If an assessment of an actual wind turbine is being made, a minimum of two measurements needs to be made at the turbine: the turbine status or availability and the electric power being produced by the wind turbine.

The turbine status simply needs to indicate whether the turbine is ready and available to produce power or if it is shut down or faulted. This is often a simple Boolean signal and may be obtained using a relay to monitor the main turbine generator circuit breaker or other suitable location.

Power may be output by the turbine control system, or SCADA, or from a wattmeter if a power signal is not available or a higher degree of accuracy is desired. Note that current transformers may be necessary, particularly for larger wind turbines and should be instrument class transformers of accuracy class 0.5 or better. Potential transformers are often not needed as the voltage is often within the range of most wattmeters, even for large wind turbines. However, if voltage transformers are used, they should also be instrument class devices with an accuracy class of 0.5 or better, depending on the accuracy desired.

21.4 EVALUATION

21.4.1 Data Cleaning

The wind data must be cleaned such that invalid data are filtered out. At a minimum, the data should be filtered for the following

1. Icing of the sensors
2. Sensor failure
3. Sensor in the wake (or “shadow”) of the met tower or of another sensor

Other filters may be applied, depending on the intended use of the data.

21.4.2 Data Calculations, Correction, and Processing

21.4.2.1 Method of Bins Wind data are evaluated using the method of bins. Each data point is sorted into wind speed bins, typically the bin size is 0.5 m/s, but bins of 1 m/s are also used.

The data is also sorted into wind direction bins. The wind direction bin size varies depending on the application. For wind resource assessment, bins of 30° are standard. For power curve measurement, a bin size of 10° is used if the terrain is complex, as the data are corrected for flow distortion due to terrain based on wind direction and more granularity is needed.

21.4.2.2 Air Density Calculation Air density may be calculated from the temperature, pressure, and humidity may be calculated from equation (x) below. If the humidity is not measured, a reference humidity typical for the site may be used.

$$\rho_{10 \text{ min}} = \frac{1}{T} \left(\frac{B}{R_{\text{air}}} - \phi P_w \left(\frac{1}{R_{\text{air}}} - \frac{1}{R_w} \right) \right) \quad (21.2)$$

where ρ = air density, kg/m³; T = temperature, K; B = barometric pressure, hPa; ϕ = relative humidity, ratio; R_{air} = gas constant for dry air (21.287.05 J/kgK); R_w = gas constant for water vapor (21.461.5 J/kgK); and P_w = water vapor pressure.

$$P_w = 0.0000205 e^{0.0631846 \times T} \quad (21.3)$$

21.4.2.3 Turbulence Intensity For each 10-min data point, the turbulence intensity is calculated as the standard deviation of the wind speed divided by the mean.

21.4.2.4 Wind Shear Wind shear is the gradient of wind speed with height above the ground. It is calculated from the wind speed measured at different heights for each data point. There are different ways of modeling and expressing wind shear. The most commonly used is the power law, equation (x).

$$v_h = v_{\text{ref}} \left(\frac{h}{h_{\text{ref}}} \right)^a \quad (21.4)$$

where v_h = velocity of wind at height h ; v_{ref} = velocity of wind at reference height h_{ref} ; and A = shear exponent (aka Hellman exponent).

Shear is often a function of atmospheric stability. Atmospheric stability often follows a diurnal pattern of unstable atmosphere during the day and stable atmosphere at night. During the day, the sun heats the ground which causes thermal rises, mixing and higher turbulence, which results in low shear. At night the atmosphere forms thermal layers with a high temperature gradient, high wind shear, and low turbulence. For example, during the day a site might have a shear exponent of 0.2 or less while at night it may be as high as 0.4 or more.

21.4.2.5 Extrapolation of Wind Speed to Hub Height If the wind speed is measured at a height other than the hub height of the turbine, the wind speed at the turbine rotor

centerline may be interpolated or extrapolated using the power law and the shear exponent. This should be done for each 10-min data point rather than adjusting the averaged results because the shear conditions will vary significantly and there may be a correlation between wind speed and certain shear conditions.

21.4.3 Wind Resource Assessment

The goal of a Wind Resource Assessment is to quantify the wind resource and to determine the inputs for the site suitability analysis. It characterizes the frequency of wind speed, the frequency of wind direction, the average site air density, the site wind shear characteristics, and the turbulence intensity profile. The data are also used to estimate 50-year maximum 10-min windspeed (V_{Ref}) and the 50-year maximum 3-s gust (V_{50}).

Wind conditions vary seasonally and so a minimum of 1 year of data is required for an accurate wind resource assessment; however, there are longer multiyear trends in weather patterns and wind conditions and so in general, more is better.

Wind data are captured at one or more locations at or near the intended wind turbine or wind farm location. For larger wind farms, more data collection locations are desirable.

The WRA data are then extrapolated to the individual turbine locations factoring in the influence of terrain, surface roughness, wakes from neighboring turbines, and other factors.

The output of the WRA is a necessary input to the site suitability analysis, which analysis the design life of a specific turbine to the site conditions, and is used to estimate the expected energy production.

21.4.3.1 Wind Speed Distribution The speed distribution is determined from the count of the occurrence of each wind speed bin. The result is then normalized to a frequency of occurrence (percent of time spent in each bin) or normalized to 1 year and expressed, as the number of hours spends in each bin (hours per year). This is expressed as a table of values, with wind speed bin in one column and hour or percent in the other.

Instead of a table of values, the data may be fitted to a Weibull distribution, which is characterized by a shape factor (k) and a scale factor (A) (see Figure 21.9).

The probability distribution function for a Weibull distribution is

$$F(V, A, k) = \frac{k}{A^k} V^{k-1} e^{-(V/A)^k} \quad (21.5)$$

where V is the wind speed; A is the Weibull scale factor; and k is the Weibull shape factor.

A site's wind distribution is sometimes referred to as its Weibull. Also it is sometimes useful to calculate the annual average wind speed from the distribution as a reference.

The wind speed distribution is calculated including all wind directions. A wind distribution is also calculated for each wind direction bin.

21.4.3.2 Wind Rose The wind direction frequency is characterized by the wind rose. The frequency of occurrence for each wind direction bin is determined. The data are typically plotted on a radar-graph. The result looks the petals of a flower, which give the wind rose its name (see Figure 21.10).

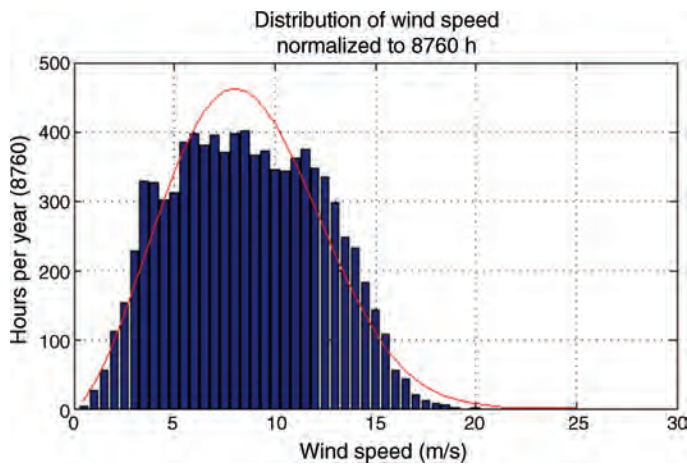


FIGURE 21.9 Example wind distribution.

The wind rose may also be expressed as an energy rose, which combines information about the wind speed in each wind direction bin to provide a further level of detail.

21.4.3.3 Average Air Density The average air density is determined from the data. It is usually not necessary to subdivide air density by wind speed or measurement sector.

21.4.3.4 Turbulence Intensity For each 10-min data point, the turbulence intensity is calculated as the standard deviation of the wind speed divided by the mean. This data are then sorted into wind speed bins and averaged, giving a profile of turbulence intensity versus wind speed. This is typically not broken down by wind direction, but this may be useful to do if the terrain is significantly different on one side of the site.

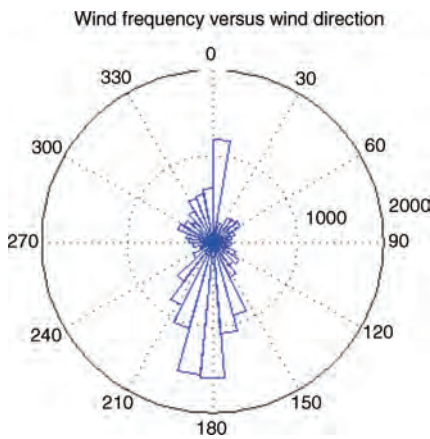


FIGURE 21.10 Example wind rose.

Note that the IEC 61400-1 defines a reference turbulence intensity profile. If a site is referred to as having a specific turbulence intensity value this usually means the turbulence intensity at 15 m/s only, for example, if a site is said to have a turbulence intensity of 12%, this means that the turbulence intensity at 15 m/s is 12%. The turbulence intensity at other wind speeds will be different. Also note that the actual site turbulence profile may be significantly different from the IEC reference profile. The standard also defines a characteristic Turbulence Intensity (CTI) as the mean Turbulence intensity plus once standard deviation.

21.4.3.5 Wind Shear The wind shear conditions for the site are determined. Shear conditions at a site are often described by the value of the shear exponent, for example, a site may be said to have an average shear exponent of 0.2. The average shear exponent is reported.

21.4.3.6 V_{Ref} The 50-year maximum 10-min average wind speed (sustained gust wind speed, also called V_{Ref}) is determined statistically from the wind speed data. V_{Ref} is an important parameter for the site suitability analysis. A wind turbine manufacturer will specify a V_{Ref} for a turbine, which is the maximum sustained wind speed for which the turbine is designed for. Thus, this is an important value to know for selecting a turbine.

As a rule of thumb, the IEC 61400-1 standard suggests that V_{Ref} is five times the annual average wind speed. This is a very rough approximation and may be somewhat limiting. There are other methods to calculate V_{Ref} statistically from the wind distribution data collected, for example, by using the method of independent storms.

Note that the extreme winds are associated with extreme weather events, specifically storms and hurricanes. It is, therefore, very important to have a long enough time series that captures a sufficient number of extreme events in order to make a good estimate of V_{Ref} .

21.4.3.7 V_{50} The 50-year maximum 3-s gust wind speed, also known as V_{50} , is another important factor for assessing the site suitability of a wind turbine and is compared with the survival wind speed of the turbine. It is calculated from the wind data similar to V_{Ref} . The IEC 61400-1 suggests that V_{50} be approximated as seven times the annual average wind speed, but as with V_{Ref} a better result may be obtained statistically from the data.

21.4.3.8 Correlation to a Known Reference Wind patterns do not just follow seasonal trends, but also multiyear trends as weather patterns change and follow longer term cycles. For example, the phenomenon known as El Niño affects weather patterns globally and occurs about once every 7 years. Therefore, it is preferable to have data sets longer than 1 year when conducting a wind resource for a site as that will provide some insight into the multiyear wind trends. Of course this is not always practical.

To estimate the longer term trends, it may be useful to correlate the wind data collected for the site to a nearby reference that goes back several years, such as a weather station at a nearby airport. For example, if the annual average wind speed at the reference station for the same time period is lower than from previous years, then it is likely that the winds at the site are lower as well. This is useful in adjusting estimates of annual energy production and also in estimating V_{Ref} and V_{50} .

21.4.3.9 Wind per Turbine If multiple wind turbines are to be installed, the data collected at the meteorological towers may be used to predict the wind conditions at each turbine, factoring terrain and surface roughness effects as well as losses due to turbines being waked by other turbines. This is typically done using computer software.

Note that adjustments must be made to factor in higher turbulence for turbines, which will be waked by other turbines.

21.4.4 Site Suitability, Wind Turbine Performance Characteristics, and Turbine Selection

Wind turbines come in many different sizes and designs. There are vertical axis turbines, horizontal axis turbines, small wind turbines and utility scale wind turbines, different designs varying numbers of blades, and so on. However, regardless of the design being deployed there are several concepts all designs have in common.

Wind turbine design requirements are specified by the international standard IEC 61400-1 Wind Turbines—Part 1: Design Requirements. Small wind turbines are also covered under IEC 61400-2. This document gives standard “engineering and technical requirements to ensure the safety of the structural, mechanical, electrical and control systems” of wind turbines. Of particular interest to this discussion is Section 21.6, which discusses the classification of wind turbines based on wind conditions. A turbine will typically be rated for an IEC wind class (I, II, or III) and for reference turbulence intensity (A or B). A turbine will also be rated for a maximum sustained wind speed, V_{Ref} , and a survival wind speed, V_{50} .

The results of the WRA are compared with the turbine class and turbine rated conditions to establish site suitability for the turbine. However, it is important to note that the actual site conditions will deviate significantly from the reference conditions specified by the IEC for turbine rating. This may allow for some flexibility in the turbine rated conditions. For example, V_{Ref} for a turbine is calculated assuming sea level air density. If the site is at a lower air density, a higher V_{Ref} may be permissible. A detailed MLA following the requirements of IEC 61400-1 may be used to verify the suitability of the turbines for the site wind conditions and planned site layout. The MLA estimates fatigue and extreme loads that the turbine will encounter at the specified location and compares them to the design conditions.

21.4.5 Estimation of Annual Energy Production

A wind turbine manufacturer will typically provide a power curve that characterizes the performance of that wind turbine versus wind speed. The y-axis of the curve is power, and the x-axis is wind speed, usually in 0.5 m/s increments. Note that from Equation 21.1, power is also proportional to air density, so the power curve will specify a reference air density that it is valid for.

For horizontal axis turbines, the wind speed is defined as the wind speed at the turbine rotor centerline, called the hub height wind speed. Note that for large wind turbines, the wind speed likely varies across the plane of the rotor and so wind speed is defined at a common location that will be most representative of the average wind speed across the entire rotor plane for most cases.

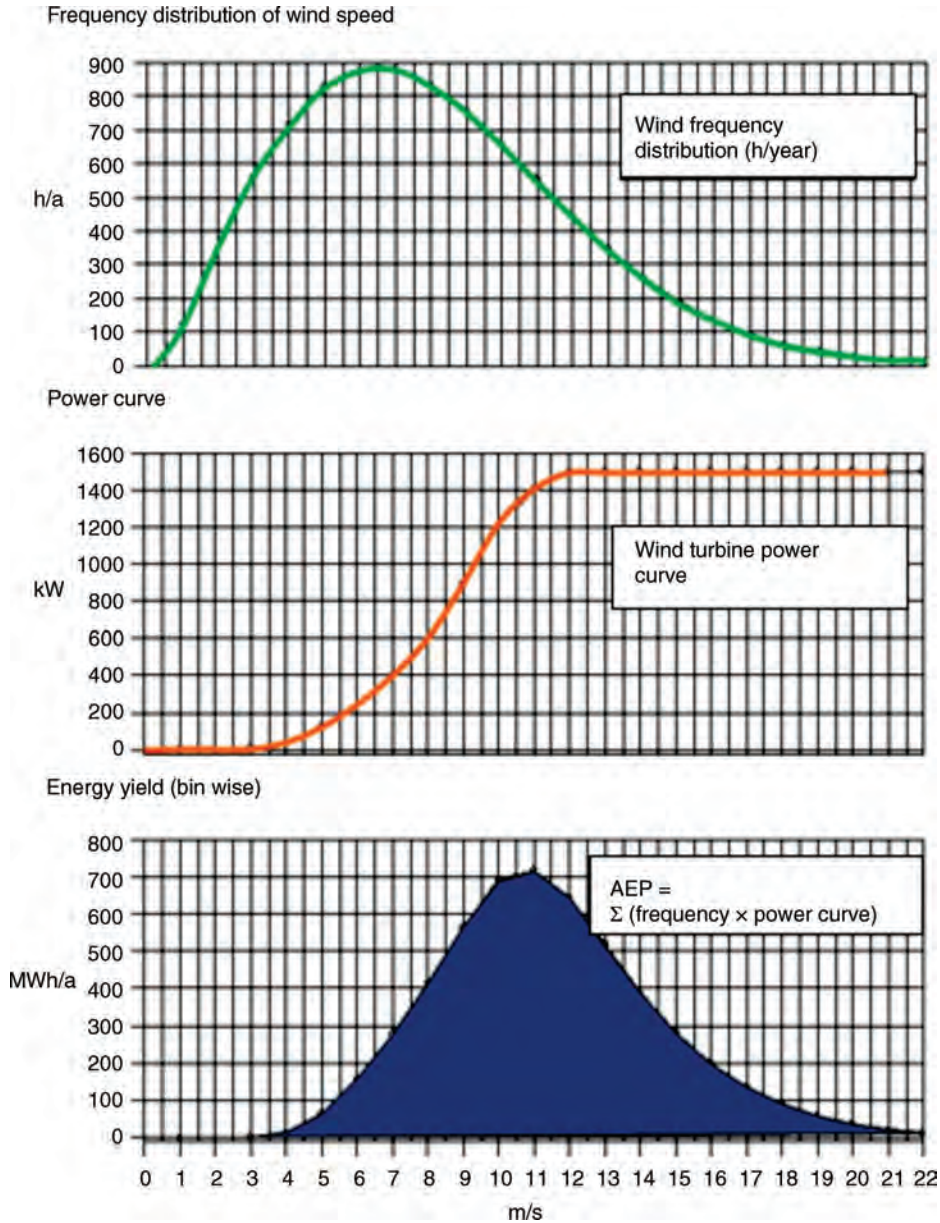


FIGURE 21.11 Calculating annual energy production.

The annual energy production (AEP) may be estimated by multiplying the power in each wind speed bin from the power curve by the frequency in each wind speed bin from the WRA expressed in hours-per-year, and summing the result (see Figure 21.11).

The AEP should be calculated using a power curve that is valid for the site average air density.

If estimating expected revenue generation, the AEP should be adjusted for turbine down time for maintenance, repair and grid and weather-related outages, transmission

and transformer losses, and other miscellaneous losses. If multiple turbines are planned, then the wind per turbine data should be used as this will include estimates for wake losses and the local terrain and roughness effects for each turbine.

21.4.6 Power Curve Measurement

If desired, the power curve of the wind turbine may be verified by measurement. The IEC 61400-12-1 provides a detailed procedure for power curve measurement; however, in general the steps are as follows.

The power curve is the electrical output of the turbine verses wind speed. It is typically not practical to measure the wind speed at the turbine location because there is significant flow distortion due to the turbine rotor and nacelle, which means that the wind speed measured at the turbine will not be reflective of the true free-stream wind speed. Therefore, wind speed must be measured at a meteorological tower away from the turbine, but too far away from the turbine and the winds measured at the met tower may not be the same as those measured at the turbine. This is particularly a problem for larger turbines because distance involved and the fact that wind speed data are recorded in 10-min intervals means that a gust at the met tower several 100 m upwind of the turbine may take several minutes to reach the turbine. The IEC 61400-12-1 specifies that the met tower must be between 2 and 4 rotor diameters from the turbine being tested.

Since the wind speed is not being measured at the turbine rotor centerline but at a location some distance away, flow distortion due to terrain may have an impact on the accuracy of the wind speed. For example, if the turbine is on top of a hill and the met tower is on the bottom, the wind conditions at the met tower will not be the same as the wind conditions at the turbine. A terrain assessment is necessary to assess the potential impact of the local terrain around the met tower and turbine. A similar assessment must be taken to identify any obstacles such as buildings, trees, or other structures will also influence the result.

If the terrain analysis shows that the wind at the met tower cannot be assumed to be the same as the wind at the turbine location, a site calibration may be undertaken to correct for this effect. In a site calibration, a met tower is erected at the turbine location prior to turbine construction, and another met tower is erected at the measurement location. Data are collected, and corrections for flow distortion are empirically developed from the data and based on wind direction.

For the power curve test, a met tower is erected between 2 and 4 rotor diameters from the test turbine (or the permanent met tower from the site calibration is used). Data are collected until sufficient data have been collected to sufficiently characterize the power curve of the wind turbine. Note that in order to characterize the power curve across the full range of wind speeds, data must be collected at all of the wind speeds of interest.

The data are filtered as follows:

1. Wind directions for which neither the turbine nor met tower are in the wake of a neighboring, operating wind turbine
2. Failure or icing of the wind sensors
3. Data when the wind turbine is shut down and unable to operate are removed
4. Other conditions may be filtered out as needed.

The power available to the wind turbine is a function of the air density (see Eq. 21.1) so the data should be density corrected to correct for density variations over the course of the test, and also to correct to the air density of the reference power curve, if applicable.

The filtered and density corrected data set are then sorted using the method of bins into 0.5 m/s wind speed bins. The data in each bin are averaged to give an average wind speed and power for that bin. The result is the measured power curve for the wind turbine. Note that the average wind speed for the bin may not necessarily be the exact bin center.

It is usually not practical to collect data for every single wind speed bin for which the turbine is rated to produce power, as the very high wind speed bins will only occur during rare storm events. If the measured power curve is being compared with a reference power curve or is used for energy production estimates, extrapolation of the power curve for the missing bins may be necessary. Not doing so assumes that the turbine produces zero power in the bins for which no data were collected, which may not be the case.

REFERENCES

- IEC 61400-1: Wind Turbines—Part 1: design requirements; 2005. <http://www.iec.ch>
- IEC 61400-2: Wind Turbines—Part 2: design requirements for small wind turbines. <http://www.iec.ch>
- IEC 61400-12-1: Wind Turbines—Part 12-1: power performance measurements of electricity producing wind turbines. <http://www.iec.ch>
- ISO 2533: Standard atmosphere.
- Manwell JF, McGowan JG, Rogers AL. *Wind Energy Explained*. Wiley; 2010 (Feb 23).

22

HUMAN MOVEMENT MEASUREMENTS

RAHMAN DAVOODI

- 22.1 Introduction
- 22.2 Characterization of human movement
- 22.3 Optical motion capture systems
- 22.4 Magnetic motion capture systems
- 22.5 Inertial motion capture systems
- 22.6 Discussion
- Acknowledgment
- References

22.1 INTRODUCTION

Making movements and detecting and interpreting the movement of others are critically important for humans. We can make movements more complex than any robot can replicate, and we can recognize and interpret the movement of others more quickly and more efficiently than any computer vision system. Using their ability to recognize and interpret the movement of others, humans can learn by imitation, enjoy athletic and artistic performances, diagnose abnormal movements, detect threats, and effectively communicate and interact with others in the society. Technology is constantly creating new opportunities where human movement must be observed and interpreted but human observers are not or could not be present or their qualitative interpretations are inadequate. It is in these areas that motion measurement technologies can be most beneficial. Motion measurement technologies can complement our own abilities because they excel where humans are deficient or inefficient. They can systematically quantify human movement by accurately measuring external and internal motion variables (Yack, 1984) and they can be used in applications for which human observers are either costly or impractical.

The constantly expanding areas of application for human motion measurement include diagnosis and treatment of movement disorders, computer-assisted surgical operations, virtual reality training of motion in sport and art, studying the motor control of movement, monitoring of physical activity in daily life activities, interactive video games, character animation in movies and video games, and surveillance. These applications have motivated the development or adoption of a number of measurement technologies including optical, magnetic, and inertial motion measurement systems. These measurement systems have been used successfully in many applications but they have limitations and their application has faced many challenges that are fueling continued research. The difficulties arise from the limitations of the measurement technologies but more importantly from the complex nature of human movement. Human movement can take many forms and can be performed in many different environments. It is generated by complex interactions between the central nervous system and the highly coupled musculoskeletal system with many redundancies and irregular joint structures. Further, the movement of the skeletal system must be inferred from the motion of sensors that can only be attached to the skin over the bones and intervening deformable tissue. Existing measurement technologies are powerful but none has the capability to overcome these challenges, and reliably and accurately measure human motion in all its variations. Because there are no silver bullets, it is important to be aware of the strengths and weaknesses of alternative measurement technologies so that the best technology for a specific requirement can be selected and applied successfully (Welch and Foxlin, 2002).

In addition to acquiring raw data about human movement, it is important to consider how the data will be analyzed and presented. Depending on the application, the same human motion can be described in varying levels of detail and from different perspective. The levels of detail in human motion measurement will be described first. This will be followed by the review of the common human motion measurement technologies including their operating principles, their strengths and weaknesses, and examples of their successful applications from the literature. Finally, the current state-of-the-art and emerging technologies for human motion measurement will be discussed.

22.2 CHARACTERIZATION OF HUMAN MOVEMENT

To fully capture the motion of the human body, one must measure the linear and angular motions of all segments in the body with respect to a ground-fixed reference frame. With these measurements one can reconstruct the motion of the whole body and derive other application-specific representations of motion such as the relative angles, velocities and accelerations between the neighboring body segments. The full motion of a body segment includes its linear and angular position, velocity, and acceleration but not all have to be measured directly. Usually one of these quantities (position, velocity, or acceleration) is measured directly and then the other motion variables are derived by numerical differentiation or integration. Full body motion measurement is seldom necessary or practical. In most applications a subset of motion variables is measured.

In some applications the motion of the whole body is measured by treating the human target as a single point. In team sports, for example, the large-scale motion of the players can be tracked while ignoring the relative motion of the individual segments

(Pers et al., 2002). Similarly, in studies of balance and falls, the whole body motion characterized by the motion of the body's center of mass is more important. The motion of the body's center of mass can be estimated by various methods such as measuring the motion of a point on the surface of the body such as waist that is close to the body's center of mass (Karantonis et al., 2006). In most other applications where the motion of the individual segments and their relative motion are important, the motion of any number of segments may be measured. These could include the measurement of the motion of all segments involved in gross movements for character animation in movies and video games, detailed study of specific limbs in reaching, grasping, standing, and gait to diagnose movement disorders, a single joint such as the knee or wrist, or a single segment such as the trunk to monitor general physical activity.

If the motions of all individual segments are available and their underlying mechanical linkage at the joints is understood, then it is possible to derive mathematically any kinematic or kinetic information that is desired. Therefore, despite the variations in methods and technologies, the goal of full body motion capture is to measure the full motion of individual body segments. This is commonly achieved by measuring the motion of multiple markers or sensors that are attached to the body segments or analyzing the remotely captured videos of the whole body movements.

22.3 OPTICAL MOTION CAPTURE SYSTEMS

Optical motion capture systems use multiple cameras to track the position of passive or active markers attached to the body segments (Figure 22.1). In the passive marker system, light sources around the camera's lens emit infrared light parallel to the axis of the camera. The reflected light from the marker is then captured by the cameras as a 2D image (Richards, 1999). Passive markers are usually spherical balls with retroreflective coating and come in various sizes. In active marker systems, that are less common, infrared light-emitting diodes actively emit the light that is imaged by the cameras (Maletsky et al., 2007). This has the disadvantage of requiring wires to each active marker but the advantage of allowing their light to be sequenced so the detection system knows exactly which marker's position is being sensed at any instant.

To identify the 3D position of a marker, the marker must be imaged by at least two cameras (Figure 22.2). One camera can identify the position of the marker in a 2D plane but it cannot provide any information on the axis perpendicular to the plane (Figure 22.2). The 2D images from two cameras with nonparallel image planes, however, can be used to reconstruct the 3D position of the marker center by stereo triangulation (Baca, 1997; Chen et al., 1994).

The full motion of a rigid body segment can be expressed by the 6D motion of a reference frame (three translations, three rotations) that is attached to and moves with the segment. Defining the three orthogonal unit vectors of such a reference frame requires at least three noncollinear points that can be provided by the positions of three optical markers whose positions in the reference frame are known (Figure 22.1). The marker positions must be recorded by the cameras at frame rates high enough to capture the movement details. The temporal and spatial resolution required of these data may be quite high for applications in which the observed motion will be used to compute the forces acting on the body through the method of inverse dynamics.

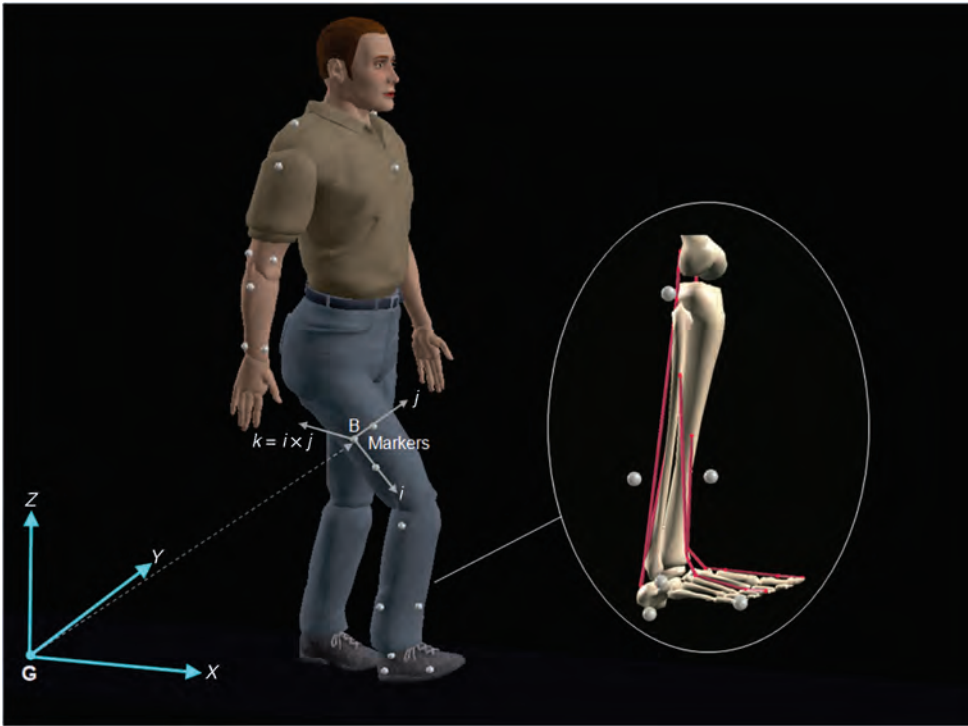


FIGURE 22.1 Optical motion capture system. Motion of the skeletal system is tracked by capturing the motion of the markers on the skin or a tight closing with respect to a ground-fixed coordinate system (G). The motion of a body segment is the same as its body-fixed coordinate system (B) that can be defined by the position of three noncollinear markers on the segment. The captured motion data can be used to create realistic character animations in movies and video games. With the motion data, internal causes of motion such as joint and muscle forces can be examined to diagnosis and treat movement disorders.

Because of the constraints on the marker attachment points, the axes of the body-fixed reference frame defined by them may not have the desired origin or alignment. For example, in biomechanical and clinical applications, a set of standard anatomical axes is defined to facilitate the comparison of anatomical data and their clinical interpretation (Wu and Cavanagh, 1995; Grood and Suntay, 1983). The misalignment between the body-fixed and anatomical reference frame is a fixed transformation that can be obtained in the calibration process. Further, the motion of the body-fixed reference frames that are described by transformation matrices, cannot be directly used or easily interpreted in most applications. For character animation for example, the motion data must be represented in the joint space as the relative angles between the neighboring segments (Xiao et al., 2009). Similarly in biomechanical applications, the relative motion between the anatomical axes of the neighboring segments are described by Euler angles with standard order of rotations (Wu and Cavanagh, 1995; Grood and Suntay, 1983). These alternative representations of the segment motion can be obtained by matrix manipulation and coordinate transformations.

Optical motion capture systems offer powerful capabilities for human motion measurement and have often been used as the gold standard for evaluation of other motion capture

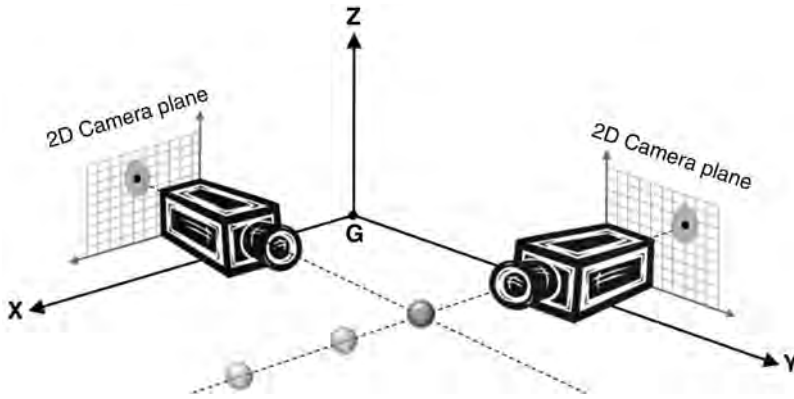


FIGURE 22.2 Reconstruction of 3D marker positions from 2D images. Each camera captures a 2D projection of the marker on its plane and cannot capture the position of the marker on the third axis perpendicular to the camera's plane. To reconstruct the 3D position of the marker, it must be imaged by, at least, two cameras. When a marker is larger or closer to the camera and its image occupies more pixels, the position of its center can be identified with more accuracy than a marker that is smaller or farther away from the camera.

systems. They can track large number of markers in a relatively large area enabling whole body tracking of human movement. They are not susceptible to noise from electromagnetic interference or electrical wiring and can use markers of different sizes to track both large and small-scale motions.

Optical motion capture systems have been applied in many areas including character animation in movies and video games (Xiao et al., 2009; Delaney, 1998), and analysis of movement in patients with gait abnormalities (Chambers and Sutherland, 2002). In gait analysis laboratories, the captured trajectories of the patient's joints are compared to the normal values to diagnose movement abnormalities and recommend treatments such as physical therapy and orthopedic surgical interventions such as osteotomy, tendon transfer, and muscle lengthening. During or after the treatment, follow up motion captures are used to systematically monitor patient's progress and evaluate treatment outcomes (Chambers and Sutherland, 2002). The systematic and objective analysis of gait with the help of motion capture systems has been shown to produce superior outcomes over treatments based on subjective analysis of clinicians, often avoiding unnecessary or inappropriate surgical interventions (Chambers and Sutherland, 2002). Further, with the captured motion, internal and invisible motion variables such as joint moments and muscle forces can be derived and used in treatment decisions that consider the underlying causes of the movement disorder (Seth and Pandy, 2007).

Optical systems have important limitations, however. To capture the motion of a marker, it must be visible to at least two cameras at all times. Therefore the marker cannot be tracked if it is occluded by the body segments or overlaps with other markers. The system may lose track of the occluded marker or confuse it with another marker. The use of more cameras can help but it adds to the cost and complexity of the system and does not completely solve the occlusion problem. Knowledge of human body constraints, signal processing techniques, and optimization methods have been used to identify plausible marker movements and track the momentarily occluded markers (Herda et al., 2001; Zakotnik et al., 2004; Cerveri et al., 2003). The required setup and the algorithms are

often too complex for application in clinical settings and often manual editing may still be necessary. Optical systems are costly and they require extensive calibration and configuration for each patient. They are usually installed in a dedicated indoor space. Although the capture area is relatively large, the precision and accuracy decreases with the distance between the marker and the camera (Maletsky et al., 2007). As the marker's distance from the camera increases, its image becomes smaller and more difficult to capture robustly and accurately (Figure 22.2). The measurement resolution itself is limited by the pixel count of the camera, the size of the marker, and its distance from the camera. Higher pixel count, larger marker size, and closer distance to the camera can result in higher resolution (Baca, 1997). The availability of high-resolution cameras and advanced center detection algorithms with subpixel resolution have made this less of a concern (Baca, 1997). Skin movement artifact is another source of significant errors, but it affects all motion capture systems that most attach their sensors to the skin above the bone. This is not an issue when the objective is to measure the motion of the soft tissue itself (e.g. capturing the motion of the key points on the face for face animation). In most applications, however, the objective is to measure the motion of the underlying skeleton, so the relative motion of the skin can introduce significant errors in the measurement. The soft tissue artifact arises from the movement associated with muscle bulging, skin sliding, and soft tissue inertia. Its amplitude depends on the physical characteristics of the individual, marker location, and the nature of the movement. A systematic review by Peters et al. (2010) found that the soft tissue artifact in thigh and tibia could be as high as 30 and 15 mm, respectively. For a specific marker attachment, the skin movement correlates with the joint motion, so methods have been developed to model and compensate for the errors (Cerveri et al., 2005). Variations in marker attachment in different sessions and laboratories, and differences in the physiognomy of the subjects make the application of these compensatory methods impractical for most clinical situations. In fact, differences in marker placement have been found to be the main source of discrepancies between movements captured by different laboratories (Gorton et al., 2009).

Because the immediate output of the optical systems are the linear and angular positions of the body segments, they can be directly used in most applications such as character animation, movies, and interactive games. Many biomechanical applications, however, require computation of kinetic variables such as joint forces and energetics that necessitates the numerical differentiation of the position data to obtain velocities and accelerations. The numerical differentiation disproportionately magnifies the noise present in the measured data, especially noise at higher frequencies (Pezzack et al., 1977). Because measurement noise has a wide bandwidth, even small amplitude noise in the position data can significantly deteriorate the signal-to-noise ratio of the differentiated data. Fortunately, the high-frequency noise can be easily removed from the motion data by low-pass filtering because the actual human motion is relatively slow. In addition, numerical differentiation algorithms have been developed that have better frequency domain characteristics and produce smoother velocity and acceleration data (Pezzack et al., 1977; Bortolami et al., 1997).

22.4 MAGNETIC MOTION CAPTURE SYSTEMS

Magnetic motion capture systems infer the 6D motion of a body segment using magnetic sensors (Figure 22.3). Only one magnetic sensor or receiver is attached to each body

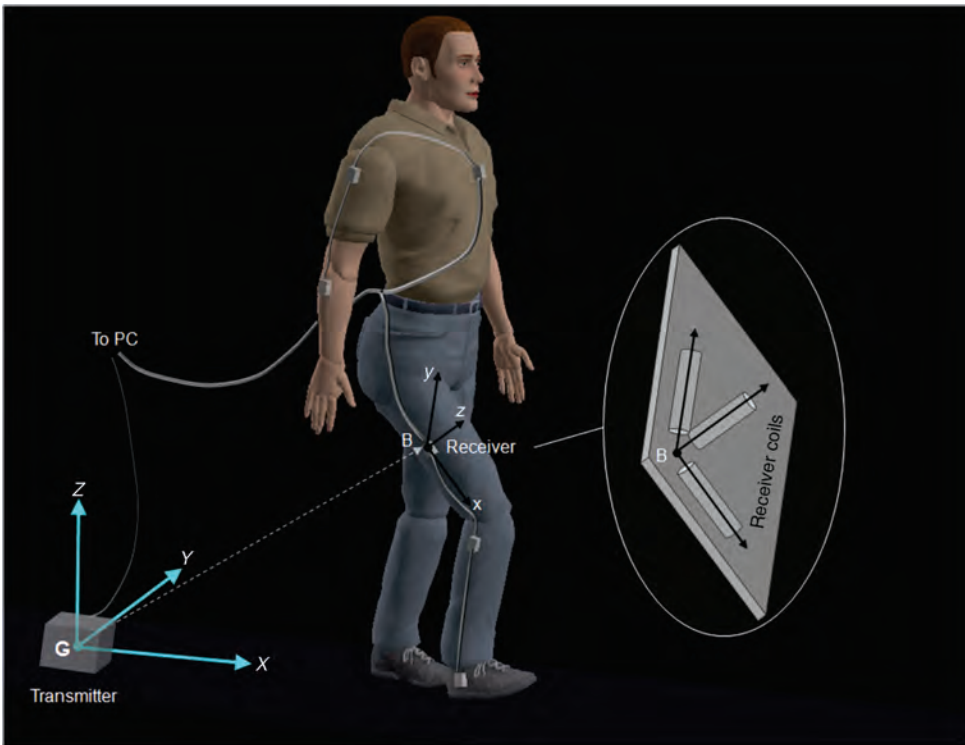


FIGURE 22.3 Magnetic motion capture system. Motion of a body segment is tracked by capturing the 3D position and 3D orientation of a magnetic receiver attached to it. Three orthogonal coils in the transmitter and the receiver define the axes of the ground-fixed (G) and body-fixed (B) coordinate systems, respectively. Magnetic systems are faster than optical systems and more appropriate for real-time interactive applications.

segment. The receiver has three orthogonal coils with ferrite core that form the axes of the body-fixed reference frame. The receivers measure the strength of the magnetic field generated by a transmitter. The transmitter is fixed to the ground and has three orthogonal magnetic coils that form the axes of the ground-fixed reference frame. It emits magnetic pulses from its coils while the corresponding magnetic inductions in the three orthogonal coils of the receivers are measured. The strength of the induced magnetic field in the receiver coils depend on and determine the receiver's relative distance and orientation with respect to the transmitter. As described earlier, once the full motion of each body segment is measured, they can be used to derive other representations of motion.

Compared to the optical systems, magnetic systems are relatively inexpensive. Unlike the optical systems that have a high initial cost but can be used with an arbitrary number of passive markers, the price of the magnetic systems grows with the number of active receivers. They are therefore more economical choices and have been used more often when limited numbers of segments must be tracked such as the motion of the upper extremities and the head movement (Kaliki et al., 2011). The receivers on the upper extremity and head are also farther away from the floor and therefore less susceptible to the magnetic distortion from the metals in the flooring. Magnetic systems have the

advantage that each receiver can measure the complete position and orientation of a body segment, whereas this normally requires three noncollinear markers in an optical system. Because the magnetic field can travel through human tissue, they are not susceptible to marker occlusion. The magnetic sensors are mounted over the skin and experience soft tissue artifact but finding one desirable attachment site on a segment with minimal soft tissue movement is easier than finding the sites for three optical markers. Because the magnetic systems directly measure the position and orientation of the body segment, they require less processing time to obtain the desired motion variables (Richards, 1999). This results in lower latencies that are important for real-time applications such as computer-assisted surgery, motion training in virtual reality, and interactive video games (Richards, 1999; Kaliki et al., 2011; Poulin and Amiot, 2002).

Errors due to the magnetic distortion caused by the metals in the flooring and any nearby instrumentation is the main disadvantage of the magnetic systems (Richards, 1999). The sources of interference may be hidden in the building and the instrumentation and therefore difficult to detect. The distortion itself is also invisible to the user and may go undetected. Errors due to magnetic field distortion can be large. Meskers et al. found that the magnetic distortion by the metal reinforced concrete in the flooring can cause position errors over 20 mm in each axis (Meskers et al., 1999). A study of various sources of interference in an operating room environment found that some devices can introduce position and orientations errors as high as 8.4 mm and 166°, respectively (Poulin and Amiot, 2002). They recommended identifying the sources of distortion and keeping them away from the capture volume to minimize the errors. By removing sources of distortion and careful calibration, measurement accuracies comparable to optical systems can be achieved (Hassan et al., 2007; Mills et al., 2007; Day et al., 2000). Because the strength of the transmitter's magnetic field drops sharply with distance (according to the inverse square law), the range of the magnetic systems is limited to a small area around the transmitter. The wiring could also be limiting because both the transmitter and the body-mounted receivers must be connected to the motion capture electronics. The useful capture area is limited to a hemisphere on either side of the transmitter with radius less than a few meters (Richards, 1999; Meskers et al., 1999). Even within the capture area, measurement errors increase with the distance between the receiver and the transmitter (Richards, 1999; Schuler et al., 2005), somewhat as they do with camera distance in optical systems. The range of the magnetic systems would be unlimited if the earth's magnetic field was used as the source. But the earth's magnetic field is weak and difficult to sense reliably and quickly and has been used mainly in hybrid inertial-magnetic systems to occasionally reset the measurements obtained by inertial sensors (see Section 22.5).

22.5 INERTIAL MOTION CAPTURE SYSTEMS

Inertial sensing with accelerometers and gyroscopes has been developed primarily for and used widely in inertial navigation of missiles, aircraft, and ships. These were too bulky and too expensive for human motion tracking until they could be produced in smaller form factor and at lower cost thanks to micro-electro-mechanical systems (MEMS) technology (Barbour and Schmidt, 2001). These miniaturized inertial sensors can be packaged in small sizes that can be attached to the body segments to track their motion. Inertial sensing of body segments' motion, however, poses specific challenges that have led to adaptations of this measurement technology.

The motion of a vehicle such as an aircraft that moves as a single rigid body can be measured by a set of three orthogonal gyroscopes and three orthogonal accelerometers. In older inertial navigation systems, the accelerometers were attached to a moving platform on the vehicle whose orientation was always aligned with a ground-fixed reference frame. This was accomplished by continually sensing the angular motion of the vehicle by the gyroscopes and rotating the platform in the opposite direction to maintain its alignment. As the result, the acceleration measurements were in the axes of the ground-fixed reference frame and could be directly integrated once to obtain the vehicle's velocity, and twice to obtain its position. The resulting computations were simple and could be performed by onboard analog computers but the precision mechanical devices were expensive and prone to occasional failure. The availability of fast microprocessors allowed the accelerometers and gyroscopes to be attached directly to the vehicle's body. This configuration, which is known as the strapdown inertial navigation system (Welch and Foxlin, 2002), does not have the expensive moving parts but requires more extensive computation. Because the accelerations are expressed in the body-fixed reference frame, they cannot be integrated directly. The angular velocities from the gyroscopes must be continually integrated to obtain the orientation of the vehicle with respect to the ground-fixed reference frame. Matrix transformations must then be used to transform the accelerations from the body-fixed to ground-fixed reference frame before they can be integrated to obtain the vehicle's velocity and position. Because of its simplicity and compact size, the strapdown configuration built from miniaturized MEMS sensors has found applications in human motion measurement. But the use of the technology is customized to the specific requirements of the human movement as described below.

A set of three orthogonal gyroscopes and three orthogonal accelerometers on a body segment can measure its angular velocity and linear acceleration. The measurements are with respect to an inertial ground-fixed reference frame but the measured quantities are projected to the axes of the body-fixed reference frame (Figure 22.4). If the initial orientation of the body segment is known (from initial calibration), it can be used to transform the acceleration and angular velocity measurements from the body-fixed to ground-fixed reference frame. Then the global acceleration minus the gravity acceleration (see below description) and angular velocity can be numerically integrated to obtain the global position and orientation of the segment at the end of the first integration step. Because the global orientation of the segment is known at the start of the next integration step, the integration process can be repeated to track the global motion of the segment over time.

As mentioned earlier, the acceleration due to gravity must be subtracted from the measured accelerations before they can be integrated. Acceleration sensors measure the strain induced by a proof mass in a supporting structure such as a cantilever beam, which reflects the sum of two forces: the weight of the mass that results from gravity and the force due to changes in the velocity of the mass that is acceleration. The accelerometers are therefore sensitive to and measure an acceleration signal that is the sum of the gravity and motion accelerations. The gravity component of the acceleration that exists even when the segment does not move must be subtracted from the measured acceleration before it can be integrated to obtain the velocity and position of the body segment. The estimate of gravity effects must be accurate because even small errors can produce large integration drifts over time. For slow movements with very small movement acceleration, however, the accelerometers measure predominantly the constant gravity acceleration vector, which allows an accurate determination of the segment's inclination with respect to the gravity vector (Luinge and Veltink, 2004; Kemp et al., 1998) (Figure 22.5). Because

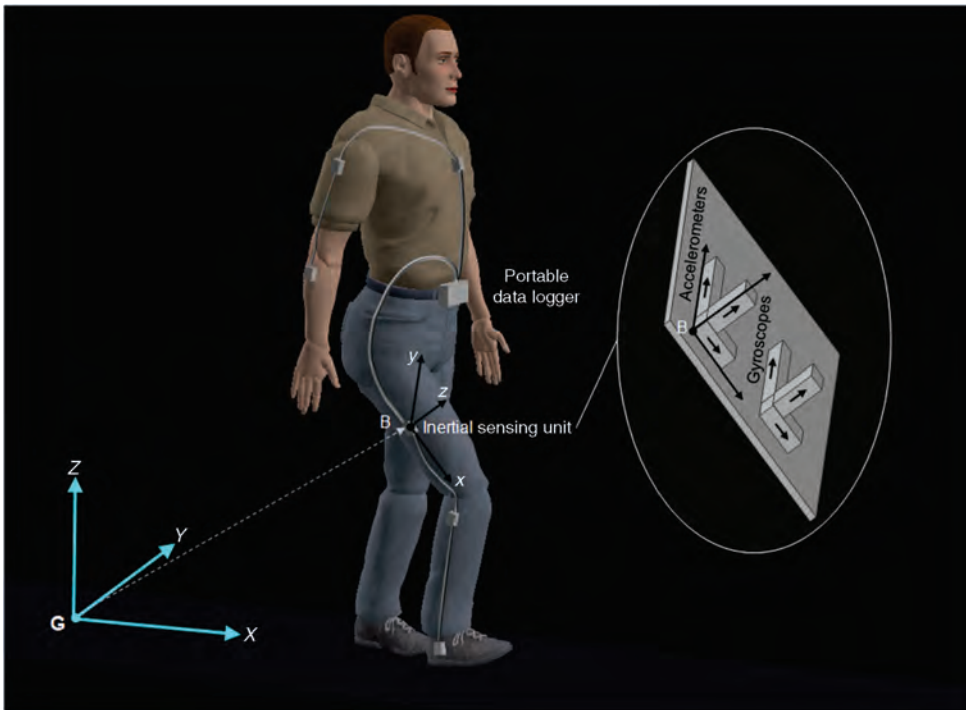


FIGURE 22.4 Inertial motion capture system. Motion of a body segment is tracked by an inertial sensing unit consisting of three orthogonal accelerometers and three orthogonal gyroscopes. 3D orientation is obtained by integrating the angular velocities from the gyroscopes and the 3D position from double integrating the linear accelerations from the accelerometers. Because inertial sensing unit can operate autonomously, it has been used widely in portable applications outside the laboratory.

many human movements such as sitting and standing are slow and the objective is to keep the body segments upright, the measurement of the body segment's deviations from the vertical direction has special utility in many applications.

The main advantage of inertial motion measurement is that it is not susceptible to signal distortion or sensor occlusion. The sensors can be made in small, wearable, and portable packages that operate independent of any external device such as cameras and transmitters. This is possible because solid-state MEMS sensors can be easily integrated with powerful microprocessors and even radio transmitters and packaged into small battery powered units. These standalone units require little power to operate and can be installed on one or more body segment as needed to measure their motion. These unique characteristics have enabled the measurement of motion in daily life activities for long periods at home and at outdoor environment (Godfrey et al., 2008). Inclination measurement by the accelerometers has also been used extensively because it provides the inclination of the body segment with little computation and can be implemented with as little as a single axis accelerometer (Luinge and Veltink, 2004; Godfrey et al., 2008).

Because of their unique capabilities, inertial sensors are finding applications in many areas of movement science, especially in measuring the motion of a limited number of body segments outside the laboratory. Many commercial devices based on inertial sensors

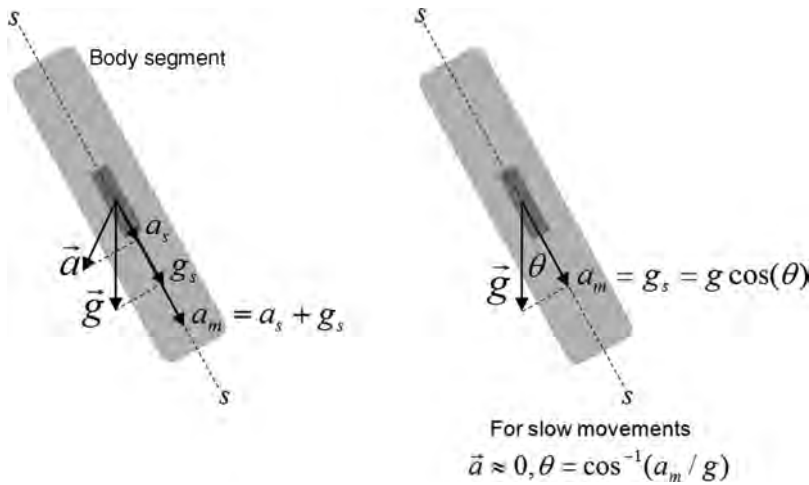


FIGURE 22.5 Components of the measured acceleration signal. The acceleration measured by an accelerometer (a_m) is the sum of the projections of movement (a_s) and gravity (g_s) accelerations on its sensitive axis (left). The gravity component of the acceleration must be accurately estimated and subtracted from the measured acceleration before it can be integrated to obtain position. Any small error in the estimation can result in large integration drifts over time. In slow movements where the movement acceleration is very small compared to the gravity acceleration, the accelerometer predominantly measures the projection of the constant gravity vector (right). This allows the determination of the segment's inclination with respect to the gravity vector, which is an informative motion variable for human movement.

have been developed and applied widely to monitor the type and intensity of physical activity (Godfrey et al., 2008; Yang and Hsu, 2010). The portable inertial sensor may be worn on a body segment relevant to the monitored activity such as the ankle for gait monitoring (Kuo et al., 2009) or on a body segment closest to the body's center of mass such as the waist to monitor the whole body movement and energetics (Karantonis et al., 2006). The measured quantities are then processed by a set of rules or more advanced statistical classification methods to identify the type of activity and estimate the energy expenditure (Godfrey et al., 2008; Yang and Hsu, 2010). Assessment and prevention of falls in the elderly has benefited from the inertial sensors' unique capability to measure inclination and dynamic accelerations. Maintaining balance depends on a limited number of motion variables such as sway distance, sway velocity, and sway acceleration of the body's center of mass that can be measured with a limited number of carefully placed inertial sensors (Sipp and Rowley, 2007). The desired values of these motion variables for maintaining stability can be obtained from simple rules or optimization (Davoodi and Loeb, 2006). Inertial sensor can then be used to detect falls by measuring whether the stability limits have been violated or train the users to avoid falls by alerting them when they are about to lose their balance (Conard et al., 2009; Maurice et al., 2010). Inertial sensors are also more practical because they can be used to monitor and prevent falls outside the laboratory environment (Sipp and Rowley, 2007). Other less common applications of inertial sensing in human motion measurement include the measurement of the upper extremity movements for rehabilitation (Zhang et al., 2008), assessment of spinal posture (Wong and Wong, 2008), measurement and analysis of the lower extremity biomechanics

(Fong and Chan, 2010), and feedback control of human movement in functional electrical stimulation (Tong et al., 2003).

The most important drawback of the inertial sensors is the integration drift and the need for initialization. Unlike optical and magnetic systems that measure absolute motion with respect to a ground-fixed reference frame, the inertial sensors are not aware of their absolute position and orientation. Inertial sensor is like a blind person in a car who knows whether the car is turning left or right but is unaware of the absolute direction with respect to the west, east, north, or south. The inertial sensors must therefore be initialized at the start of the session. Because inertial measurements must be integrated over time to obtain the position of the body segment, even small bias errors in the measurement or errors in the gravity compensation can accumulate resulting in large errors over time (Figure 22.6). Most accelerometers are known to suffer from fluctuating offset or bias error that are difficult to remove completely by calibration especially in clinical applications (Luinge and Veltink, 2004). The most effective method to minimize the effect of integration drift is to periodically reset the integrated motion variables by independent measurements (Figure 22.6). The main reason for building such hybrid systems is that the independent measurements are not always available or they can be made only at much slower rates than is required. A relatively successful example for human motion measurement is the hybrid

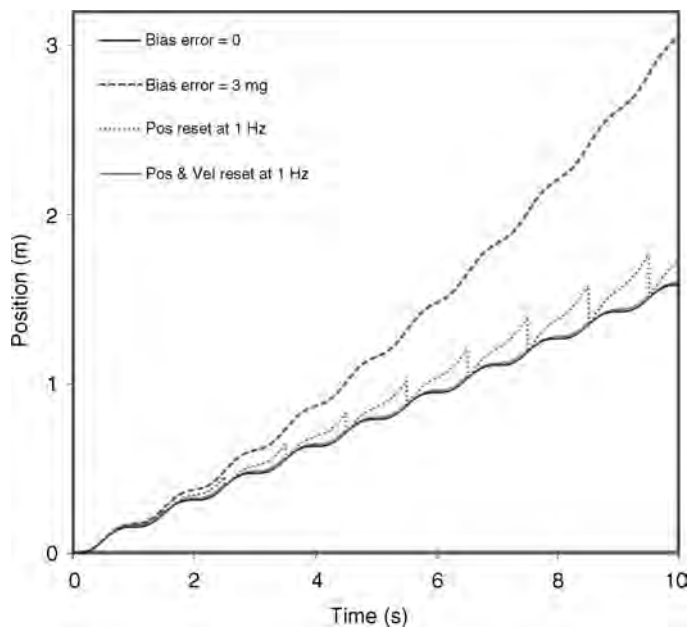


FIGURE 22.6 Integration drift caused by acceleration bias error. Unlike optical and magnetic systems that directly measure the position and orientation, inertial measurements must be integrated twice to obtain these quantities. Shown in the figure is the position of an imaginary point on the body obtained by double-integration of its acceleration (sinusoidal trajectory, 1 m/s^2 amplitude) with no bias error, with a typical bias error of 3 mg for industrial grade MEMS accelerometers, and with resetting of the integrated position or both position and velocity at the rate of 1 Hz . Over only 10 s of movement, the integrated position drifts by 1.5 m . Resetting the integrated position or both position and velocity with independent measurements reduces the max error within 10 s of integration to 0.26 m and 0.014 m , respectively.

inertial–magnetic system in which the integrations of the faster inertial measurements are periodically reset (reinitialized) by the slower measurements of the magnetic system. If the magnetic system senses the magnetic field generated by a small battery-powered transmitter (Roetenberg et al., 2007) or the earth’s magnetic field (Zhang et al., 2008), it can be packaged with inertial sensors in small portable units that can be used in ambulatory applications outside the laboratory. These are increasingly marketed as inexpensive “electronic compasses” or “digital compasses” for use in cell phones, video game controllers, and other portable, consumer electronic devices.

22.6 DISCUSSION

The growing applications in different areas of movement science, entertainment, and medicine are the motivating forces for the development of increasingly sophisticated human motion measurement systems. These often demand measurement systems that are cheaper, smaller, faster, and more accurate, and at the same time capable of measuring human motion in its natural environment reliably and without any encumbrance by markers, sensors, or wires. None of the existing measurement systems satisfy all of these requirements but they offer range of possibilities that must be carefully considered for a specific application.

For whole body motion measurement, optical systems are the gold standard. They are expensive but they can make accurate and reliable measurements when necessary. Because the markers are small and there are no wires, the movements can be made with little encumbrance in relatively large spaces. As the result, they have been used widely to capture the movement of actors for movies and video games and motion of patients with movement disorders. They can also be used in smaller scale measurements of the body segments if their expense can be justified. Magnetic systems have been less accurate and more limited in terms of their capture volume but they are faster and less expensive. Therefore, they have been used in smaller scale measurements of few body segments such as head and arms. They are especially suited for real-time interactive applications where the speed may be more important than the accuracy of the measurements. Magnetic systems are catching up and, with careful calibration, can achieve measurement accuracies that are comparable to the optical systems. But to achieve wider acceptance in character animation and biomechanics, the range of the system must be extended and the electromagnetic distortions must be minimized in a more reliable manner. The most distinguishing aspect of the inertial systems is that they can operate autonomously in indoor and outdoor environments and they do not suffer from the main sources of interference in other systems, that is, marker occlusion and electromagnetic interference. In addition, inertial sensors can directly measure acceleration and inclination of the body segments that are important aspects of many human movements. Inertial sensors have therefore been used more widely in portable applications to monitor physical activity in free-living environments. Inertial sensors can measure the complete 6D motion of body segments but integration drift has limited their use in this capacity. Hybrid inertial–magnetic systems have been built to minimize the integration drift but they need further development before they can be used in multi-segment motion measurement.

Measurement of human motion challenges even the most advanced measurement technologies. For example, inertial sensing is a mature and successful technology for tracking the motion of missiles and airplanes but its application to human motion has faced many

challenges (see Section 22.5). Some of the difficulties arise from the anatomical complexity and large variations of human movement that cannot be changed. But others arise from the limitations of the measurement technologies that are the subject of intense research and development. One of the main limitations and the cause of many errors is the requirement to attach the markers or sensors to the body segments. They are difficult to attach and calibrate consistently; they measure the skin motion on top of the skeletal motion; they suffer from occlusion and interference; they are tedious to attach and they encumber the natural motion of the human target. The goal of one of the promising technologies, markerless motion capture, is to extract the motion of the human target from the 2D images of one or more cameras. Markerless systems are very appealing because they allow the human targets to move freely in their natural environment without the encumbrance of the markers or sensors. They can be applied in areas that have traditionally used marker-based systems and they enable new areas of application such as surveillance where marker-based systems are impractical. But extracting accurate skeletal motion from the video images alone is a challenging task. Sophisticated algorithms must be used to separate the human target's body from the background in the image, identify individual segment's boundary and shape, infer the location of the joint centers, and deal with segment occlusion (Kehl and Gool, 2006; Mundermann et al., 2006). Markerless motion capture has its origin in computer vision research and has been applied mainly to applications with lower accuracy requirement such as surveillance (Mundermann et al., 2006; Aggarwal and Cai, 1999). But the use of more cameras and sophisticated feature extraction algorithms are enabling measurements that are approaching the accuracy requirements of some clinical applications (Corazza et al., 2010; Corazza et al., 2007).

ACKNOWLEDGMENT

The author would like to thank Dr. Gerald E. Loeb for his excellent comments.

REFERENCES

- Aggarwal JK Cai Q. Human motion analysis: a review. *Computer Vision and Image Understanding* 1999;73(3):428–440.
- Baca A. Spatial reconstruction of marker trajectories from high-speed video image sequences. *Medical Engineering & Physics* 1997;19(4):367–374.
- Barbour N, Schmidt G. Inertial sensor technology trends. *IEEE Sensors Journal* 2001;1(4):332–339.
- Bortolami SB, Riley PO, Krebs DE. Numerical differentiation of tracking data of human motion: the virtual accelerometer. *Journal of Dynamic Systems Measurement and Control Transactions of the ASME* 1997;119:355–358.
- Cerveri P, Pedotti A, Ferrigno G. Robust recovery of human motion from video using Kalman filters and virtual humans. *Human Movement Science* 2003;22(3):377–404.
- Cerveri P, Pedotti A, Ferrigno G. Kinematical models to reduce the effect of skin artifacts on marker-based human motion estimation. *Journal of Biomechanics* 2005;38(11):2228–2236.
- Chambers HG, Sutherland DH. A practical guide to gait analysis. *The Journal of the American Academy of Orthopaedic Surgeons* 2002;10(3):222–231.
- Chen L, Armstrong CW, Raftopoulos DD. An investigation on the accuracy of three-dimensional space reconstruction using the direct linear transformation technique. *Journal of Biomechanics* 1994;27(4):493–500.

- Corazza S, Mundermann L, Andriacchi T. A framework for the functional identification of joint centers using markerless motion capture, validation for the hip joint. *Journal of Biomechanics* 2007;40(15):3510–3515.
- Corazza S, Gambaretto E, Mundermann L, Andriacchi TP. Automatic generation of a subject-specific model for accurate markerless motion capture and biomechanical applications. *IEEE Transactions on Biomedical Engineering* 2010;57(4):806–812.
- Davoodi R, Loeb GE. Evolutionary methods for analysis of human movement. In: Begg R, Palaniswami M, editors. *Computational Intelligence for Movement Sciences: Neural Networks and other Emerging Techniques*. 1st ed. Idea Group Publishing;2006. p. 281–298.
- Day JS, Murdoch DJ, Dumas GA. Calibration of position and angular data from a magnetic tracking device. *Journal of Biomechanics* 2000;33(8):1039–1045.
- Delaney B. On the trail of the shadow woman: the mystery of motion capture. *IEEE Computer Graphics and Applications* 1998;14–19.
- Fong DTP, Chan YY. The use of wearable inertial motion sensors in human lower limb biomechanics studies—a systematic review. *Sensors* 2010;10:11556–11565.
- Godfrey A, Conway R, Meagher D, O'laighin G. Direct measurement of human movement by accelerometry. *Medical Engineering & Physics* 2008;30(10):1364–1386.
- Gorton IIIG, Hebert DA, Gannotti ME. Assessment of the kinematic variability among 12 motion analysis laboratories. *Gait & Posture* 2009;29(3):398–402.
- Grood ES, Suntay WJ. A Joint Coordinate system for the clinical description of three-dimensional motions—application to the knee. *Journal of Biomechanical Engineering*. 1983;105:136–144.
- Hassan EA, Jenkyn TR, Dunning CE. Direct comparison of kinematic data collected using an electromagnetic tracking system versus a digital optical system. *Journal of Biomechanics* 2007;40(4):930–935.
- Herda L, Fua P, Plankers R, Boulic R, Thalmann D. Using skeleton-based tracking to increase the reliability of optical motion capture. *Human Movement Science* 2001;20(3):313–341.
- Kaliki R, Davoodi R, Loeb GE. Evaluation of non-invasive command scheme for upper limb prostheses in a virtual reality reach and grasp task. *IEEE Transactions on Circuits and Systems* in press, 2011.
- Karantonis DM, Narayanan MR, Mathie M, Lovell NH, Celler BG. Implementation of a real-time human movement classifier using a triaxial accelerometer for ambulatory monitoring. *IEEE Transactions on Information Technology in Biomedicine*. 2006;10(1):156–167.
- Kehl R, Gool LV. Markerless tracking of complex human motions from multiple views. *Computer Vision and Image Understanding* 2006;104:190–209.
- Kemp B, Janssen, AJMW, van derKamp. B. Body position can be monitored in 3D using miniature accelerometers and earth-magnetic field sensors. *Electroencephalography and Clinical Neurophysiology* 1998;109:484–488.
- Kuo YL, Culhane KM, Thomason P, Tirosh O, Baker R. Measuring distance walked and step count in children with cerebral palsy: an evaluation of two portable activity monitors. *Gait & Posture* 2009;29(2):304–310.
- Luinger HJ, Veltink PH. Inclination measurement of human movement using a 3-D accelerometer with autocalibration. *IEEE Transactions on Neural Systems and Rehabilitation Engineering* 2004;12(1):112–121.
- Maletsky LP, Sun J, Morton NA. Accuracy of an optical active-marker system to track the relative motion of rigid bodies. *Journal of Biomechanics* 2007;40(3):682–685.
- Maurice J, Robert S, Jos A, Rob, Van L, Herman. K. Salient and placebo vibrotactile feedback are equally effective in reducing sway in bilateral vestibular loss patients. *Gait & Posture* 2010;31:213–217.

- Meskers CG, Fraterman H, van der Helm FC, Vermeulen HM, Rozing PM. Calibration of the "Flock of Birds" electromagnetic tracking device and its application in shoulder motion studies. *Journal of Biomechanics* 1999;32(6):629–633.
- Mills PM, Morrison S, Lloyd DG, Barrett RS. Repeatability of 3D gait kinematics obtained from an electromagnetic tracking system during treadmill locomotion. *Journal of Biomechanics* 2007;40(7):1504–1511.
- Mundermann L, Corazza S, Andriacchi TP. The evolution of methods for the capture of human movement leading to markerless motion capture for biomechanical applications. *Journal of Neuroengineering Rehabilitation* 2006;3:6.
- Pers J, Bon M, Kovacic S, Sibila M, Dezman B. Observation and analysis of large-scale human motion. *Human Movement Science* 2002;21(2):295–311.
- Peters A, Galna B, Sangeux M, Morris M, Baker R. Quantification of soft tissue artifact in lower limb human motion analysis: a systematic review. *Gait & Posture* 2010;31(1):1–8.
- Pezzack JC, Norman RW, Winter DA. An assessment of derivative determining techniques used for motion analysis. *Journal of Biomechanics* 1977;10(5–6):377–382.
- Poulin F, Amiot LP. Interference during the use of an electromagnetic tracking system under OR conditions. *Journal of Biomechanics* 2002;35(6):733–737.
- Richards JG. The measurement of human motion: a comparison of commercially available systems. *Human Movement Science* 1999;18:589–602.
- Roetenberg D, Slycke PJ, Veltink PH. Ambulatory position and orientation tracking fusing magnetic and inertial sensing. *IEEE Transactions on Biomedical Engineering* 2007;54(5):883–890.
- Schuler NB, Bey MJ, Shearn JT, Butler DL. Evaluation of an electromagnetic position tracking device for measuring in vivo, dynamic joint kinematics. *Journal of Biomechanics* 2005;38(10):2113–2117.
- Seth A, Pandy MG. A neuromusculoskeletal tracking method for estimating individual muscle forces in human movement. *Journal of Biomechanics* 2007;40:356–366.
- Sipp AR, Rowley BA. Alternative devices for the quantification of human motion. *Critical Reviews in Biomedical Engineering* 2007;35(5):413–442.
- Tong KY, Mak AFT, Ip WY. Command control for functional electrical stimulation hand grasp systems using miniature accelerometers and gyroscopes. *Medical & Biological Engineering & Computing* 2003;41:710–717.
- Welch G, Foxlin E. Motion tracking: no silver bullet, but a respectable arsenal. *IEEE Computer Graphics and Applications*, 2002;22(6):24–38.
- Wong WY, Wong MS. Detecting spinal posture change in sitting positions with tri-axial accelerometers. *Gait & Posture* 2008;27:168–171.
- Wu G, Cavanagh R. ISB recommendations for standardization in the reporting of kinematic data. *Journal of Biomechanics* 1995;28(10):1257–1261.
- Xiao Z, Nait-Charif H, Zhang JJ. Real time automatic skeleton and motion estimation for character animation. *Computer Animation and Virtual Worlds* 2009;20:523–531.
- Yack HJ. Techniques for clinical assessment of human movement. *Physical Therapy* 1984;64(12):1821–1830.
- Yang CC, Hsu YL. A review of accelerometry-based wearable motion detectors for physical activity monitoring. *Sensors* 2010;10:7772–7788.
- Zakotnik J, Matheson T, Durr V. A posture optimization algorithm for model-based motion capture of movement sequences. *Journal of Neuroscience Methods* 2004;135(1–2):43–54.
- Zhang S, Hu H, Zhou. H. An interactive Internet-based system for tracking upper limb motion in home-based rehabilitation. *Medical & Biological Engineering & Computing* 2008;46:241–249.

23

FLOW MEASUREMENT

ARNOLD A. FONTAINE, KEEFE B. MANNING, AND STEVEN DEUTSCH

- 23.1 Introduction
- 23.2 Flow measurement applications
 - 23.2.1 Volume flow measurement
 - 23.2.2 Velocity measurements
- References

23.1 INTRODUCTION

Fluid flow occurs throughout biomedical engineering in areas as different as airflow in the lungs, Fung (1990) and Primiano (1998), for diffusion of nutrients or wastes through membranes [Hampers et al. (1973), Neuman (1998), and Tedgui and Lever (1984)]. Flow-related problems can involve fluid media in the form of gas, liquid, or multiphase flows of liquids, and gas together or in combination with solid matter. Biomedical flows occur in both *in vivo*, Jin et al. (2003), and *in vitro*, Hochareon et al. (2004). They can involve relatively benign flows like that of saline through an intravenous tube to a biochemically active flow of a non-Newtonian fluid such as blood. Many biomedical or bioengineering processes require the quantification of some flow field that may be directly or indirectly related to the process. Such quantification can involve the measurement of volume or mass flow, the static and dynamic pressures, the local velocity of the fluid, the motion (speed and direction) of particles such as cells, the flow-related shear, or the diffusion of a chemical species.

Interest in understanding fluid flows, and attempts to measure flow-related phenomena has had a long history in the scientific and medical communities with early work by Newton, DaVinci, and others. Early studies often involved observations of flow-related phenomena that can be characterized as simple flow visualization or particle tracking (Merzkirch, 1974), or the estimation of a pressure by the displacement of a fluid. Rouse

and Ince (1957) provide a historical review of these early works. Throughout the years, flow measurement techniques have advanced significantly in capability (what is measured and how), versatility, and in many ways, complexity. Some techniques, such as photographic flow visualization, have changed little in over 100 years, whereas others are only possible because of advances in electronics, optics, and physics. Improved capability and versatility are evidenced through the increased ease of use in some systems, and the ability to measure more quantities with increased accuracy and resolution. However, this improved capability and versatility has, in some cases, come at the cost of increased complexity in system hardware, calibration requirements, and application complexity.

Measurement techniques can be characterized as invasive or noninvasive, and direct or indirect. Invasive measurement techniques require the insertion of a sensing or a sampling element directly into the flow field. As a result of this direct insertion, invasive probes may alter the flow field characteristics or induce bias errors associated with the presence of the probe in the flow field or by the operation of the probe (Latto et al., 1981; Lauchle et al., 1989). Invasive probes are often designed to minimize flow disturbance by miniaturizing the sensing elements or by displacing the sensing elements some distance from the hardware holding the probe in the flow, as illustrated in Figure 23.1. Invasive probes also require some type of closure at the penetration site through the boundary of the flow, which must be accounted for in the test design, and can be particularly important in *in vivo* applications to prevent fluid loss or infection. White et al. (1994) measured wall shear stress in the abdominal aorta of dogs with an invasive, flush-mounted hot-film probe. These authors describe how the probe tip is modified to provide an effective entry mechanism through the arterial wall with an adequate seal.

Noninvasive techniques do not involve direct insertion of a sensor into the flow but provide sensing capability through either access to the flow at the flow boundary or through the use of some form of electromagnetic radiation (EMR) transfer or propagation. Wall-mounted thermal sensors, static pressure taps, and transducers, or surface sampling probes are examples of noninvasive techniques that require access to the boundary of the flow field through a wall penetration (White et al., 1994). Ultrasound (Povey, 1997), magnetic resonance (MR), X-ray, and optical techniques are all examples of EMR that can be used to probe flow fields (Siedband, 1998). Unlike the wall-mounted, invasive probes described above, EMR-based measurement systems do not require physical penetration into the flow field or access through the flow field boundary. They do, however, require a “window” into the flow through the boundary enclosing the flow field of interest. This

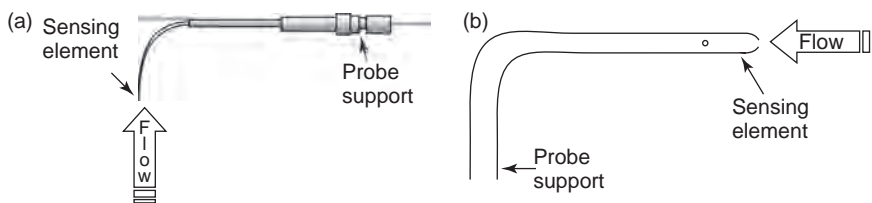


FIGURE 23.1 Examples of invasive velocity measurement sensors with displaced sensing elements relative to their probe supports. (a) A boundary layer style hot-wire probe for thermal anemometry (Picture from TSI Inc. catalog, probe catalog #1261A) (b) Pitot static probe for velocity measurement.

“window” depends on the type of technique being used. Optical-based techniques require an optically clear window that may not be suitable for *in vivo* applications, whereas ultrasound and X-ray techniques require the material properties of the flow boundaries be transparent to these forms of EMR waves. For example, lead will shield X-ray penetration, and metal objects on a surface or in the flow may create local artifacts (noise or error) in MR measurements.

Direct and indirect measurements are defined by how quantities of interest are measured. The displacement of a particle or cell in a flow can be directly measured by photographing the movement of the particle over a finite time interval (Adrian, 1991). Most flow-related measurement systems, however, are indirect. In general, velocity or flow is indirectly calculated from the direct measurement of a quantity and the application of a calibration that relates the magnitude of the measured quantity to the parameter of interest. This calibration may involve not only a conversion of a measured quantity like a voltage to a physical quantity such as a pressure, but may also involve the application of a functional relationship (i.e., Bernoulli’s equation), which requires assumptions about the flow. For example, volume flow probes often assume a characteristic velocity profile at the location of the probe (Webster, 1998). Blood flow in the microcirculation can be estimated by indirectly measuring the cell velocity with a time-of-flight optical technique where the time a cell takes to move a known distance is measured and the cell velocity is calculated by the ratio of the distance divided by the transit time (Lipowski et al., 1989).

Indirect measurement can also impact the uncertainty in the estimated quantity. The measurement of velocity with a Pitot probe is one example of an indirect measurement (Shaughnessy et al., 2005). The Pitot probe measures the local dynamic and static pressures in the flow. These pressures are most often measured with a pressure transducer that provides a voltage output in response to an applied pressure. A calibration is then applied to the measured voltage to convert it to pressure. Velocity is indirectly calculated from the estimated pressures using the Bernoulli equation. The error in Pitot probe velocity measurements includes the pressure transducer calibration uncertainty, noise, and statistical uncertainty during the pressure measurement, electronic noise in the acquisition of the transducer output voltage, transducer drift, and potential bias caused by the physical size of the probe relative to the flow scales being measured. These errors are nonlinearly propagated into the estimate of the velocity uncertainty.

Measurement accuracy is also a function of the physical and operating characteristics of the probe itself. Many flows exhibit a range of spatial and temporal scales. The physical size and the frequency response of the sensing element must be taken into account when choosing a measurement system for a particular application. A large sensing element or an element with poor frequency response has the effect of low-pass filtering the measured signal (Bendat and Peirsol, 1986). This low-pass filtering will cause a bias in the measured quantity. The total measurement uncertainty must also take into account statistical errors associated with random processes, cycle-to-cycle variability in pulsatile systems, and noise. The reader is referred to the texts by Coleman and Steele (1999) and Montgomery (1991) for a detailed approach to experimental uncertainty analysis. The focus of this chapter will be on measurement techniques, their fundamentals of operation, their advantages and disadvantages, and examples of their use.

The name “Flow Measurement” is a broad term that can encompass the measurement of many different flow-related parameters. In this chapter, we will focus on the measurement of those parameters that are most often desired in a biomedical/bioengineering application, volume flow rate, and velocity. Imaging, Doppler echocardiography, and MR

techniques are addressed in other chapters within the encyclopedia and, thus, will only be briefly introduced in this chapter when applicable. This chapter is subdivided into sections that will address volume flow, and velocity separately, with a detailed presentation of systems that are available for the measurement of each. Although ultrasound and MR techniques are often used to measure flow-related parameters, a detailed discussion of the principles of operation will not be presented here as these topics are covered in depth in other chapters of this encyclopedia.

23.2 FLOW MEASUREMENT APPLICATIONS

23.2.1 Volume Flow Measurement

In both the clinical environment and the laboratory environment, the measurement of the volume flow rate of a fluid as a function of time can be an important parameter. In internal flow applications, which comprise most biomedical flows of interest, the volume flow of a fluid (Q) is related to the local fluid velocity (V) through the integration of the velocity over the crosssectional area of the duct or vessel (Shaughnessy et al., 2005; White, 1979)

$$Q = \int V dA \quad (23.1)$$

The flow rate Q , velocity V , and area A have dimensions of volume per time, length per time, and length squared, respectively. The SI units are typically used in the bioengineering field with mass units of grams or kilograms, length units of meters (millimeter and centimeter), and time units of seconds. The mass flow (M) is directly related to the flow volume through the fluid density, ρ , with units of mass/volume,

$$M = \rho \cdot Q \quad (23.2)$$

Fluid pressure and velocity are related through the Navier–Stokes equations, which govern the flow of fluids in internal and external flows (see White, 1979).

The volume flow rate of blood is often measured in many cardiovascular applications (Caro et al., 1978). For example, cardiac output (CO) is the integrated average of the instantaneous volume flow rate of blood (Q_b) exiting the aortic valve over one cardiac cycle (T_c):

$$CO = \left[\int Q_b dt \right] / T_c \quad (23.3)$$

The cardiac output has units of volume flow rate, volume per unit time. The following subsections will provide an overview of measurement techniques typically used for volume flow and velocity measurement in *in vivo*, and *in vitro* studies. This chapter will be limited to those flow measurement techniques most often used in the biomedical and bioengineering fields. Specialized measurement techniques, such as concentration or species measurement through laser-induced fluorescence (LIF) or mass spectrometry, will not be addressed.

23.2.1.1 Electromagnetic Flow Probes Carolina Medical, Inc. developed the first commercially available electromagnetic flow meter in 1955. The design provided scientists

and clinicians with a noninvasive tool that could directly measure the volume flow rate (Webster, 1998). Clinical probes were developed that could be attached to a vessel for extravascular measurement of blood flow without the need for cannulation of a surgically exposed vessel.

Electromagnetic flow meters are volumetric flow measuring devices designed to measure the flow of an electrically conducting liquid in a closed vessel or pipe. Commercial meters, used in both the biomedical and general engineering fields, come in a variety of sizes and designs that can measure flow rates from ~ 1 mL/min to $>100,000$ L/min. Reported uncertainties are typically on the order of a few percent but can be as low as 0.5% of reading in some specialized meter designs. Low uncertainties are dependent on proper use and installation of the meter. Improper use or installation (mounting the meter in a location with a complex flow profile) will increase measurement uncertainty. Most clinical quality meters that are mounted externally to a vessel exhibit uncertainties that can approach 15% in some applications, (Carolina Medical, Inc., product support literature). In cardiovascular applications, these meters may also be susceptible to electrical interference by the heart or by measurement anomalies caused by motion of the probe.

The principle governing the operation of an electromagnetic flow meter is Faraday's Law of Electromagnetic Induction. This law states that an induced electrode voltage is proportional to the velocity of flow of a conductor through a magnetic field of a known density. Mathematically, this is:

$$E_e = K[V \cdot B \cdot L_e] \quad (23.4)$$

Here, E_e is the induced voltage between two electrodes (with units of volts) separated by a known conductor length L_e (provided by the conducting fluid between the electrodes) in units of millimeters or centimeters, B is the magnetic field strength in units of Tesla's, and V the conducting fluid average velocity in units of length per time. The parameter K is a dimensionless constant. Meter output is linear and proportional to flow velocity.

Fluid properties such as viscosity and density are absent from Equation (23.4). Thus, the output of an electromagnetic flow meter is independent of these properties and its calibration is independent of the type of fluid, provided the fluid meets minimum conductivity levels. This meter can then be used for highly viscous fluids, Newtonian fluids, and non-Newtonian fluids such as blood. The requirement of an electrically conductive fluid can disqualify an electromagnetic meter in some applications. Typical meters require a minimum fluid conductivity of ~ 1 mS/cm. However, low conductivity designs are capable of operating with fluid conductivities as low as 0.1 mS/cm. The presence of gas bubbles in the fluid can cause erratic behavior in the meter output.

A typical meter design has electromagnetic coils mounted on opposing sides of an electrically insulated duct with two opposing electrodes mounted 90° relative to the electromagnets. The two electrodes are mounted such that they are in contact with the conducting fluid. Figure 23.2 illustrates the typical configuration.

The meter is designed to generate a magnetic field that is perpendicular to the axis of motion of the flowing fluid. A voltage is generated when the conducting fluid flows through the magnetic field. This voltage is sensed by the two opposing electrodes. The supporting material around the meter is made of a nonconducting material to prevent leakage of the voltage generated in the moving fluid into the surrounding material. In practice, the conductor length, L_e , is not the simple path illustrated, but is rather the integral of all possible path lengths between the two electrodes across the crosssection of the duct or vessel. The signal generated along each path length is proportional to the fluid

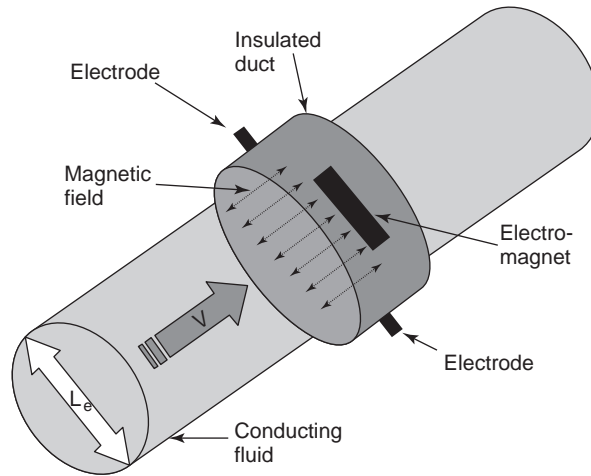


FIGURE 23.2 Illustration of the principle of operation of the electromagnetic flow meter. Note, the magnetic field, electrodes, and flow direction are all mutually perpendicular to one another.

velocity across that path. Thus, the two electrodes measure the integrated sum of all velocities across every possible path in the vessel crosssection. This signal is then directly proportional to the volume flow rate of the fluid passing through the magnetic field.

The magnetic field generated in commercial meters may be anything from a uniform field to a specifically designed field with prescribed characteristics. Meters with uniform magnetic fields can exhibit some sensitivity to the conducting liquid's velocity profile. Fluid flowing through a vessel does not have the same velocity at all locations across the vessel. The no-slip condition at the walls of the duct ensures that the fluid velocity at the wall is zero (Shaughnessy et al., 2005). Viscosity then generates a gradient between the flowing fluid in the vessel and the stationary fluid at the wall. In complex geometries, secondary flows may occur, and velocity gradients in other directions may also develop (Pedley et al., 1971; Mori and Nakayama, 1965). This variation in fluid velocity across the magnetic field coupled with variations in the conductor length generates a variation in the magnitude of the voltage measured across the duct. As a result, installation of these meters must be carefully performed to ensure that the velocity profile of the liquid in the tube is close to that used during calibration.

A number of commercial meters shape the magnetic field coils to generate a magnetic field that exhibits a field strength with a prescribed pattern across the duct. This field shaping compensates for some velocity variations in the duct, and provides a meter with a reduced sensitivity to flow profile. As a result, this type of meter is better able to measure in vessels with upstream and downstream characteristics, such as curvature and non-uniformity in the vessel crosssection, that generate asymmetric flow profiles with secondary flow velocity components.

Commercial meters generate magnetic fields with either an AC excitation or a pulsed DC excitation. The AC excitation generates a magnetic field with a strength that varies with the frequency of the applied AC voltage. This configuration produces a meter with a relatively high-frequency response but with the disadvantage that the output signal not only varies with the flow velocity but also with the magnitude of the alternating excitation

voltage. Thus, the output of the meter for a flow with constant velocity across the vessel will exhibit a sinusoidal pattern. In addition, zero flow will produce an offset output because of the presence of the nonmoving conductor in a moving magnetic field. Quadrature signal rejection techniques can be used to filter the unwanted signal generated by the AC excitation from the desired signal generated by the flowing liquid, but this correction requires careful compensation and zeroing of the meter in the flow field before data acquisition.

The pulsed DC excitation was developed to reduce or eliminate the zero shift encountered with AC excitation. This improvement has the cost of reduced frequency response, and increased sensitivity to the presence of particulates in a fluid. Particles that impact the electrodes in a pulsed DC-operated meter produce output fluctuations that can be characterized as noise. The accuracy in each system is comparable.

A low sensing voltage at the electrodes requires a signal-conditioning unit to provide a measurable output with a good signal-to-noise. Meter calibrations typically involve one of two calibration techniques. The meter and signal-conditioning unit are calibrated separately, or the meter and signal-conditioning unit are calibrated as a system. The latter provides the most accurate calibration with accuracies that can approach 0.5% of reading in certain applications. The reader is referred to literature by various manufacturers of electromagnetic flow meters for a more comprehensive discussion of the techniques, operation, and use for specific applications.

23.2.1.2 Ultrasound Techniques—Transit Time Volume Flow Meters Ultrasonic Transit Time flow meters provide a direct measure of volume flow by correlating the change in the transit time of sound waves across a pipe or vessel with the average velocity of the liquid flowing through the pipe (Webster, 1998; Drost, 1978; Lynnworth, 1979). Transit time ultrasonic flow meters are widely used in clinical cardiovascular applications. In recent years, a number of studies have been performed to evaluate and compare transit time ultrasonic flow measurement techniques with other techniques used clinically (Beldi et al., 2000). The typical configuration for an ultrasonic flow probe involves one or two ultrasonic transducers (transmitters/receivers) and possibly an ultrasonic reflector. Transducers and reflectors are positioned in opposing configurations across a tube or vessel as illustrated in Figure 23.3. The time it takes an ultrasound wave to propagate across a fluid depends on the distance of propagation and the acoustic velocity in the fluid. If the fluid is moving, the motion of the fluid will positively or

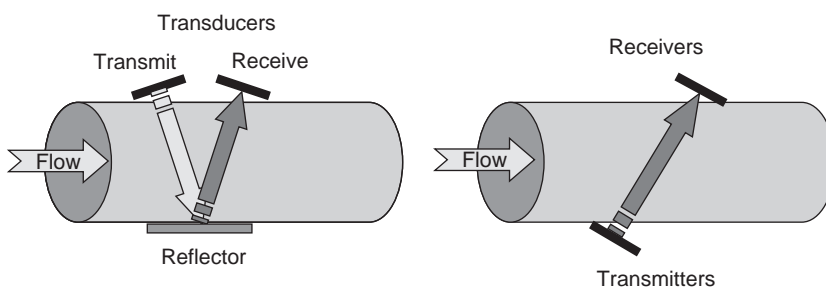


FIGURE 23.3 Illustration of principal of transit time ultrasonic flow probe operation.

negatively add a phase shift to the time of propagation (transit time) through the fluid, which can be written mathematically as:

$$T_t = D_p / [c \pm V \cdot \cos(\theta)] \quad (23.5)$$

Here, T_t is the measured transit time (s), D_p is the total propagation distance of the wave (length units), c is the acoustic speed of the fluid (units of length per time), V is the average velocity of the fluid (units of length per time) over the propagation length, and θ is the angle between the flow direction and the propagation direction of the wave. The configurations illustrated in Figure 23.3 have an inherent dependency of the measured transit time on the coupling of the transducer with the vessel. Acoustic impedance characteristics of the vessel wall, and mismatches in impedance at the vessel wall/transducer and wall/fluid interfaces will affect the accuracy of the flow rate measurement.

The approach to using Equation (23.5) in a metering device is to incorporate bidirectional wave propagation in opposing directions, as shown in Figure 23.4, which will produce two independent transit time measurements (T_{t1} and T_{t2}), one from each direction of propagation. The forward direction transit time T_{t1} is defined by Equation (23.5) with a plus sign before V and T_{t2} by the minus sign. The fluid velocity can then be obtained by taking the difference of the transit times ($T_{t1} - T_{t2}$). It can be shown that, for fluid velocities small relative to the acoustic velocity of the fluid ($V^2 \ll c^2$), this difference reduces to:

$$(T_{t1} - T_{t2}) = 2D_p(V \cdot \cos(\theta)) / c^2 \quad (23.6)$$

With the probe geometry defined (the propagation distance and propagation angle relative to the vessel or flow direction) and with the fluid properties known (acoustic speed of the fluid), the average speed of the fluid along the propagation path of a narrow beam can be calculated from the transit times. Wide beam illumination, where the beam width is wider than the vessel diameter, effectively integrates the measured transit time phase shift over the crosssection of the vessel. The wide beam transmission can be approximated by the summation of many infinitesimally narrow beams adjacent to one another. Thus, the measured wide beam phase shift is proportional to the sum of these narrow beams propagating through the vessel. As the phase shift encountered in a narrow beam transmission is proportional to the average fluid velocity times the beam path length, integrating or summing over all narrow beams across the vessel results in a measured total transit time phase shift

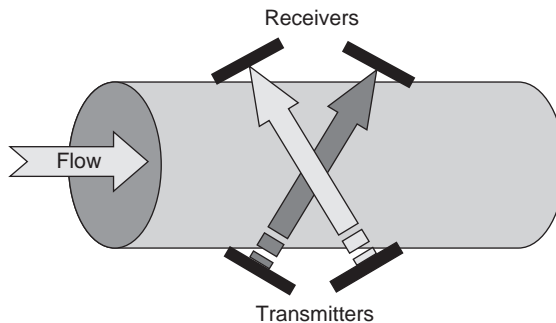


FIGURE 23.4 Bidirectional wave propagation.

that is proportional to the average fluid velocity times the area of the vessel sliced by the ultrasound beam, or the volume flow rate.

The popular Transonic Systems, Inc. flow meter uses bidirectional transmission with two transducers operating in transmit and receive modes, and a single reflector configured as illustrated in the left schematic of Figure 23.4. This approach increases the propagation length while effectively reducing sensitivity to wall coupling or misalignment of the probe with the wall. The increased path length improves uncertainty and provides a probe body with a relatively small footprint, an advantage in *in vivo* or surgical applications.

The basic operation of a bidirectional transit time meter involves the transmission of an ultrasound plane wave at a specific frequency. This wave propagates through the vessel wall and fluid where it is either received at the opposite wall or is reflected to another transducer operating as an acoustic receiver. This received signal is recorded, processed, and digitized before the transducer is reconfigured to transmit a second pulse in the opposite direction. The overall frequency response of such a probe is dependent on the pulse time, the time delay between the forward and reverse pulses, the acoustic speed through the medium, the propagation distance, and the signal-conditioning electronics, which can include analog signal acquisition, and filtering. The meter size governs the propagation distance and, thus, the size of the vessel that the meter can be mounted on.

The frequency response of commercial probes varies from approximately 100 Hz to more than 1 KHz, where the highest frequency responses are obtained in the smaller probes. As a result, commercial probes have sufficient frequency response for most clinically or biomedically relevant flow regimes. Velocity and flow resolution is governed, in part, by the propagation length over which the flow is integrated and the resolution of the transit time measurement. The reader is referred to the meter manufacturers for detailed information about the operating specifications of particular meters. Reported uncertainties in transit time meters can be better than 15%. Actual uncertainties will depend on meter use, the experience of the operator, the meter calibration, and the acoustic properties of the fluid measured and how these properties differ from those of the calibration fluid.

23.2.1.3 Ultrasound Techniques—Doppler Volume Flow Meters Flow can also be measured by ultrasound using the Doppler shift in a propagating sound wave generated by moving objects in a fluid flow (Webster, 1998; Weyman, 1994). The primary difference is in the principal of operation. Devices using the Doppler approach measure the Doppler frequency shift of the transmitted beam caused by the motion of particles encountered along the beam path, as illustrated in Figure 23.5. The Doppler shift caused by reflection of an incident wave by a moving particle is given by:

$$F_D = 2F_o V \cos(\theta) / c \quad (23.7)$$

The shift frequency F_D (units of 1/s) is linearly related to the component of the speed of a particle, V , in + the direction of the wave propagation, the initial transmission frequency of the wave, F_o , and the speed of sound in the fluid, c . The reader is referred elsewhere in this encyclopedic series, and to the text by Weyman (1994) for a detailed presentation of the Doppler technique.

The Doppler meter provides a direct measure of velocity that can be used to calculate the volume flow rate indirectly. Most biomedical applications involving volume flow measurement are performed on flow through a duct or vessel of given shape and size. Thus, the volume flow is the integral of the measured velocity profile across the vessel

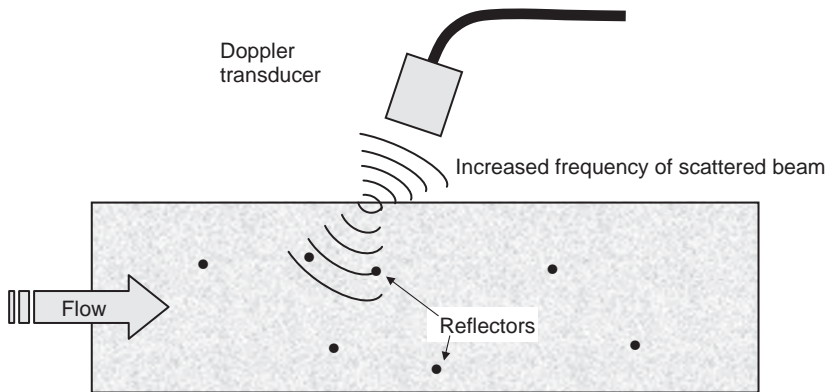


FIGURE 23.5 Schematic of the operation of a Doppler ultrasound probe.

crosssectional area as defined in Equation (23.1). The integral in Equation (23.1) can be related to the average velocity \bar{U} across the duct multiplied by the crosssectional area of the duct (White, 1979). The Doppler technique then requires not only an estimate of the average velocity in the vessel but knowledge of the vessel area as well.

Commercially available Doppler volume flow meters, although not commonly used in biomedical applications, can be attached to a pipe or duct wall as with transit time meters. The commercial Doppler flow meters measure volume flow by integrating the measured Doppler shift frequency generated by particles throughout the flow. This integration is performed over a predefined path length in the flow and is dependent on the number, type of particles, their size, and distribution. The meter accuracy is also dependent on the velocity profile in the flow. Careful *in situ* calibrations are often needed to obtain accuracies of less than 10%. The Doppler meter has several disadvantages when compared with the transit time meter.

It requires a fluid that contains a sufficient concentration of suspended particles to act as scattering sites on the incident ultrasound wave. These particles must be large enough to scatter the incident beam with a high intensity level, but small enough to ensure that they follow the fluid flow (Crowe et al., 1998). As a result of their dependence on flow profile, Doppler flow meters are not well suited for measurement of flow in vessels with curvature or branching. Doppler flow meter measurements in blood rely on blood cells to act as scatterers. Clinical Doppler ultrasound machines, commonly used in echocardiography, can also be used to indirectly infer volume flow through the direct measure of the fluid velocity, and will be discussed later in the subsection on velocity measurements.

23.2.1.4 Invasive or Inline Volume Flow Measurement Invasive or inline flow meters must be installed inline as a part of the piping or vessel network and involve hardware that is in contact with the fluid. These meters often have a non-negligible pressure drop and may adversely interact with the flowing fluid. As a result, these meters are not often used in *in vivo* applications. Meters that fall in this category are: variable area rotameters, turbine/paddle wheel meters, and vortex shedding meters. The primary advantage of these meters is low cost and ease of use. However, these meters typically exhibit sensitivity

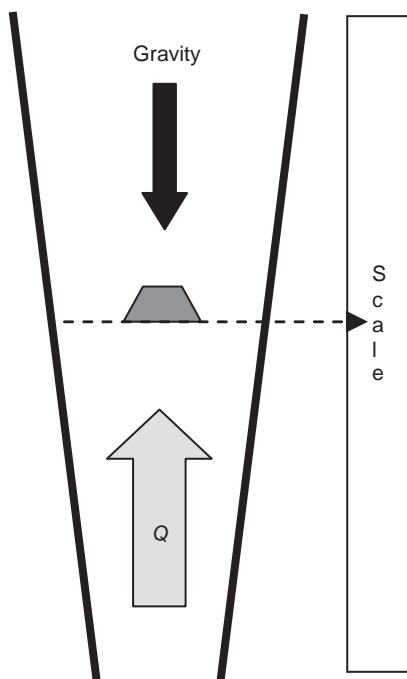


FIGURE 23.6 Schematic of a variable area flowmeter.

both to fluid properties, which can be dependent on temperature and pressure, and to flow profile (White, 1979).

Variable area rotameters are simple measurement devices that can be used with a variety of liquids and gases. The flow of fluid through the meter raises a float in a tapered tube, as shown in Figure 23.6. The higher the float is raised, the larger the diameter of the tapered tube, increasing the crosssectional area of the tube for passage of the fluid. As the flow rate increases, the float is elevated higher in the tube. The height of the float is directly proportional to the fluid flow rate. In liquid flow, the float is raised by the combination of the buoyancy of the liquid and the fluid drag on the float. Buoyancy is negligible in gaseous flows and the float moves in response to the drag by the gas flow. For constant flow, the float reaches a stable position in the tube when the upward force generated by the flowing fluid equals the downward force of gravity. The float will move to a new equilibrium position in response to a change in flow. These meters must be installed vertically to operate properly; however, spring-loaded meters have been designed to eliminate the need for gravity, and permit installation in other orientations.

Rotameters are designed and calibrated for the type of fluid (fluid properties such as viscosity and density) and flow range expected. They do not function properly in non-Newtonian fluids. Use of a meter with a fluid different from that which the meter was calibrated for, or with a fluid at a different temperature or pressure, requires a correction to the meter reading. Meter uncertainty and repeatability will vary with operation of the meter, but can approach a few percent with proper operation.

Turbine and paddle wheel meters measure volume flow rate through the rotation of a vaned rotor in contact with the flowing fluid. These meters are intrusive flow

measurement devices that produce higher pressure drops than others in the class of invasive flow probes. The turbine meter has a turbine mounted across the pipe or duct in full contact with the flow, whereas the paddle wheel meter has a vaned wheel mounted on the side of the duct with half of the wheel in contact with the flow. Accuracy and repeatability is best with the turbine meter, but the pressure drop is also higher. An AC voltage is induced in a magnetic pickup coil as the turbine or paddle wheel rotates. Each pulse in the AC signal represents the passage of one blade of the turbine. As the turbine fills the flow path, a pulse represents a distinct volume of fluid being displaced between two turbine blades. This design provides an accurate direct measure of the volume flow rate.

Flowmeter selection must take into account the type of fluid, the flow rate range under study, and the acceptable pressure drop for a given flow application. In general, these meters have a sensitivity to flow profile, and the pressure drop is dependent on the fluid properties. The meters incorporate moving parts within the flow and thus use bearings to ensure smooth operation. Bearing wear will affect the meter accuracy and must be monitored for the life of the meter. The paddle wheel meter operates in a similar manner as the turbine meter. The primary difference is that only part of the rotor is in contact with the fluid and, thus, although the paddle wheel meter is more sensitive to flow profile, it has a smaller pressure drop. Installation of these meters often involves a specified number of straight pipe sections upstream and downstream of the meter, and may also require installation of a flow straightener inline upstream of the meter.

Vortex meters operate on the principal of Strouhal shedding. Separating flow over an obstruction such as a cylinder or sharp-edged bar results in a pulsatile or oscillatory pattern as shown in Figure 23.7. The shedding frequency, ω (units of 1/s), is related to the fluid velocity by:

$$\omega = V \cdot S_t / L \quad (23.8)$$

where S_t is the Strouhal number, which is a nondimensional number that is a function of the flow Reynolds number and geometry of the obstruction, and L is a characteristic length scale (White, 1979). For a cylinder, L is the diameter of the cylinder. The Reynolds



FIGURE 23.7 Vortex shedding from a circular cylinder (Picture from White (1979), courtesy of the U.S. Naval Research Laboratory)

number is a dimensionless number that is the ratio of inertial to viscous forces in the flow and is defined as

$$Re = V \cdot L / \nu \quad (23.9)$$

Here, V and L are defined as in Equation (23.8), and ν is the kinematic viscosity of the fluid with units of length squared per time.

The vortex meter is an intrusive meter that has a "shedder bar" installed across the diameter of the duct. The flow separates off this bar and generates a shedding frequency that is transmitted through the bar to a piezoelectric sensor attached to the bar. The meter is sensitive to flow and fluid properties, and rated accuracy and pressure drop depend on application.

Volume flow rate can also be estimated through an indirect measure of the velocity profile in the flow and the use of Equation (23.1). A number of instruments are available that measure fluid velocity in biomedical engineering applications. Doppler ultrasound and MR phase velocity encoding are standard clinical techniques used to measure velocity of a flowing fluid noninvasively. *In vitro* systems that are commonly used to measure fluid velocity, in addition to Doppler and MR, are laser Doppler velocimetry (LDV), particle image velocimetry (PIV), and thermal anemometry. Besides an estimate of volume flow rate, fluid velocity measurement can provide quantification of flow profiles, fluid wall shear, pressure gradient, and flow mixing. The following section summarizes velocity measurement techniques commonly used in biomedical/bioengineering applications.

23.2.2 Velocity Measurements

23.2.2.1 Thermal Anemometry Thermal anemometry is an invasive technique used to measure fluid velocity or wall shear. A heated element is inserted into the flow and specialized electronic circuitry is used to measure the rate of change in heat input into the element in response to changes in the flow field (Comte-Bellot, 1976). Thermal anemometry, when used properly, is characterized by high accuracy, low noise, and high spatial, and temporal resolution. Its main disadvantages are sensitivity to changes in fluid temperature and properties, particulates and bubbles suspended in a fluid, nonlinear response to velocity, and its invasive characteristics (geometry, size, vibration, etc.).

Hot-film anemometry has been used to measure blood velocity and wall shear in biomedical and bioengineering applications both *in vitro* and *in vivo*. Arterial blood flow velocity measurements were performed by Nerem et al. 1976, 1974 in horses and by Falsetti et al. (1983) in dogs. Tarbell et al. (Baldwin et al., 1988) used flush-mounted hot films to measure wall shear in the abdominal aorta of a dog. *In vitro* applications of hot-film anemometry include the measurement of wall shear in an LVAD device (Baldwin et al., 1988) and the *in vitro* measurement of flow in rigid models of arterial bifurcations by Batten and Nerem (1982). Although rarely used in biomedical applications now, we will briefly present hot-film anemometry here for completeness. The reader is referred to the text by Brunn (1995) and the symposium proceedings by Stock (1993) for a detailed presentation of thermal anemometry.

Thermal anemometry operates on the principal of convective cooling of a heated element. A thin wire or coated quartz element, mounted between supports, is heated and exposed to a fluid flow, as shown in Figure 23.8. The element is heated by passing a current through the wire or element. The amount of heat generated is proportional to the

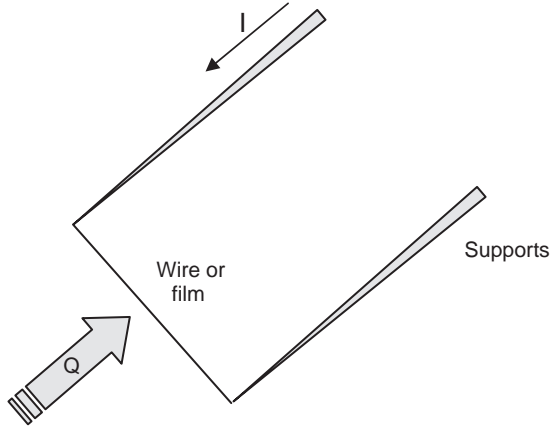


FIGURE 23.8 Illustration of a hot wire or film element.

current, I (measured in units of amps), and the resistance of the element, R (ohms), by I^2R . The element is convectively cooled by flow until an equilibrium state is reached between electrical heating of the element and flow-induced convective cooling; $DE/Dt = W + H$, where E is the thermal energy stored in the element, W is the heat added by Joule heating, and H is the heat loss to the environment by cooling. At equilibrium, $DE/Dt = 0$ and $W = H$.

Changes in flow velocity will increase or decrease the amount of convective cooling and produce changes in the equilibrium state of the element or the temperature of the element. Commercial anemometers employ a four-arm electronic bridge circuit to maintain constant element temperature, current, or voltage in response to changes in convective cooling. As convective cooling changes, the anemometer output, current, or voltage changes in response to maintaining the desired set condition. This equilibrium condition assumes that radiation losses are small, conduction to supports is small, temperature is uniform over the length of the sensor, velocity impinges normally on the sensor, velocity is uniform over the sensor length and is small compared with the sonic speed, and, finally, the fluid temperature and density are constant.

An energy balance between convective heat cooling and Joule heating can be performed to derive a set of governing equations that relate input current, I , to convective velocity, V . The “King’s Law” is the classic result of this energy balance:

$$I^2 R^2 = V_o^2 = (T_w - T_a)(A + B \cdot V^n) \quad (23.10)$$

where V_o is the measured voltage drop in response to a velocity, V , and T_w and T_a are the wire and ambient fluid temperatures (degrees Centigrade), respectively. The coefficients, A and B , and power, n , are determined through careful calibration over the velocity and temperature range that will be observed experimentally. In the event of a three-component flow, the probe must be calibrated for yaw and pitch angles between the probe and the flow velocity vector, and the velocity, V , in Equation (23.10) must be replaced by a term related to the velocity vector magnitude. Bridge-type circuits are also prone to stable and unstable performance under unsteady operation. Thus, the overall calibration of a

hot-wire/film system must involve the element and electronics as a system, and must also involve dynamic calibrations to characterize the frequency response of the system.

Hot-wire/film probes come in a variety of sizes, shapes, and configurations. Probes are manufactured from platinum, gold-plated tungsten, or nickel-plated quartz, and comes in single-or multi-element configurations for measurement in a variety of flow conditions. The reader is referred to the hot-wire/film manufacturers for a complete summary of probe types and conditions for use. In general, wire probes are used when possible because of lower cost, improved frequency response, and ease of repair. However, wire probes are more fragile when compared with film-type probes and are usually restricted to airflows. Film probes are used in rough environments, such as liquid flows.

The following considerations should be addressed to ensure accurate measurements when using thermal anemometry. The type of flow should be assessed for velocity range, flow scales, and fluid properties (clean gas or particle contaminated liquid, etc.). The flow characteristics will define the right probe, anemometer configuration, and A/D setup to use. Appropriate calibrations with complete hardware setup should be performed. Perform the experiment and post calibrations to ensure that the anemometer/probe calibration has not changed.

23.2.2.2 Doppler Ultrasound and Magnetic Resonance Flow Mapping The focus of this subsection is to introduce the concept of Doppler ultrasound and MR flow mapping for local velocity measurement. Flow measurement with clinical Doppler can suffer from the same limitations as the small Doppler meter, but has several advantages over these small meters. Most ultrasound machines can operate in continuous wave or pulsed Doppler modes of operation; see Weyman (1994) for a more detailed discussion of the modes of operation.

Pulsed Doppler ultrasound offers the advantage of localizing a velocity measurement within a flow and can be used to measure the velocity profile across a vessel or lesion. This information coupled with echocardiographic imaging of the geometry can be used to calculate the flow rate from Equation (23.1). Unfortunately, the implementation of this technique is not straightforward because of limitations in resolution, velocity aliasing, and the need to know the relative angle between the transmitted ultrasound beam and the local flow.

Velocity aliasing in pulsed-mode Doppler occurs because the signal can only be sampled once per pulse transmission (e.g., the pulse repetition frequency). Frequency aliasing, or the ambiguous characterization of a waveform, occurs in signal processing when a waveform is sampled at less than one half of its fundamental frequency, referred to as the Nyquist condition in signal processing. In a pulsed Doppler system, velocity aliasing will occur when the Doppler shift of the moving particles exceeds half of the pulse repetition frequency. As the pulse repetition frequency is a function of the depth at which a sample is measured, the alias velocity will vary with imaging depth. Increasing the imaging depth will lower the velocity at which aliasing will occur. Continuous wave Doppler signals are typically digitized at higher sampling frequencies, limited by the Nyquist frequency associated with the frequency of the transmission wave. Velocities observed clinically produce Doppler shifts that are generally lower than sampling frequencies used in continuous wave Doppler. As a result, velocity alias is not usually observed with continuous mode Doppler in clinical applications.

Aliased signals that are processed by the measuring system are not directly proportional to the Doppler shift generated by the velocity of the particle but can be related to

the Doppler shift. Undersampling a wave under estimates the frequency and produces a phase shift. Most clinical Doppler machines use the frequency and phase of the sampled wave to infer both velocity magnitude and direction. Velocity alias will produce a lower velocity magnitude with increasing Doppler shift above the alias limit. As frequency is by definition positive and Doppler machines use the signal phase to determine direction, the measured frequency is usually reported as a negative velocity above the alias limit, which is often displayed as an increasing positive velocity magnitude with increasing Doppler shift up to the alias limit. Further increases in the Doppler shift (particle velocity) result in a sign change at the velocity magnitude of the alias velocity with a continued decrease in velocity with increasing Doppler shift. Velocity alias can be reduced or eliminated by frequency unwrapping and baseline shifting, or through the careful selection of machine settings during data acquisition.

Frequency unwrapping is simply correcting the reported aliased velocity by a factor that is related to the alias velocity limit and the magnitude of the reported negative velocity. This correction is, roughly speaking, adding the relative difference in magnitude between the measured aliased velocity and the velocity alias limit to the velocity alias limit. This method of addressing velocity alias is often accomplished by baseline shifting in commercial Doppler machines. In baseline shifting, the phase angle at which a negative velocity is defined is shifted with the effect of a relative shift in the reported alias velocity. Baseline shifting or frequency unwrapping does not eliminate velocity alias but provides a correction to extend the measurement to higher frequencies.

Velocity alias can be "eliminated" by reducing the Doppler frequency of moving particles and thereby shifting the measurable range below the alias limit, which can be accomplished by reducing the carrier frequency of the ultrasound wave, which will in turn reduce the Doppler frequency shift induced by a moving particle and increase the maximum velocity that can be recorded before reaching the Nyquist limit. Alternatively, the Doppler shift frequency can be reduced by increasing the angle between the propagation of the ultrasound wave and the velocity vector, which reduces the magnitude of the Doppler shift frequency by the cosine of this included angle. Angle correction has limitations in that the flow direction must be known and the uncertainty in the correction increases with increasing included angle. As color flow mappers operate in the pulsed Doppler mode, they are subject to velocity alias. Color flow mappers indicate velocity direction by a color series (for example, shades of red or blue). Velocity alias is displayed as a change in a color series from red-to-blue or blue-to-red.

The clinical measurement of many velocity ensembles across a vessel and the integration across the vessel geometry can be time-consuming and problematic in pulsatile flow through a compliant duct. Furthermore, lesions are often complex in shape and cannot be adequately defined by echo. Doppler echocardiographers and scientists have exploited the physics of fluid flow to develop diagnostic tools that complement these capabilities of commercial Doppler systems. The text by Otto (1997) provides an excellent review of these diagnostic tools. Techniques, such as the PISA (proximal isovelocity surface area) or proximal flow convergence, use the capability of color Doppler flow mapping machines to estimate volume flow through an orifice, such as a heart valve. Figure 23.9 illustrates the concept of the proximal flow convergence method.

The flow accelerating toward a small circular orifice will increase in velocity V_a , until a maximum velocity at the orifice V_j is reached. This acceleration occurs in a symmetric pattern around the orifice and is characterized by hemispheres of constant velocity. As the

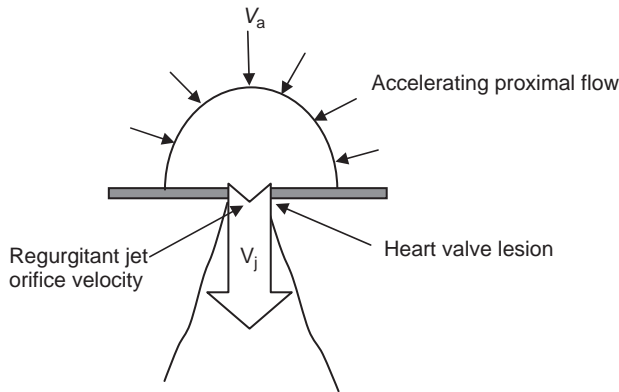


FIGURE 23.9 Illustration of the proximal isovelocity surface area (PISA) concept.

orifice is approached, the velocity increases and the radius of the hemisphere decreases. The regurgitant flow through the orifice can then be calculated as:

$$Q = (2\pi r^2)V_a \quad (23.11)$$

The combined imaging and Doppler characteristics of color Doppler flow mapping are exploited in the PISA approach. The location of the alias velocity in the flow map provides a measure of V_a , which is then coupled with a measure of the radial location from the orifice using the imaging capability of color flow mapping. As flow is velocity times area, the hemispheric assumption provides a shell with a surface area of $2\pi r^2$ that velocity V_a is passing through. Figure 23.10 illustrates the concept of PISA with a color Doppler flow image of valvular regurgitation in a patient.

The PISA approach assumes a round orifice with a hemispherical acceleration zone. In clinical applications, orifices are rarely round and the acceleration zone is not

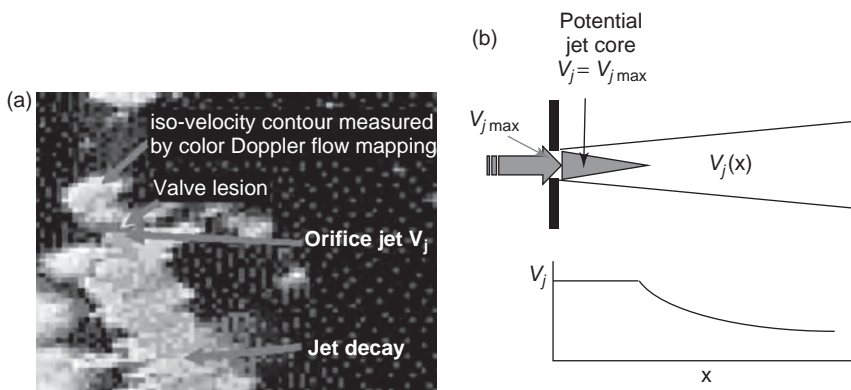


FIGURE 23.10 (a) Color Doppler flow map image of the proximal isovelocity surface area (PISA) in valvular regurgitation. (b) Illustration of the jet decay downstream of an orifice. V is the Jet velocity and x is measured from the orifice.

hemispherical with the result of under or over estimation of the flow rate depending on what radial velocity contour is used in Equation (23.11). The semielliptic method is one approach at considering nonhemispheric geometries in an attempt to correct for errors associated with the PISA technique.

The combination of continuous wave and pulsed Doppler ultrasound is exploited in the turbulent jet decay method of measuring flow through a stenotic lesion or a regurgitant valve. Although continuous wave Doppler does not suffer from velocity alias as does pulsed Doppler, it cannot provide spatial localization of the velocity. The turbulent jet decay method uses continuous wave Doppler to measure the high velocity at the lesion orifice and then uses pulsed.

Doppler to measure the velocity decay at specified location downstream of the orifice. Turbulent jet theory can be used to relate the flow rate of the turbulent jet to the decay of the jet velocity downstream of the orifice, as in Equation (23.12):

$$Q = (\pi V_m^2 x^2) / 160 V_j \quad (23.12)$$

The velocity V_m is measured by pulsed Doppler at location x measured from the jet orifice, whereas the orifice velocity, V_j , is measured by continuous wave Doppler, Figure 23.10b illustrates this decay phenomenon. This equation is valid for round jets and has been extended to jets with other geometries by Cape et al. (1993) with the resulting change to Equation (23.12):

$$Q = (V_m^2 H x) / 5.78 V_j \quad (23.13)$$

where H is the width of the jet measured by color Doppler.

Doppler velocity measurements are also used to estimate pressure gradients in various cardiovascular applications. The Bernoulli equation can be used to estimate the pressure drop across a stenotic lesion or through a valve by measuring the velocity upstream and downstream of the lesion or valve. The Bernoulli equation is

$$\Delta P = (P_1 - P_2) = 1/2 \cdot \rho (V_2^2 - V_1^2) \quad (23.14)$$

where position 1 is often measured upstream of the lesion and position 2 is at the lesion or downstream. In this equation, the pressure drop ΔP has units of Pascals's (Pa). A Pascal is a Newton (N) per square meter, where a Newton has units of mass (kg) times length (m) per time squared. Bioengineering and biomedical applications often use the units of millimeters of mercury (mmHg) in defining a pressure value. A mmHg is related to Pa by the conversion $1 \text{ mmHg} = 133.32 \text{ Pa}$.

Magnetic resonance flow mapping has the advantage over Doppler that it can measure the full three component velocity field over a volume region (Hahn, 1960; Ku et al., 1990; Pettigrew, 1993), which eliminates the uncertainty in flow direction and enables the use of standard fluid dynamic control volume analysis. The advantages of MR flow mapping come at the cost of long imaging times and increased sensitivity to motion artifacts in *in vivo* applications, where phase locking to the heart rate or breathing cycle can increase complexity.

The velocity of moving tissue can be detected by a time of flight technique (Singer and Crooks, 1983) and by phase velocity encoding (Moran et al., 1985; Bryant et al., 1984).

The time-of-flight method tracks a selected number of protons in a plane and measures the displacement of the protons over a time interval defined by the imaging rate. *In vivo* (Matsuda et al., 1987), and phantom (Edelman et al., 1989) studies have shown that the time-of-flight technique is capable of accurate velocity measurement up to velocities at least as high as 0.5 m/s. However, the time-of-flight method requires a straight length of vessel on the order of several centimeters for accurate velocity estimation. This requirement reduces its usability in most *in vivo* applications. The phase velocity encoding method has become the preferred technique in most clinical applications.

Phase velocity encoding directly relates the local velocity of nuclei to the induced phase shift in an imaging voxel. Properly defined bipolar magnetic field gradients are produced in the direction of interest for velocity measurement. The velocity of hydrogen nuclei is then encoded into the phase of the detected signal (Moran, 1982). Chatzimavroudis et al. (2001) and Firmin et al. (1990) provide a discussion of the phase encoding technique with an assessment of its accuracy and limitations for flow measurement.

The technique uses two image acquisitions to provide velocity compensation and velocity encoding. Velocity information in each voxel is obtained by a voxel-by-voxel subtraction of the two images with respect to the phase of the signal. Like Doppler ultrasound, phase velocity encoding can suffer from aliasing effects, alignment error, and limits in spatial and temporal resolution. Velocity estimation using phase shift measurement is limited to a maximum range of phase of 2π radians without ambiguity or aliasing. However, the estimation of the phase shift using phase subtraction between two images reduces that sensitivity to this problem. Studies have been conducted that show MR phase velocity encoding can measure velocities covering the complete physiologic range up to several meters per second (Zhang et al., 2004). Misalignment of the flow direction with the encoding direction will produce a negative bias in the measured flow where the measured velocity will be lower than the true velocity. Like Doppler, this bias follows a cosine behavior where $V_{\text{meas}} = V_{\text{act}} \cos(\theta)$, where V_{meas} is the measured velocity, V_{act} is the actual velocity, and θ is the misalignment angle. This error is typically less than 1% in most applications.

The size of a voxel and the sampling capabilities of the hardware characterize the spatial and temporal resolution of the system. Spatial resolution affects the size of a flow structure that can be measured without spatially filtering or averaging the structure or velocity measurement. Spatial velocity gradients that are small relative to the voxel size will not be adequately resolved and will be averaged over the voxel volume (Kraft et al., 1992). In addition, rapidly varying velocity fluctuations in time will produce a similar low-pass frequency filtering effect if these fluctuations occur with a time scale that is much smaller than the imaging time scale of the measurements. Turbulent flow can produce spatial and temporal scales that could be small relative to the imaging characteristics and can result in what is referred to as signal loss in the image (Suzuki et al., 1990). Stenotic lesions and valvular regurgitation are clinical examples where turbulent flow can occur with spatial and temporal scales that could compromise measurement accuracy.

Phase velocity encoding has the drawback of fairly long imaging or magnet residence times, which is particularly true for three-component velocity mapping. Although long imaging times acceptable for *in vitro* testing with flow loop phantoms, it can present problems and concerns with clinical measurements. Patients can be exposed to long time intervals in the magnetic with the associated problems of patient comfort and care. *In vivo* velocity measurements are often phase-locked with cardiac cycle or breathing rhythm.

Long imaging times can increase potential for measurement error caused by patient movement and variability in the cardiac cycle or breathing rhythm, which can cause noise in the phase-averaged, three-component velocity measurements. Research, in recent years, has focused on hardware and software improvements to increase spatial resolution and reduce imaging time, for example, Zhang et al. (2002).

Magnetic resonance phase velocity encoding provides coupled 3D geometric imaging using traditional MR imaging methods with three-component velocity information. This coupled database provides a powerful diagnostic tool for blood flow analysis and has been used extensively in *in vitro* and clinical applications. Jin et al. (2003) used this coupled imaging flow-mapping capability to study the effects of wall motion and compliance on flow patterns in the ascending aorta. Standard imaging was used to measure aortic wall movement and the range of lumen area and diameter change over the cardiac cycle. This aortic wall motion was phase-matched with phase velocity encoded axial velocity distributions in the ascending aorta. Similar to the PISA approach in Doppler ultrasound, a control volume approach using phase velocity encoded MR velocities can be applied to the assessment of valvular regurgitation (Walker et al., 1995; Chatzimavroudis et al., 1998). The control volume approach is illustrated in Figure 23.11.

23.2.2.3 Laser Doppler Velocimetry The Doppler shift of laser light scattered by particles or cells in a fluid is the basis of laser Doppler velocimetry. Detailed presentations of the LDV technique are provided in the works by Adrian (1983), Drain (1980), and Durst et al. (1976). The scattered radiation, from a laser beam directed at moving particles in a fluid, has a Doppler-shifted frequency defined as

$$f_D \sim (1 - V/C_1)f' \quad (23.15)$$

where C_1 is the speed of light in a vacuum, V is the particle velocity, and f' incident light frequency. The Doppler-shifted frequency is very small relative to the frequency of

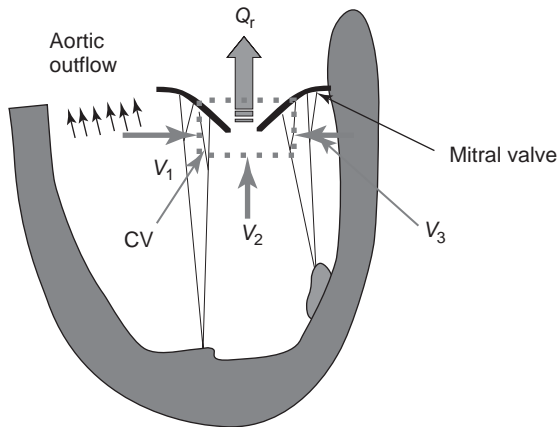


FIGURE 23.11 Illustration of the control volume method in MR phase velocity assessment of valvular regurgitation. The control volume (CV) is the heavy dotted line box around the mitral regurgitant orifice. The box edges are usually selected to correspond with rows and columns in the MR image. V_i represents the 3-component velocities measured with MR through the i faces of the box. Faces 4 and 5 are in the plane of the image at $\pm Z$ offsets from the plane of the image. The regurgitant flow Q_r is the sum of the $V_i A_i$ on each face.

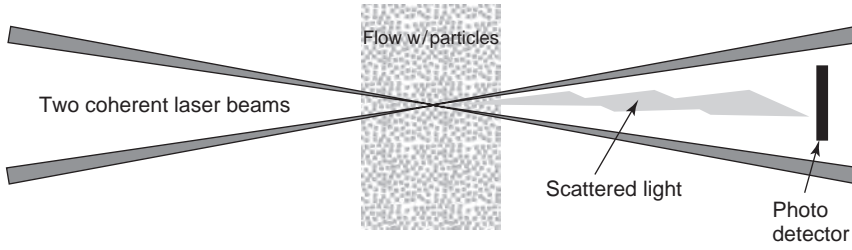


FIGURE 23.12 Illustration of the dual-beam or fringe mode LDV setup.

light and, thus, dual-beam or fringe mode LDV is the system configuration of choice. The dual-beam mode of operation is schematically shown in Figure 23.12. In fringe mode LDV, two coherent laser beams of the same wavelength or frequency are focused to a common point (control volume) in the flow field. The scattered light from a particle moving through the control volume is received by a photodetector.

The crossing of two, coherent, collimated laser beams forms interference fringes as the propagating light waves constructively and destructively interfere with one another. This interference creates a series of light and dark bands with spacing, d_f , of:

$$d_f = \lambda / 2\sin(\kappa) \quad (23.16)$$

The number of fringes, N_{FR} , in the measurement volume is given by:

$$N_{FR} = 1.27d/D_e^{-2} \quad (23.17)$$

where d is the spacing between the two parallel laser beams before the focusing lens and D_e^{-2} is the beam diameter before the lens. Figure 23.13 illustrates the probe geometry generated by the intersection of two focused coherent laser beams with a common wavelength.

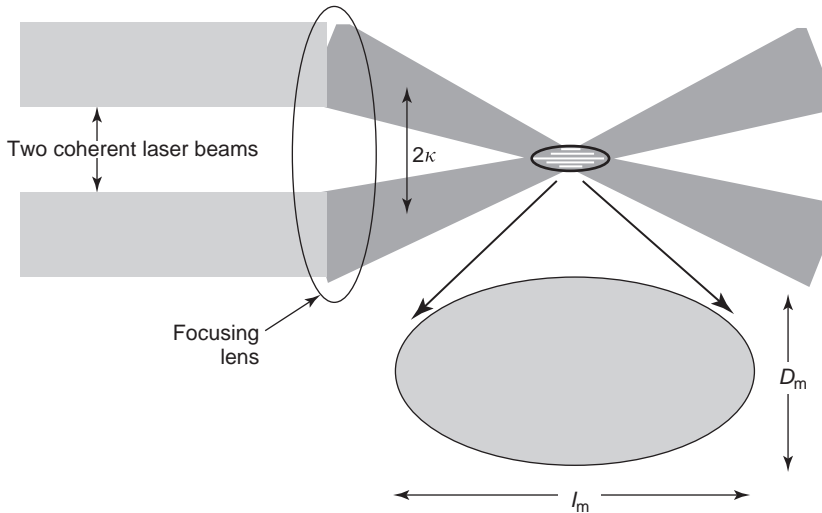


FIGURE 23.13 Illustration of the measurement volume generated in fringe mode LDV.

The spatial resolution of a dual-beam system is affected by the distribution of the light intensity at the intersection of the two focused beams, referred to as the probe or measurement volume. When the laser is operating in the TEM₀₀ mode, the laser cavity sustains a purely longitudinal standing wave oscillation along its axis with no transverse modes. The laser output has an axisymmetric intensity profile that is approximately a Gaussian function of radial distance from the axis. In the far field, the beam divergence is small enough to appear as a spherical wave from a point source located at the front of the lens. A lens is used to focus the beam into a converging spherical wave. The radius of this wave decreases until the focal point of the lens is reached. At the focal point, the beam has almost a constant radius and planar behavior. The beam is focused to its minimum diameter or focal waist, d_e^{-2} and is defined as:

$$d_e^{-2} = (4\lambda f) / (\pi D_e^{-2}) \quad (23.18)$$

where λ is the wavelength of the laser beam and f is the focal length of the lens. A single pair of laser beams generates an ellipsoidal geometry having dimensions of major axis l_m and minor axis d_m given by:

$$l_m = d_e^{-2} / \sin(\kappa) \quad \text{and} \quad d_m = d_e^{-2} / \cos(\kappa) \quad (23.19)$$

where κ is the half angle between the two laser beams, as illustrated in Figure 23.13.

The particle velocity is calculated by the fluctuating light intensity collected by the receiver as the particle passes through the measurement volume and scatters light from the fringes. The intensity change of the scattered light from the light and dark fringes is converted into an electrical signal by a photomultiplier tube (PMT). The electrical signal represents an amplitude-modulated sine wave, with frequency proportional to the Doppler frequency shift (f_D) of the particle traveling through the measurement volume. The particle velocity is then equal to the Doppler frequency multiplied by the fringe spacing. In a two-beam LDV system, the measured velocity component is in the plane of the two laser beams and in the direction perpendicular to the long axis of the measurement volume.

Coherent laser beams with the same frequency produce stationary fringes. A particle moving in either direction across the fringes will produce the same frequency independent of sign, such that a stationary fringe system can only determine the magnitude of the velocity, not the direction. To avoid this directional ambiguity, one of the laser beams of a beam pair is shifted to a different frequency, using a Bragg cell, to provide a moving fringe pattern. One laser beam from each beam pair passes through a transparent medium such as glass, in which acoustic waves, generated by a piezoelectric transducer, are traveling. If the angle between the laser beam and the acoustic waves satisfies the Bragg condition, reflections from successive acoustic wave fronts reinforce the laser beam. The beam exits at a higher frequency and a prism directs the beam to its original direction. The Bragg shift causes the fringes in the probe volume to move at a constant speed in either the positive or negative direction relative to the flow. The measured frequency by the PMT and processor is then the sum or difference of the Bragg cell frequency (typically 40 MHz) and the Doppler shift frequency. This measured frequency is then downmixed with a frequency that is a percent of the Bragg frequency (called the shift frequency) producing a frequency that has a zero shifted to a higher baseline frequency (usually on the order of several mega hertz). This zero shift eliminates directional ambiguity in LDV signal processing.

Laser Doppler velocimetry has excellent spatial and temporal frequency response compared with most other measurement systems. It is considered a gold standard measurement technique in biomedical applications and is the noninvasive measurement system of choice for turbulence measurements. Two disadvantages of LDV worth noting are: (1) LDV noise and (2) velocity bias. The LDV is noisy when compared with other turbulence measurement systems, such as thermal anemometry, because of the use of photomultiplier tubes. These optical detectors, used for their sensitivity and high-frequency response, suffer from higher noise floors than other photo detectors.

Velocity bias is a result of the random sampling characteristics of LDV. As a velocity ensemble is randomly recorded when a particle passes through a probe volume, the statistics of the measured velocity ensembles are not independent of the particle velocity. A greater number of higher speed particles will cross the measurement volume over a specified time than will slower speed particles. Standard ensemble averaging will produce mean velocity estimates that are biased toward higher velocities. This velocity bias can have a significant impact on the velocity statistics, particularly in turbulent flow. In addition to velocity bias, two other biases may occur, fringe bias and gradient bias.

Fringe bias is an error that is minimized by frequency shifting. This type of bias is created by not having enough fringe crossings to satisfy processor validation criteria when calculating a velocity, which occurs when a particle crosses the edge of the probe volume or if the particle velocity is nearly parallel to the fringes. Thus, velocity ensemble averages weight velocities from particles traveling near the center of the measurement volume or those particles that cross more fringes than others. By frequency shifting with a fringe velocity at least two times greater than the flow velocity, particles moving parallel to the fringes can cross the minimum number of fringes for validation by a processor.

Gradient bias results from a non-negligible mean gradient across the probe volume. This bias depends on the fluid flow and the measurement volume dimensions. The mean velocity and the odd order moments are the only statistics affected by gradient bias. In general, LDV transmitting optics are chosen to provide as small a measurement volume as possible to increase spatial resolution and reduce gradient bias. As the LDV measurement volume is longer than it is wide, experiments should be designed to ensure that the LDV optical setup is oriented to position the measurement volume diameter in the direction of the maximum gradients in the flow.

Several postprocessing techniques have been developed to reduce velocity bias. The recommended technique is to use a transit time weighting when computing the velocity statistics. This transit time weighting approximates the ensemble average as a time average. The reader is referred to the reference by Edwards (1987) for a detailed discussion of the transit time technique and its implementation in LDV data processing.

Multiple pairs of laser beams with different wavelength (color) or polarization can be used to produce a multicomponent velocity measuring system. Two or three laser beam pairs can be focused to the same point in the flow. Each beam pair can then be used to independently measure a different component of the velocity vector. As more than one particle can pass through a measurement volume at one time, it is possible to get valid velocity component estimates from different particles. The ellipsoidal geometry of the measurement volumes exaggerates this problem. As a result, LDV data are often processed in one of two modes, random and coincident.

Random mode processing records every valid velocity ensemble as it arrives at the measurement volume, which can generate uneven sample distributions in the different velocity components or LDV channels being measured. Random mode processing has a

negligible impact on mean velocity statistics but can be detrimental to turbulence estimates. Coincident mode processing uses hardware and or software filters to ensure that each velocity component ensemble is measured from the same particle. Filters are used to ensure that the Doppler bursts measured on the different LDV channels overlap in time. Bursts generated by one particle should be measured on each channel with perfect overlap. Time window filters are used to reject bursts that do not occur within a window defined by a percentage of the burst length. The effect of coincident mode processing is usually a reduction in the overall data rate by a factor of at least two, but provides the necessary data quality for turbulence measurements.

Laser Doppler velocimetry is primarily an *in vitro* tool, although systems have been developed for blood flow measurement (Webster, 1998; Tomonaga et al., 1981). Blood is a multiphase fluid composed of a carrier liquid, plasma, and a variety of cells and other biochemical components. Plasma is optically clear to the wavelengths of light used in LDV. The optical opacity of blood is caused by the high concentration of cells, in particular red cells. On the microscopic level, however, blood can transmit light over a short distance because of the plasma carrier fluid. Clinical style probes have been developed to measure the velocity of blood cells in blood using catheter type insertion into vessels of suitable size or through transcutaneous measurement of capillary flow below the skin. These *in vivo* systems are designed with very short focal length transmitting lenses providing a small measurement volume located a very short distance from the transmitting lens. Laser light is propagated through the plasma and focused a few millimeters from the probe tip. Blood cells act as particles in the fluid and scatter light that is collected by the transmitting lens and directed to a PMT system for recording of the Doppler bursts. Manning et al. (2005) and Ellis et al. (1996) have used LDV to measure the velocity fields around mechanical heart valves in *in vitro* studies. Figure 23.14 shows the measured velocity distributions associated with impact of a Bjork–Shiley monostrut valve (Manning et al., 2005).

23.2.2.4 Particle Image Velocimetry and Particle Tracking Velocimetry Particle image velocimetry (PIV) and particle tracking velocimetry (PTV) have been applied in fluid flow applications for over a decade. They are noninvasive experimental techniques that provide a quantitative, instantaneous velocity vector field with good spatial resolution; an appealing feature when studying complex, time-dependent fluid flows that can occur in biomedical applications. The methods allow the quantitative visualization of instantaneous flow structures over a spatial region, as opposed to a point measurement like LDV, which can provide insight into the flow physics. The two techniques, PIV and PTV, differ in the way particle displacements are measured. Particle tracking follows and tracks the motion of individual particles, whereas PIV measures the displacement of groups of particles within the flow. PTV, although not commonly used, is a subset of the more common PIV technique and is still used in specific applications of PIV. Raffel et al. (1998) provides a comprehensive discussion of the PIV and PTV techniques with a detailed presentation of the physics, processing tools, capabilities, and errors.

The instantaneous velocity field is computed by measuring an instantaneous particle displacement field over a specified, finite time interval. Laser-based PIV and PTV are non-invasive velocity measurement tools that require good optical access to the flow field. As a result, they are essentially *in vitro* tools (Hochareon et al., 2004; Oley et al., 2005) that are of limited use *in vivo*. Figure 23.15 shows an example of the use of PIV in a bioengineering application of flow through one chamber of an artificial heart (Hochareon, 2003). X-ray-based PTV systems are being developed and will be capable of *in vivo* use. In this

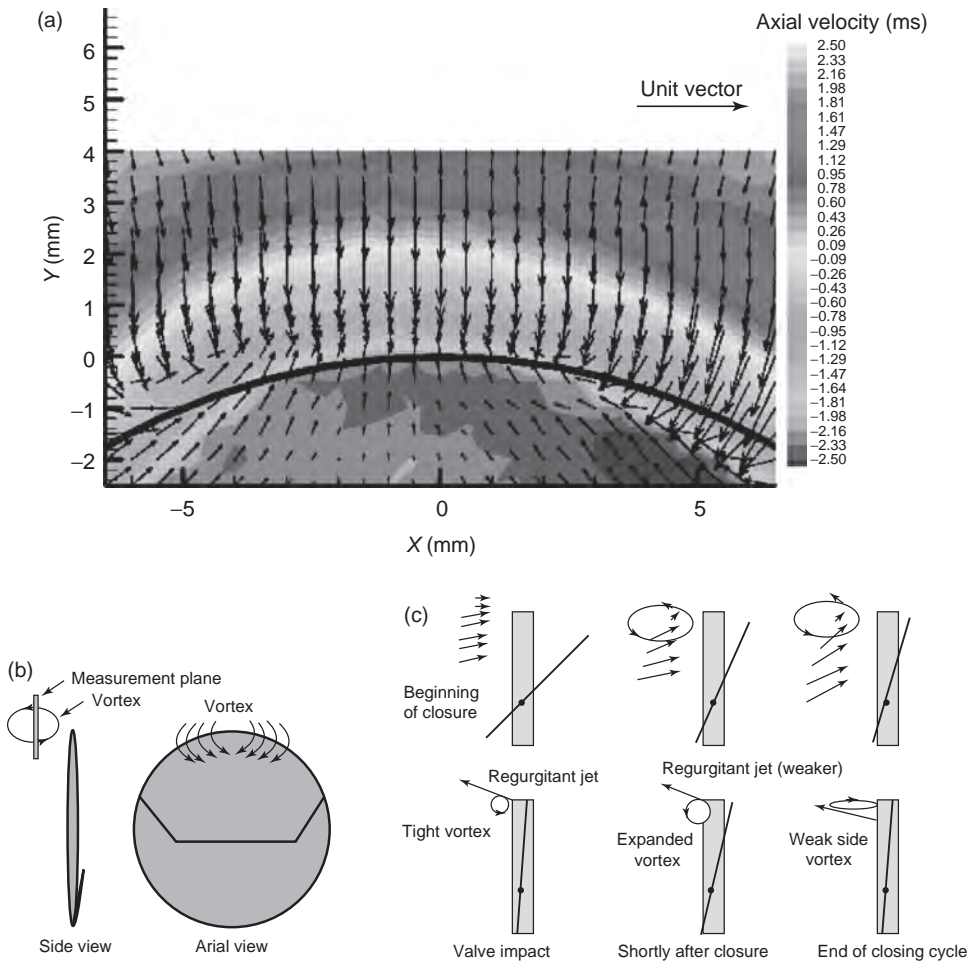


FIGURE 23.14 3D phase averaged velocity map of major orifice regurgitant flow in Björk–Shiley monostrut valve. (a) 3 mm from valve housing, 4 ms after impact. (b) Illustration of measurement plane and vortex flow pattern. (c) Flow field schematic during valve closure determined from multi-component LDV measurements (Manning et al., 2005).

section, we will focus on PIV; however, system concepts (seeding, acquisition, processing, noise, and errors) would be applicable to some degree to systems like X-ray PTV (Lee and Kim, 2003).

Particle image velocimetry uses a double-pulsed light source, usually a laser, to illuminate a thin sheet in the flow field. Particles suspended in the fluid scatter light during each pulse, and this scattered light is recorded on a digital camera. The optimal setup has the recording device located 90° to the illumination sheet. Figure 23.16 illustrates the typical PIV setup.

Two lasers with coincident beam paths are used to illuminate a desired plane of the flow by incorporating optics to produce thin laser sheets. During image acquisition, the two lasers are pulsed with the specified time separation (typically $1\text{--}1000\ \mu\text{s}$). A trigger system, referred to as a synchronizer, controls the firing of the two lasers relative to the shuttering

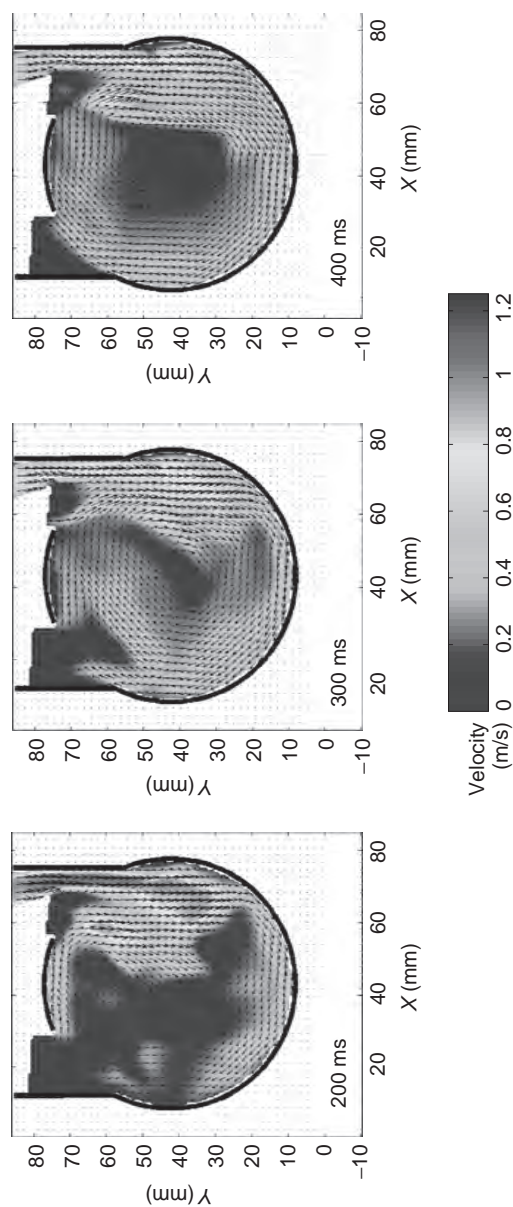


FIGURE 23.15 Phase-average velocity maps from mid to late diastole for a prototype artificial heart ventricular chamber (time reference is from the onset of diastole, 4.7 L/min CO, 75 bpm) (Hochareon, 2003).

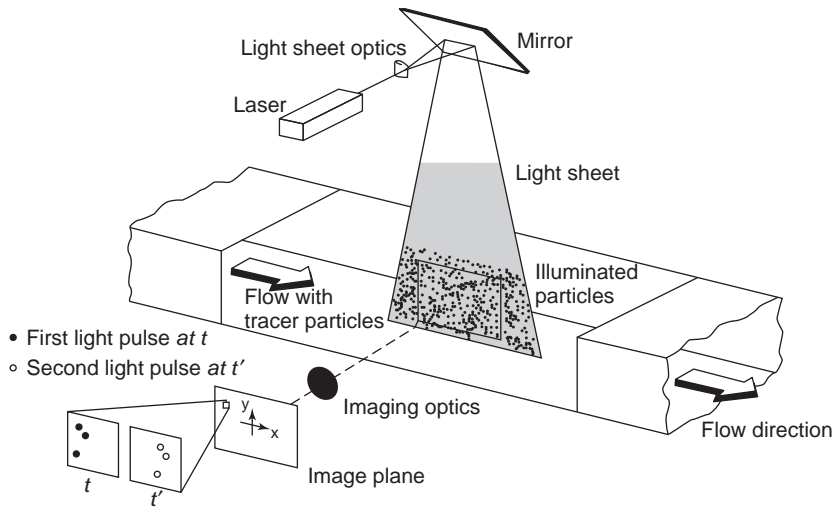


FIGURE 23.16 Schematic of a PIV setup. (Raffel et al., 1998; Chapter 1, p. 4, Fig. 1.4., with kind permission of Springer Science and Business Media.)

of a CCD camera. The camera, usually placed orthogonal to the laser sheet, collects the light scattered by tracer particles in the flow and records an image. The synchronizer, used in crosscorrelation-based PIV systems, delays the firing of the first laser such that the camera shutter is centered between the firing of the two lasers. This synchronization technique is called frame straddling and produces two sequential images of each laser beam pulse.

Although the time between successive camera frames may be much larger than the time duration between laser pulses, the two images of the particle field created are separated by the specified time interval between the two laser pulses.

A discussion of PIV must begin with a brief introduction of the terminology commonly used. Figure 23.17 provides a schematic representation of geometric imaging. The “light plane” is the “volume” of the fluid illuminated by the light sheet. The “image plane” is the image from the light plane captured on the CCD sensor. It is important to note that the light plane is a 3D space or “volume,” whereas the image plane is a 2D space or “surface.” The subvolume selected from the light plane for cross correlation is called the “interrogation subvolume.” The corresponding location of this interrogation volume captured on the image plane is called the “interrogation subregion.” Please note that the displacement vectors in an interrogation volume are three-component vectors, whereas those in an interrogation area are two-component vectors. “Particle” is the physical particle suspended in the fluid volume. “Particle image” is the image of each particle in the image plane. Particle density, intensity, and pair refer to particle properties, whereas image density, intensity, and pair refer to particle image properties.

Most commercial systems use a crosscorrelation-based image processing technique to compute the particle displacement. Images are subdivided into small “interrogation regions,” typically a fraction of the overall image size, and the same two subregions are crosscorrelated between the two images to obtain the average displacement of the particles within that subregion. From this displacement and the known time delay, the velocities within the interrogation region are obtained.

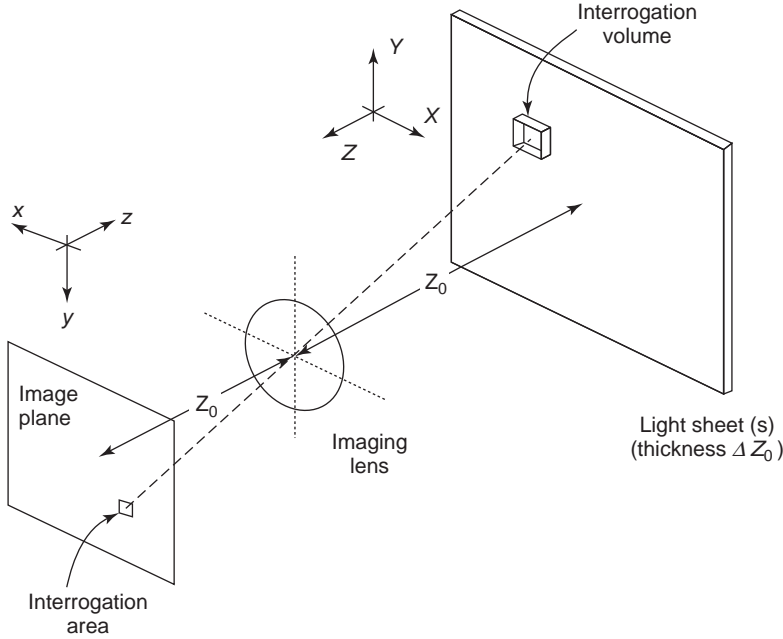


FIGURE 23.17 Schematic representation of geometric imaging. (Raffel et al., 1998; Chapter 3, p. 62, Fig. 3.1., with kind permission of Springer Science and Business Media.)

Statistical PIV assumes all particles in an interrogation subregion move a similar distance and direction. The processing algorithm then computes the mean displacement vector for the particles in the interrogation volume. Therefore, the local particle distribution pattern captured on each exposure should be similar; but the group of local particles is displaced from image to image. Statistical PIV is then “pattern recognition” of the particle distribution within an interrogation subregion, instead of the averaged particle displacements. Sophisticated pattern recognition schemes have been developed by a number of researchers; however, the crosscorrelation tends to be the algorithm of choice. The use of a crosscorrelation as opposed to an autocorrelation eliminates directional ambiguity in the velocity measurement. Most commercial systems use advanced crosscorrelation algorithms, such as the Hart correlation, developed to improve signal to noise in the correlation estimate and enhance resolution (Hart, 1999, 2000).

The crosscorrelation function for two interrogation subregions of frames A and B is defined by:

$$R_{II}(s, \Gamma, D) = \langle I(x, \Gamma) I'(x + s, \Gamma' D) \rangle \quad (23.20)$$

where s is the shifting vector of the second interrogation window, Γ is the series of location vectors for each particle in the interrogation volume, D is the displacement vector for each particle, x is the spatial domain vector within the interrogation area, I is the intensity matrix for the interrogation area from frame A, and I' is the intensity matrix for the interrogation area from frame B. A detailed mathematical derivation of the crosscorrelation for (group) particle tracking is beyond the scope of this presentation. The location of the maximum value of crosscorrelation function represents the mean particle displacement

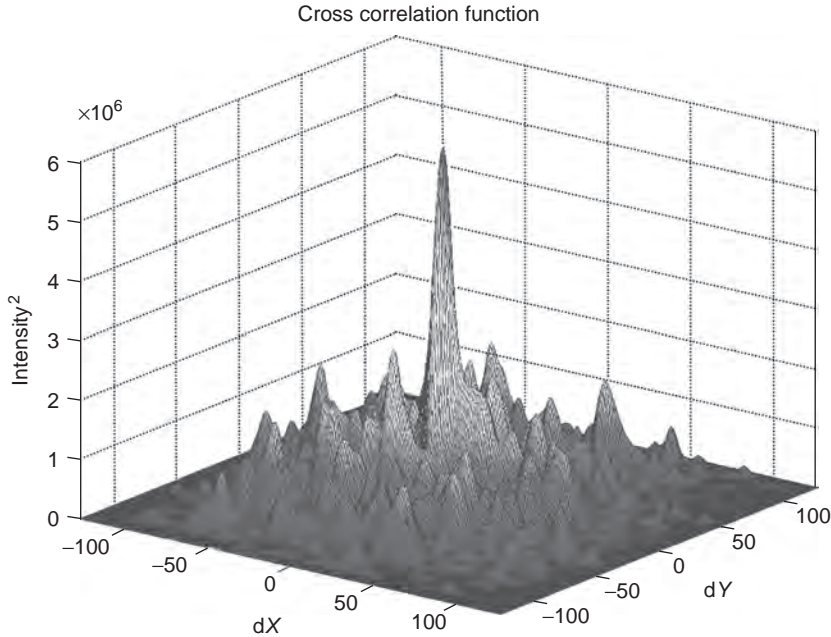


FIGURE 23.18 Representative crosscorrelation map between frames A and B.

for the interrogation image. Figure 23.18 is an example of a crosscorrelation function between two images.

The location of the crosscorrelation peak should be determined with subpixel accuracy. Several curve fitting algorithms have been developed to identify the peak in the crosscorrelation. Gaussian, Parabolic, and Centroid are three common methods in commercial software, although others exist. A Gaussian peak fit is most commonly used because the crosscorrelation function of two Gaussian functions also yields a Gaussian function, which means if the intensities of individual particle images in the interrogation area fit a Gaussian shape, the crosscorrelation function can be accurately approximated by the Gaussian function, which occurs only under the condition of low displacement gradient, so that the particle distribution pattern is preserved in windows A and B. The crosscorrelation function in distorted, particle image intensity distribution patterns are less accurately approximated by a Gaussian distribution. A three-point estimator for Gaussian fit works better with a narrow and symmetric peak. Centroid peak finding should be considered when the crosscorrelation peak is expected to have asymmetric and irregular patterns. Such cases occur for particle images larger than 2–3 pixels in diameter, for a low intensity ratio of particle image to background noise, or for a high gradient displacement field. For “correlation-based correction,” the centroid peak finding might be more suitable than Gaussian because the multiplications could distort the crosscorrelation peak.

The use of a digital CCD camera presents an error source known as “peak locking.” This error impacts the accuracy of the subpixel estimation in the correlation peak and thus impacts the velocity measurement. This error will be discussed later.

Like LDV and Doppler, PIV and PTV require that the fluid be seeded with tracer particles that follow the fluid motion. The particle density, the number of particles per unit volume of fluid, determines what technique should be used, PIV or PTV. Flows with a low particle density are more suited to PTV, whereas PIV works best in high particle density flows. It is assumed that the tracer particles follow the flow accurately to give an exact representation of the flow field at various times and locations. The particle density, however, should be sufficiently low to preserve the original flow dynamics. Such a dilute condition is expressed by the inequality:

$$(\rho_p \pi d_p^4 v_r) / (18 \mu \delta_p^3) < 1 \quad (23.21)$$

where d_p and ρ_p are the particle diameter and density, respectively, μ is fluid viscosity, v_r is the averaged particle velocity relative to neighboring particles, and δ_p is the average distance between the particles.

Particles must be small enough to follow the fluid flow but large enough to be seen. The particle relaxation time, τ_s , should be small relative to the flow time scales.

$$\tau_s = d_p^2 (\rho_p / 18 \mu) \quad (23.22)$$

In practice, τ_s is made small by using very small particles. The Stokes number for PIV experiments, St_{PIV} , can be defined as τ_s / τ_{PIV} , where τ_{PIV} is the small finite separation time between two observations (pulse separation time). St_{PIV} should be much less than 1 to assure negligible particle-fluid velocity differences over the pulse separation. However, particles must be large enough to scatter sufficient light energy to be visualized on the recording device (e.g., a CCD camera) with the goal that a particle image is at least several pixels in size. Increasing the light source energy can improve visibility, but a saturation point is reached where increasing light source energy does not help. Furthermore, high energy can damage windows and plastic test models.

The time separation of the two laser pulses must be small enough to minimize particle loss through too large a particle displacement between the first and second frames of the interrogation window. However, the time separation must be long enough to permit adequate displacement of particles at the lowest measurable velocities in each velocity component and to minimize the impact of pixel peak locking (Christensen, 2004). Complex and highly 3D flows can be biased in a 2D PIV system. PIV can provide very high spatial resolution, but suffers from low temporal resolution. Furthermore, high magnification imaging used in high resolution PIV introduces additional constraints and limitations that must be overcome to achieve high-quality vector maps.

The challenge for PIV is to correctly track particle motion. Figure 23.19 shows an example of a PIV particle image. The statistical crosscorrelation approach is used to track the displacement of a “group” of particles within an indicated small volume or subregion. The location of a velocity vector is at the center of the subregion spot. The spatial resolution for a “velocity vector” in PIV is the size of the interrogation subregion. Overlapping adjacent interrogation subregions is commonly used to reduce the distance between adjacent vectors and provide additional vectors in the overall vector map. However, this does not increase the spatial resolution of the velocity field, which is limited to the size of the interrogation subregion.

Commercial PIV systems use a multigrid recursive method to reduce interrogation subregion size. In the hierarchical approach, a PIV measurement with large interrogation

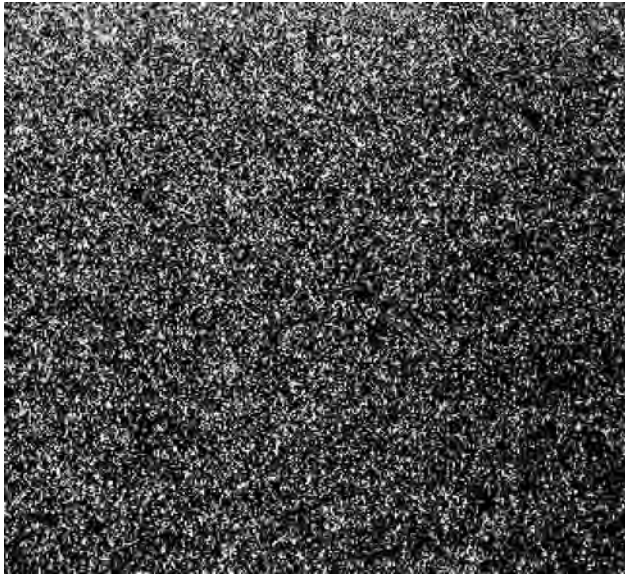


FIGURE 23.19 Example of a PIV particle image.

subregions is first computed. Subsequently, the initial interrogation area is evenly divided into smaller areas for processing. Each smaller interrogation area is offset by the displacement obtained from its parent interrogation area. This process is repeated until the smallest possible interrogation size is reached (for example, $128 \times 128 \rightarrow 64 \times 64 \rightarrow 32 \times 32 \rightarrow 16 \times 16$) for the given flow field. An iterative method can be applied at the final grid level. Similar to the multigrid method, the iterative method uses the obtained displacement from the first crosscorrelation to offset the second window for the next crosscorrelation. The difference is that it does not break down the interrogation areas into smaller windows. This process is repeated until the difference between the displacements from successive crosscorrelations is less than one pixel. If the window B is shifted by the converged displacement, windows A and B should be virtually the same, as long as the gradient is sufficiently low. Iterative crosscorrelation is another way to increase accuracy.

A minimum number of particle pairs (on the order of 10) are required in PIV processing. The particle density in the flow will determine the minimum size subregion that can be used to obtain adequate vector maps. Thus, the spatial resolution is governed by the particle density. In near-solid surfaces, the particle density is often lower in flows with strong wall gradients. Reducing the interrogation window size increases spatial resolution. However, an overly small window causes in-plane particle loss because of particles moving out of the interrogation spot. Several techniques exist to capture the particles moving out of the window without enlarging the interrogation spot. The first is simply to enlarge the second window to cover the expected displaced particle. The original interrogation window (frame A) is enlarged to the same size as window B and zero-padded at the extended part. The second technique is to offset the second window to the location of anticipated displacement.

Errors in PIV processing can occur from several sources. The spatial resolution for a velocity vector is the dimension of the interrogation volume. If the particles are evenly

distributed, the center of an interrogation volume can be used as the vector location. The accuracy of the displacement depends on both the subpixel accuracy of the peak finding algorithm and the image quality. A one-tenth pixel is the best accuracy (Raffel et al., 1998). Time resolution for the velocity is the separation time between two pulses, as the information during this period is not recorded. The velocity error is composed of systematic and residual error. Systematic errors come from the algorithm and experiment setting or image quality, which can be minimized and uncertainty in the time separation. Residual errors are inherent in the processing, such as errors caused by the peak finding algorithm. The residual errors are usually not a function of Δt . Therefore, a too small separation time increases the velocity error as this error is proportional to $1/\Delta t$.

The following discussion is relevant to the effect of image quality on PIV accuracy. Large particle images can result in wide crosscorrelation peaks, which can reduce the accuracy of the peak finding algorithm. In addition, large particles require larger interrogation spots to contain an appropriate number of particles, which leads to a reduction in spatial resolution. Particle images smaller than two pixels in diameter, or particle displacements that are less than two pixels, can introduce a displacement bias, called "peak locking." The displacement peaks tend to be biased toward integer values. Peak locking presents itself as a "staircased" velocity pattern, in a region with a velocity gradient where the velocity distribution should be smooth. The calculation of spatial derivatives of this vector map then produces a mosaic pattern in the gradient map. Figure 23.20 illustrates these patterns. Techniques, such as multigrid or continuous window-shifting or iterative image deformation, have been proposed to overcome peak locking. Image preconditioning, such as filtering, or defocusing can optimize the image diameter. The resolution of the CCD sensor, therefore, also limits the use of smaller particles to increase the velocity resolution.

The methods developed to increase displacement accuracy rely on the assumption of low displacement gradient. High gradient tends to bias the displacement toward low values because the particles with smaller displacements are more likely to remain in the interrogation volume longer than those with higher displacements. This bias can be minimized by reducing the size of the interrogation volume and separation time. For high distortion of the particle pattern in a high gradient spot, the centroid peak finding algorithm is more suitable than the Gaussian. However, PTV, as it follows an individual particle, is not affected by high gradient. Several research groups use the displacement results from PIV to guide the local search areas for PTV in a coupled PIV/PTV processing algorithm. These coupled techniques relieve the gradient limit in PIV and increases resolution of both velocity vectors and velocity fields.

The motion across the light sheet in highly 3D flows can bias the local velocity estimation caused by perspective projection, if the particle has not left the light sheet. The effect of perspective projection velocity bias, illustrated in Figure 23.21, is usually more severe at the edges of the image, where the projection angle increases. At high image magnification, the focal length becomes shorter and the projection angle increases, which worsens the perspective projection. Strong perspective projection could vary the magnification factor through the image plane resulting in an image distortion.

In general, the light sheet thickness is smaller than the depth of focus, δ_z . The light sheet thickness, therefore, determines the thickness of the effective interrogation volume: All illuminated particles are well focused. Most commercial systems using standard-grade Yag lasers with appropriate optics can generate light sheets that have a thickness on the order of one to two hundred microns, although light sheets can be as thick as a 1 mm.

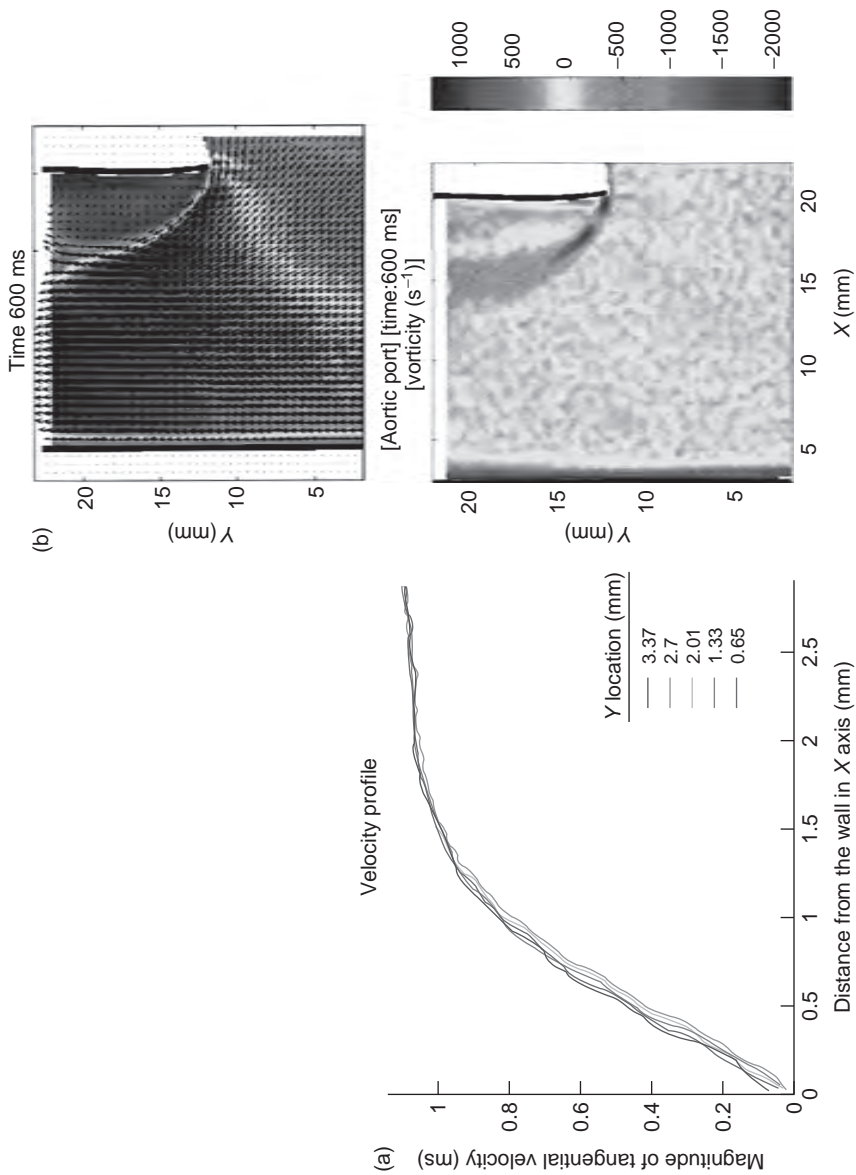


FIGURE 23.20 (a) Staircased pattern in a velocity profile at a wall. (b) Gradient field calculation (bottom image) of a measured velocity field (top image) showing mosaic pattern (Hochareon, 2003).

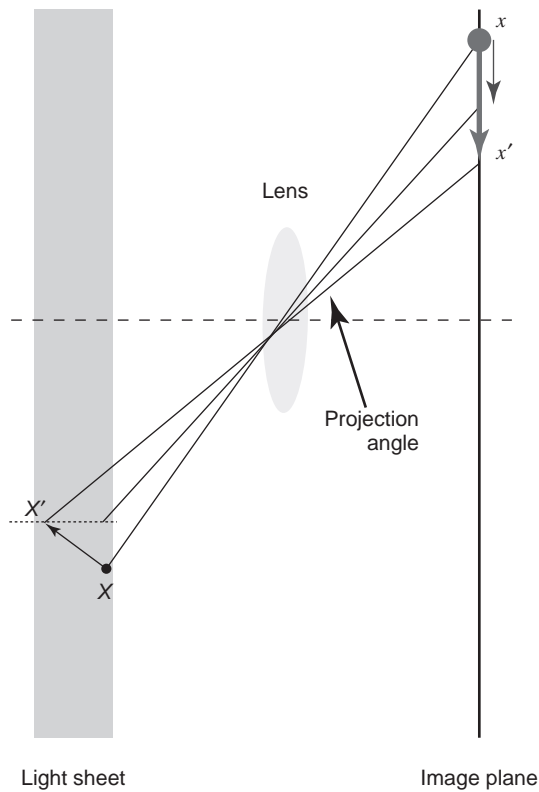


FIGURE 23.21 Illustration of the perspective projection; red arrow is the vector obtained from perspective projection; blue is the correct projection of displacement vector on the XY plane (Hochareon, 2003).

In high magnification imaging, the depth of focus can become smaller than the light sheet thickness. The thickness of the effective interrogation volume is then constrained to the depth of focus. In this case, particles located beyond the focal plane but within the illuminated plane are out-of-focus and appear as highly nonuniform background image noise, and can affect the crosscorrelation. In addition, the thickness of the effective interrogation volume determines the tolerance to out-of-plane motion. A smaller effective volume thickness increases the probability for out-of-plane particle loss. In general, the estimated maximum out-of-plane motion should be less than one fourth of the effective volume thickness.

Data validation is another source of uncertainty in PIV. Bad velocity vectors will ultimately appear within a vector map because of noise in the displacement estimation. Several filtering algorithms have been developed to remove these bad vectors. These filter routines operate on thresholding velocity at a particular location and using the magnitude of the velocity estimate; the mean, the median, or the rms within a predefined subregion; or other more complicated thresholds to low-pass filter the estimates. Improper application of a validation scheme can over filter the velocity map and throw away good data. For example, rms validation techniques should be carefully used in turbulent shear layers where high rms values are normally encountered. It is possible to inadvertently filter good

instantaneous turbulent velocity ensembles with a tight rms filter setting. In general, some knowledge of the flow under study is needed to accurately perform vector validation.

Commercial PIV systems can be two-component or three-component, planar or volume systems. A two-component, planar PIV system provides information on two components of the velocity vector. In two-component PIV, the measured displacement vector is the projection of the three-component velocity vector on the 2D plane of the light sheet. Flow information for highly three-component flows can be inaccurately represented by planar PIV images. Stereographic and holographic PIV systems have been developed for three-component measurement in a plane or volume, respectively. Although the instantaneous velocity field obtained by PIV is an advantage over LDV (or Doppler), two-exposure PIV only provides information on the particle motion during the two exposures and also suffers from poor temporal frequency response in the measurement of adjacent vector maps in time. Particle acceleration cannot be measured by direct two-exposure PIV. Four-exposure systems have been developed to permit calculation of the particle acceleration by Hassan and Phillip (1997) and Lui and Katz (2004), although the temporal resolution for the acceleration is not yet comparable to that of LDV.

REFERENCES

- Adrian RJ. Particle imaging techniques for experimental fluid mechanics. *Annual Review of Fluid Mechanics* 1991;23:261–304.
- Adrian RJ. Laser velocimetry. In: *Fluid Mechanics Measurements*. New York: Hemisphere Publishing; 1983.
- Baldwin JT, Tarbell JM, Detusch S, Gaselowitz DB, Rosenberg G. Hotfilm wall shear probe measurements inside a ventricular assist device. *Journal of Biomechanical Engineering* 1988;110(4):326–333.
- Baldwin JT, Tarbell KM, Deutsch S, Geselowitz DB. Wall shear stress measurements in a ventricular assist device. *Journal of Biomechanical Engineering* 1988;110:326–333.
- Batten JR, Nerem RM. Model study of flow in curved and planar arterial bifurcations. *Cardiovascular Research* 1982;16(4):178–186.
- Beldi G, Bosshard A, Hess OM, Althaus U, Walpoth BH. Transit time flow measurement: experimental validation and comparison of three different systems. *Annals of Thoracic Surgery* 2000;70:212–217.
- Bendat JS, Peirsol AG. *Random Data: Analysis and Measurement Procedures*. New York: John Wiley & Sons; 1986.
- Brunn HH. *Hot Wire Anemometry: Principles and Signal Analysis*. New York: Oxford University Press; 1995.
- Bryant DJ, Payne JA, Firmin DN, Longmore DB. Measurement of flow with NMR imaging using a gradient pulse and phase difference technique. *Journal of Computer Assisted Tomography* 1984;8:588–593.
- Cape EG, Nanda NC, Yoganathan AP. Quantification of regurgitant flow through Bileaflet heart valves: theoretical and in vitro studies. *Ultrasound in Medicine and Biology* 1993;19:461–468.
- Caro CG, Pedley TJ, Schroter RC, Seed WA. *The Mechanics of the Circulation*. New York: Oxford University Press; 1978.
- Chatzimavroudis GP, Oshinski JN, Franch RH, et al. Evaluation of the precision of magnetic resonance phase velocity mapping for blood flow measurements. *Journal of Cardiovascular Magnetic Resonance* 2001;3:11–19.

- Chatzimavroudis GP, Oshinski JN, Pettigrew RI, et al. Quantification of mitral regurgitation with magnetic resonance phase velocity mapping using a control volume method. *Journal of Magnetic Resonance Imaging* 1998;8:577–582.
- Christensen KT. The influence of peak-locking errors on turbulence statistics computed from PIV ensembles. *Experiments in Fluids* 2004;36(3):484–497.
- Coleman HW, Steele WG. *Experimentation and Uncertainty Analysis for Engineers*. New York: John Wiley & Sons; 1999.
- Comte-Bellot G. Hot-wire anemometry. *Annual Review of Fluid Mechanics* 1976;8:209–232.
- Crowe CT, Sommerfeld M, Tsuji Y. *Multiphase Flows with Droplets and Particles*. Boca Raton, FL: CRC Press; 1998.
- Drain LE. *The Laser Doppler Technique*. New York: John Wiley & Sons; 1980.
- Drost CJ. Vessel diameter-independent volume flow measurements using ultrasound. Proceedings of San Diego Biomedical Symposium 1978;17:299–302.
- Durst F, Melling A, Whitelaw JH. *Principles and Practice of Laser Doppler Anemometry*. San Diego, CA: Academic Press; 1976.
- Edelman RR, Heinrich PM, Kleefield J, Silver MS. Quantification of blood flow with dynamic MR imaging and presaturation bolus tracking. *Radiology* 1989;171:551–556.
- Edwards R. V. Report of the special panel on statistical particle bias problems in laser anemometry. *Journal of Fluids Engineering and Transactions of the ASME* 1987;109(2):89–93.
- Ellis JT, Healy TM, Fontaine AA, Westin MW, Jarret CA, Saxena R, Yoganathan AP. An *in vitro* investigation of the retrograde flow fields of two bileaflet mechanical heart valves. *Journal of Heart Valve Disease* 1996;5:600–606.
- Falsetti HL, Carroll RJ, Swope RD, Chen CJ. Turbulent blood flow in the ascending aorta of dogs. *Cardiovascular Research* 1983;17(7):427–436.
- Firmin DN, Nayler GL, Kilner PJ, Longmore DB. The application of phase shifts in NMR for flow measurement. *Magnetic Research Medicine* 1990;14:230–241.
- Fung YC. *Respiratory gas flow*. In: *Biomechanics: Motion, Flow, Stress and Growth*. New York: Springer-Verlag; 1990, Chapter 7.
- Hahn EL. Detection of sea-water motion by nuclear pre-emission. *Journal of Geophysical Research* 1960;65:776–777.
- Hampers CL, Schuback E, Lowrie EG, Lazarus JM. Clinical engineering in hemodialysis and anatomy of an artificial kidney unit. In: *Long Term Hemodialysis*. New York: Grune and Stratten; 1973.
- Hart DP. Super-resolution PIV by recursive local correlation. *Journal of Visualization* 1999;10:1–10.
- Hart DP. PIV error correction. *Experimental Fluids* 2000;29(1):13–22.
- Hassan YA, Phillip OG. A new artificial neural network tracking technique for particle image velocimetry. *Experimental Fluids* 1997;23(2):145–154.
- Hochareon P, Manning KB, Fontaine AA, Tarbell JM, Deutsch S. Wall shear-rate estimation within the 50cc Penn State artificial heart using particle image velocimetry. *Journal Biomechanical Engineering* 2004;126:430–437.
- Hochareon P, Manning KB, Fontaine AA, Tarbell J, Deutsch S. Correlation of *in vivo* clot deposition with the flow characteristics in the 50cc Penn State artificial heart: a preliminary study. *Journal of ASAIO* 2004;50(6):537–542.
- Hochareon P. Development of particle image velocimetry (PIV) for wall shear stress estimation within a 50cc Penn State artificial heart ventricular chamber, Ph.D. Thesis, Bioengineering Department, Penn State University; University Park, PA; 2003.
- Jin S, Oshinski J, Giddens DP. Effects of wall motion and compliance on flow patterns in the ascending aorta. *Journal of Biomechanical Engineering* 2003;125:347–354.

- Kraft KA, Fei DY, Fatouros PP. Quantitative phase-velocity MR imaging of in-plane laminar flow: effect of fluid velocity, vessel diameter, and slice thickness. *Medical Physics* 1992;19:79–85.
- Ku DN, Biancheri CL, Pettigrew RI, et al. Evaluation of magnetic resonance velocimetry for steady flow. *Journal of Biomechanical Engineering* 1990;112:464–472.
- Latto B, El O, Riedy, Vlachopoulos J. Effect of sampling rate on concentration measurements in nonhomogeneous dilute polymer solution flow. *Journal of Rheology* 1981;25:583–590.
- Lauchle GC, Billet ML, Deutsch S. Hydrodynamic measurements in high speed liquid flow facilities. In: Gadel-Hak M, editor *Lecture Notes in Engineering, 46, Experimental Fluid Mechanics*. New York: Springer Verlag; 1989.
- Lee SJ, Kim GB. X-ray particle image velocimetry for measuring quantitative flow information inside opaque objects. *Journal of Applied Physics* 2003;94:3620–3623.
- Lipowski HH, McKay CB, Seki J. Transit time distributions of blood flow in the microcirculation. In: Lee and Skalak, editors *Microvascular Mechanics*. New York: Springer Verlag; 1989, p. 13–27.
- Lui X, Katz J. Measurements of pressure distribution in a cavity flow by integrating the material acceleration. Proceedings of 2004 Heat Transfer and Fluids Engineering Conf., ASME HTFED04-56373, July, 2004.
- Lynnworth LC. *Ultrasonic flow meters*. In: Mason WP, Thurston RN, editors. *Physical Acoustics*. New York: Academic Press; 1979.
- Manning KB, Przybysz TM, Fontaine AA, Tarbell JM, Deutsch S. Near field flow characteristics of the Bjork–Shiley monostrut valve in a modified single shot valve chamber. *ASAIO Journal* 2005;51(2):133–138.
- Matsuda T, Shimizu K, Sakurai T et al. Measurement of aortic blood flow with MR imaging: comparative study with Doppler ultrasound. *Radiology* 1987;162:857–861.
- Merzkirch W. *Flow Visualization*. New York: Academic Press; 1974.
- Montgomery DC. *Design and Analysis of Experiments*, 3rd ed., New York: John Wiley & Sons; 1991.
- Moran PR, Moran RA, Karstaedt N. Verification and evaluation of internal flow and motion: true magnetic resonance imaging by the phase gradient modulation method. *Radiology* 1985;154(2):433–441.
- Moran PR. A flow velocity zeugmatographic interlace for NMR imaging in humans. *Magnetic Resonance Imaging* 1982;1:197–203.
- Mori Y, Nakayama W. Study on forced convective heat transfer in curved pipes. *International Journal of Heat Mass Transfer* 1965;8:67–82.
- Nerem RM, Rumberger JA, Gross DR, Muir WW, Geiger GL. Hot film coronary artery velocity measurements in horses. *Cardiovascular Research* 1976;10(3):301–313.
- Nerem RM, Rumberger JA, Gross DR, Hamlin RL, Geiger GL. Hot film anemometer velocity measurements of arterial blood flow in horses. *Circulation Research* 1974;34(2):193–203.
- Neuman MR. Therapeutic and prosthetic devices. In: Webster, editor *Medical Instrumentation: Application and Design*. New York: John Wiley & Sons; 1998, Chapter 8.
- Oley LA, Manning KB, Fontaine AA, Deutsch S. Off design considerations of the 50cc Penn State ventricular assist device. *Artificial Organs*, 2005, in print.
- Otto CM. *The Practice of Clinical Echocardiography*. Philadelphia PA: W.B. Saunders Co.; 1997.
- Pedley TJ, Schroter RC, Sudlow MF. Flow and pressure drop in systems of repeatedly branching tubes. *Journal of Fluid Mechanics* 1971;46 (part 2):365–383.
- Pettigrew RI. Magnetic resonance in cardiovascular imaging. In: Zaret BL et al. editors. *Frontiers in Cardiovascular Imaging*. New York: Raven Press; 1993, Chapter 9.
- Povey MJW. *Ultrasonic Techniques for Fluids Characterization*. San Diego, CA: Academic Press; 1997.

- Primiano FP. Measurements of the respiratory system. In: Webster, editor *Medical Instrumentation: Application and Design*. New York: John Wiley & Sons; 1998, Chapter 9.
- Raffel M, Willert CE, Kompenhans J. *Particle Image Velocimetry: A Practical Guide*. New York: Springer; 1998.
- Rouse H, Ince S. *History of hydraulics. Iowa Institute of Hydraulics Research Report*, Iowa State University; Ames, Iowa, 1957.
- Shaughnessy EJ, Katz IM, Schaffer JP. *Introduction to Fluid Mechanics*. New York: Oxford University Press; 2005.
- Siedband MP. Medical imaging systems. In: Webster, editor *Medical Instrumentation: Application and Design*. New York: John Wiley & Sons; 1998, Chapter 12.
- Singer JR, Crooks LE. Nuclear magnetic resonance blood flow measurements in the human brain. *Science* 1983;221:654–656.
- Stock DE, editor Thermal anemometry 1993. Proceedings of 3rd Int. Symposium on Thermal Anemometry/ASME Fluids Engineering Conference; Washington, DC, FED Vol. 167, 1993.
- Suzuki J, Caputo GR, Kondo C, Higgins CB, Cine MR imaging of valvular heart disease: display and imaging parameters affect the size of the signal void caused by valvular regurgitation. *American Journal of Roentgenology* 1990;155:723–727.
- Tedgui A, Lever MJ. Filtration through damaged and undamaged rabbit thoracic aorta. *American Journal of Physiology* 1984;247:784.
- Tomonaga G, Mitake H, Hoki N, Kajiya F. Measurement of point velocity in the canine coronary artery by laser Doppler velocimeter with optical fiber. *Japanese Journal of Surgery* 1981;11(4):226–231.
- Walker PG, Oyre S, Pedersen EM, Houlind K, Guenet FS, Yoganathan AP. A new control volume method for calculating valvular regurgitation. *Circulation* 1995;92:579–586.
- Webster JG,. Measurement of flow and volume of blood. In: Webster, editor *Medical Instrumentation: Application and Design*. New York: John Wiley & Sons; 1998, Chapter 8.
- Weyman AE. *Principles and Practice of Echocardiography*, 2nd ed., Philadelphia, PA: Lea & Febiger Publishers; 1994.
- White KC, Kavanaugh JF, Wang DM, Tarbell JM. Hemodynamics and wall shear rate in the abdominal aorta of dogs. *Circulation Research* 1994;75(4):637–649.
- White FM. *Fluid Mechanics*. New York: McGraw-Hill Publishers; 1979.
- Zhang H, Halliburton SS, White RD, Chatzimavroudis GP. Fast measurements of flow through mitral regurgitant orifices with magnetic resonance phase velocity mapping. *Annals of Biomedical Engineering* 2004;32(12):1618–1627.
- Zhang H, Halliburton SS, Moore JR, Simonetti OP, et al. Ultrafast flow quantification with segmented k-space magnetic resonance phase velocity mapping. *Annals of Biomedical Engineering* 2002;30:120–128.

PART III

INDUSTRIAL ENGINEERING

24

STATISTICAL QUALITY CONTROL

MAGD E. ZOHDI

- 24.1 Measurements and quality control
 - 24.2 Dimension and tolerance
 - 24.3 Quality control
 - 24.3.1 \bar{X} , R , and σ charts
 - 24.4 Interrelationship of tolerances of assembled products
 - 24.5 Operation characteristic (OC) curve
 - 24.6 Control charts for attributes
 - 24.6.1 The p and np charts
 - 24.6.2 The c and u charts
 - 24.7 Acceptance sampling
 - 24.7.1 Double sampling
 - 24.7.2 Multiple and sequential sampling
 - 24.8 Defense department acceptance sampling by variables
- Further readings

24.1 MEASUREMENTS AND QUALITY CONTROL

The metric and English measuring systems are the two measuring systems commonly used throughout the world. The metric system is universally used in most scientific applications but, for manufacturing in the United States, has been limited to a few specialties, mostly items that are related in some way to products manufactured abroad.

24.2 DIMENSION AND TOLERANCE

In dimensioning a drawing, the numbers placed in the dimension lines are only approximate and do not represent any degree of accuracy unless so stated by the designer.

To specify the degree of accuracy, it is necessary to add tolerance figures to the dimension. Tolerance is the amount of variation permitted in the part or the total variation allowed in a given dimension.

Dimensions given close tolerances mean that the part must fit properly with some other part. Both must be given tolerances in keeping with the allowance desired, the manufacturing processes available, and the minimum cost of production and assembly that will maximize profit. In general, the cost of a part is increased as the tolerance is decreased.

Allowance, which is sometimes confused with tolerance, has an altogether different meaning. It is the minimum clearance space intended between mating parts and represents the condition of tightest permissible fit.

24.3 QUALITY CONTROL

When parts must be inspected in large numbers, 100% inspection of each part is not only slow and costly but does not eliminate all of the defective pieces. Mass inspection tends to be careless; operators become fatigued; and inspection gages become worn out or out of adjustment more frequently. The risk of passing defective parts is variable and of unknown magnitude, whereas, in a planned sampling procedure, the risk can be calculated. Many products, such as bulbs, cannot be 100% inspected, since any final test made on one results in the destruction of the product. Inspection is costly, and nothing is added to a product that has been produced to specifications.

Quality control enables an inspector to sample the parts being produced in a mathematical manner and to determine whether the entire stream of production is acceptable, provided that the company is willing to allow up to a certain known number of defective parts. This number of acceptable defectives is usually taken as 3 out of 1000 parts produced. Other values might be used.

24.3.1 \bar{X} , R , and σ Charts

To use quality techniques in inspection, the following steps must be taken (Table 24.1):

- 1. sample the stream of products by taking m samples, each of size n ;
- 2. measure the desired dimension in the sample, mainly the central tendency;
- 3. calculate the deviations of the dimensions;
- 4. construct a control chart;
- 5. plot succeeding data on the control chart.

TABLE 24.1 Computational Format for Determining \bar{X} , R , and σ

Sample Number	Sample Values	Mean \bar{X}	Range R	Standard Deviation σ'
1	$X_{11}, X_{12}, \dots, X_{1n}$	\bar{X}_1	R_1	σ'_1
2	$X_{21}, X_{22}, \dots, X_{2n}$	\bar{X}_2	R_2	σ'_2
.	\dots	.	.	.
.	\dots	.	.	.
.	\dots	.	.	.
m	$X_{m1}, X_{m2}, \dots, X_{mn}$	\bar{X}_m	R_m	σ'_m

The arithmetic mean of the set of n units is the main measure of central tendency. The symbol \bar{X} is used to designate the arithmetic mean of the sample and may be expressed in algebraic terms as

$$\bar{X}_i = (X_1 + X_2 + X_3 + \cdots + X_n)/n \quad (24.1)$$

where X_1, X_2, X_3 , and so on represent the specific dimensions in question. The most useful measure of dispersion of a set of numbers is the standard deviation σ . It is defined as the root-mean-square deviation of the observed numbers from their arithmetic mean. The standard deviation σ is expressed in algebraic terms as

$$\sigma_i = \sqrt{\frac{(X_1 - \bar{X})^2 + (X_2 - \bar{X})^2 + \cdots + (X_n - \bar{X})^2}{n}} \quad (24.2)$$

Another important measure of dispersion, used particularly in control charts, is the range R . The range is the difference between the largest and the smallest observed values in a specific sample

$$R = X_i(\max) - X_i(\min) \quad (24.3)$$

Even though the distribution of the X values in the universe can be of any shape, the distribution of the \bar{X} values tends to be close to the normal distribution. The larger the sample size and the more nearly normal the universe is, the closer will the frequency distribution of the average \bar{X} 's approach the normal curve, as in Figure 24.1.

According to the statistical theory (the Central Limit Theory), in the long run, the average of the \bar{X} values will be the same as μ , the average of the universe. And in the long run, the standard deviation of the frequency distribution \bar{X} values, $\sigma_{\bar{X}}$, will be given by

$$\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}} \quad (24.4)$$

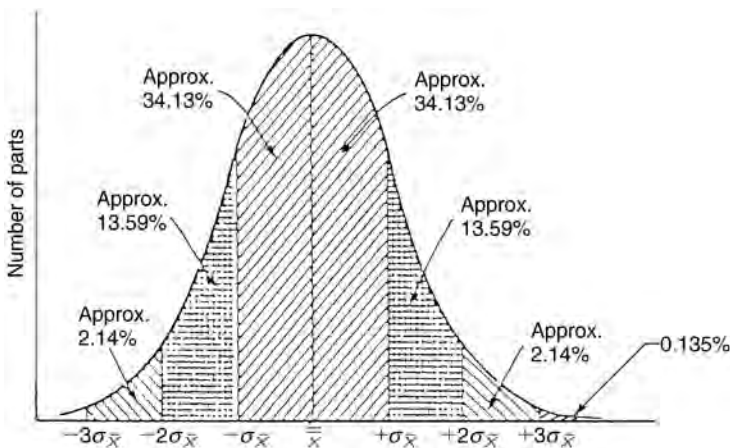


FIGURE 24.1 Normal distribution and percentage of parts that will fall within σ limits.

where σ is the standard deviation of the universe. To construct the control limits, the following steps are taken:

1. calculate the average of the average $\bar{\bar{X}}$ as follows

$$\bar{\bar{X}} = \sum_1^m \bar{X}_i / m \quad i = 1, 2, \dots, m \quad (24.5)$$

2. calculate the average deviation, $\bar{\sigma}$, where

$$\bar{\sigma} = \sum_1^m \sigma'_i / m \quad i = 1, 2, \dots, m \quad (24.6)$$

Statistical theory predicts the relationship between $\bar{\sigma}$ and $\sigma_{\bar{X}}$. The relationship for the $3\sigma_{\bar{X}}$ limits or the 99.73% limits is

$$A_1 \bar{\sigma} = 3\sigma_{\bar{X}} \quad (24.7)$$

This means that control limits are set so that only 0.27% of the produced units will fall outside the limits. The value of $3\sigma_{\bar{X}}$ is an arbitrary limit that has found acceptance in industry.

The value of A_1 calculated by probability theory is dependent on the sample size and is given in Table 24.2. The formula for 3σ control limits using this factor is

$$CL(\bar{X}) = \bar{\bar{X}} \pm A_1 \bar{\sigma} \quad (24.8)$$

Once the control chart (Figure 24.2) has been established, data (\bar{X}_i 's) that result from samples of the same size n are recorded on it. It becomes a record of the

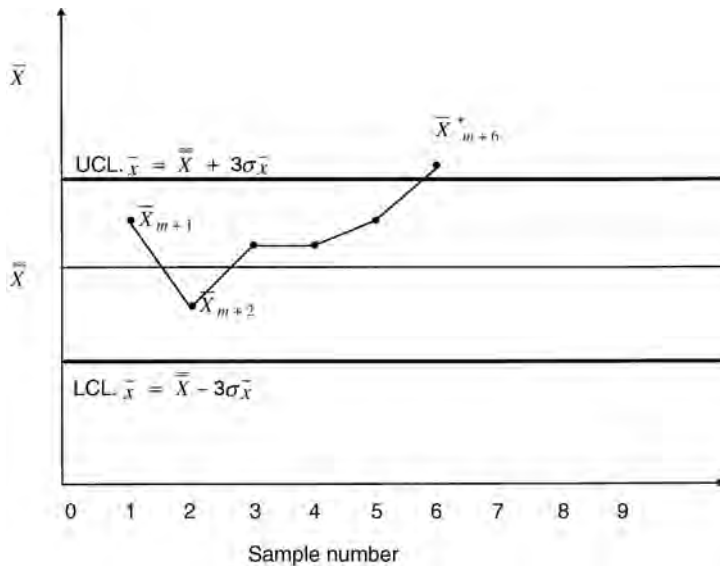


FIGURE 24.2 Control chart \bar{X} .

TABLE 24.2 Factors for \bar{X} , R , σ , and X Control Charts

Sample Size n	Factors for \bar{X} Chart		Factors for R Chart		Factors for σ' Chart		Factors for X Chart		$\sigma = \bar{R}/d_2$ d_2
	From $\bar{R} A_2$	From $\bar{\sigma} A_1$	Lower D_3	Upper D_4	Lower B_3	Upper B_4	From $\bar{R} E_2$	From $\bar{\sigma} E_1$	
2	1.880	3.759	0	3.268	0	3.267	2.660	5.318	1.128
3	1.023	2.394	0	2.574	0	2.568	1.772	4.146	1.693
4	0.729	1.880	0	2.282	0	2.266	1.457	3.760	2.059
5	0.577	1.596	0	2.114	0	2.089	1.290	3.568	2.326
6	0.483	1.410	0	2.004	0.030	1.970	1.184	3.454	2.539
7	0.419	1.277	0.076	1.924	0.118	1.882	1.109	3.378	2.704
8	0.373	1.175	0.136	1.864	0.185	1.815	1.054	3.323	2.847
9	0.337	1.094	0.184	1.816	0.239	1.761	1.011	3.283	2.970
10	0.308	1.028	0.223	1.777	0.284	1.716	0.975	3.251	3.078
11	0.285	0.973	0.256	1.744	0.321	1.679	0.946	3.226	3.173
12	0.266	0.925	0.284	1.717	0.354	1.646	0.921	3.205	3.258
13	0.249	0.884	0.308	1.692	0.382	1.618	0.899	3.188	3.336
14	0.235	0.848	0.329	1.671	0.406	1.594	0.881	3.174	3.407
15	0.223	0.817	0.348	1.652	0.428	1.572	0.864	3.161	3.472
16	0.212	0.788	0.364	1.636	0.448	1.552	0.848	3.152	3.532
17	0.203	0.762	0.380	1.621	0.466	1.534	0.830	3.145	3.588
18	0.194	0.738	0.393	1.608	0.482	1.518	0.820	3.137	3.640
19	0.187	0.717	0.404	1.597	0.497	1.503	0.810	3.130	3.687
20	0.180	0.698	0.414	1.586	0.510	1.490	0.805	3.122	3.735
21	0.173	0.680	0.425	1.575	0.523	1.477	0.792	3.114	3.778
22	0.167	0.662	0.434	1.566	0.534	1.466	0.783	3.105	3.819
23	0.162	0.647	0.443	1.557	0.545	1.455	0.776	3.099	3.858
24	0.157	0.632	0.451	1.548	0.555	1.445	0.769	3.096	3.895
25	0.153	0.619	0.459	1.540	0.565	1.435	0.765	3.095	3.931

variation of the inspected dimensions over a period of time. The data plotted should fall in random fashion between the control limits 99.73% of the time if a stable pattern of variation exists.

So long as the points fall between the control lines, no adjustments or changes in the process are necessary. If five to seven consecutive points fall on one side of the mean, the process should be checked. When points fall outside of the control lines, the reason must be located and corrected immediately.

Statistical theory also gives the expected relationship between $\bar{R}(\Sigma R_i/m)$ and $\sigma_{\bar{X}}$. The relationship for the $3\sigma_{\bar{X}}$ limits is

$$A_2 \bar{R} = 3\sigma_{\bar{X}} \quad (24.9)$$

The values for A_2 calculated by probability theory, for different sample sizes, are given in Table 24.2.

The formula for 3σ control limits using this factor is

$$CL(\bar{X}) = \bar{\bar{X}} \pm A_2 \bar{R} \quad (24.10)$$

In control chart work, the ease of calculating R is usually much more important than any slight theoretical advantage that might come from the use of σ . However, in some cases where the measurements are costly, and it is necessary that the inferences from a limited number of tests be as reliable as possible, the extra cost of calculating σ is justified. It should be noted that, because Figure 24.2 shows the averages rather than individual values, it would have been misleading to indicate the tolerance limits on this chart. It is the individual article that has to meet the tolerances, not the average of a sample. Tolerance limits should be compared with the machine capability limits. Capability limits are the limits on a single unit and can be calculated by

$$\begin{aligned}\text{Capability limits} &= \bar{\bar{X}} \pm 3\sigma \\ \sigma &= \bar{R}/d_2\end{aligned}\tag{24.11}$$

Since $\sigma' = \sqrt{n} \sigma_{\bar{x}}$, the capability limits can be given by

$$\text{Capability limits } (X) = \bar{\bar{X}} \pm 3\sqrt{n} \sigma_{\bar{x}}\tag{24.12}$$

$$= \bar{\bar{X}} \pm E_1 \bar{\sigma}\tag{24.13}$$

$$= \bar{\bar{X}} \pm E_2 \bar{R}\tag{24.14}$$

The values for d_2 , E_1 , and E_2 calculated by probability theory, for different sample sizes, are given in Table 24.2.

Figure 24.3 shows the relationships among the control limits, the capability limits, and assumed tolerance limits for a machine that is capable of producing the product with this specified tolerance. Capability limits indicate that the production facility can produce 99.73% of its products within these limits. If the specified tolerance limits are greater than the capability limits, the production facility is capable of meeting the production requirement. If the specified tolerance limits are tighter than the capability limits, a certain percentage of the production will not be usable and 100% inspection will be required to detect the products outside the tolerance limits.

To detect changes in the dispersion of the process, the R and σ charts are often employed with \bar{X} and X charts.

The upper and lower control limits for the R chart are specified as

$$\text{UCL } (R) = D_4 \bar{R}\tag{24.15}$$

$$\text{LCL } (R) = D_3 \bar{R}\tag{24.16}$$

Figure 24.4 shows the \bar{R} chart for samples of size 5.

The upper and lower control for the T chart are specified as

$$\text{UCL } (\sigma) = B_4 \bar{\sigma}\tag{24.17}$$

$$\text{LCL } (\sigma) = B_3 \bar{\sigma}\tag{24.18}$$

The values for D_3 , D_4 , B_3 , and B_4 calculated by probability theory, for different sample sizes, are given in Table 24.2.

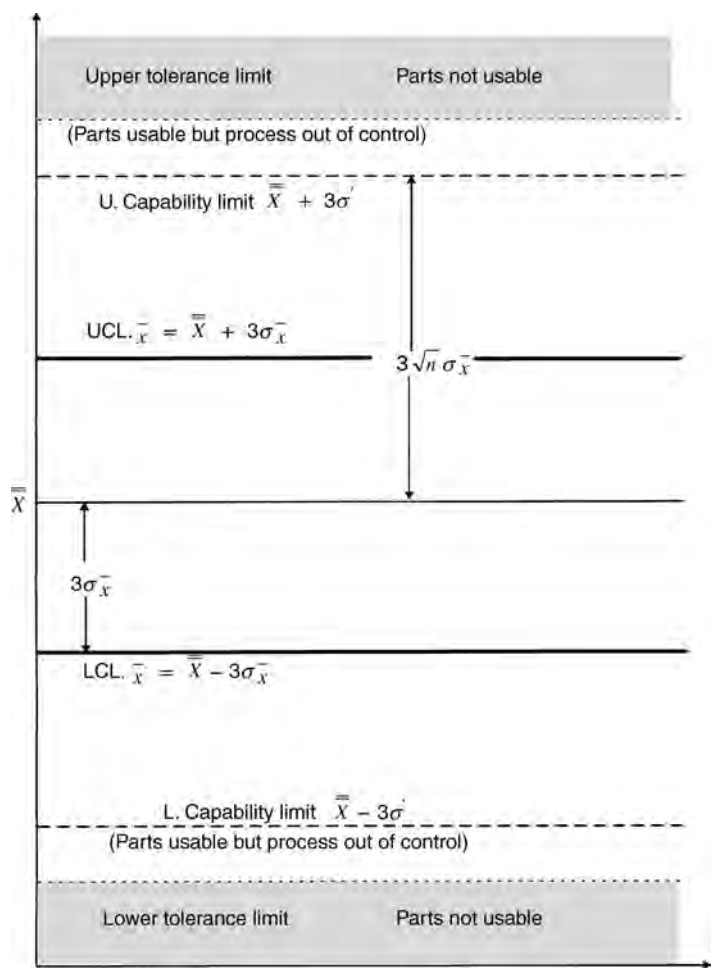


FIGURE 24.3 Control, capability, and tolerance (specification limits).

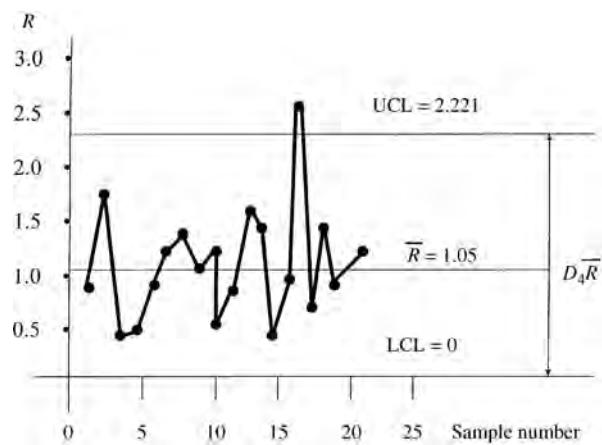


FIGURE 24.4 R Chart for samples of five each.

24.4 INTERRELATIONSHIP OF TOLERANCES OF ASSEMBLED PRODUCTS

Mathematical statistics states that the dimension on an assembled product may be the sum of the dimensions of the several parts that make up the product. It states also that the standard deviation of the sum of any number of independent variables is the square root of the sum of the squares of the standard deviations of the independent variables. So if

$$X = X_1 \pm X_2 \pm \cdots \pm X_n \quad (24.19)$$

$$\bar{X} = \bar{X}_1 \pm \bar{X}_2 \pm \cdots \pm \bar{X}_n \quad (24.20)$$

$$\sigma(X) = \sqrt{(\sigma_1)^2 + (\sigma_2)^2 + \cdots + (\sigma_n)^2} \quad (24.21)$$

Whenever it is reasonable to assume that the tolerance ranges of the parts are proportional to their respective σ' values, such tolerance ranges may be combined by taking the square root of the sum of the squares

$$T = \sqrt{T_1^2 + T_2^2 + T_3^2 + \cdots + T_n^2} \quad (24.22)$$

24.5 OPERATION CHARACTERISTIC (OC) CURVE

Control charts detect changes in a pattern of variation. If the chart indicates that a change has occurred when it has not, Type I error occurs. If three-sigma limits are used, the probability of making a Type I error is approximately 0.0027.

The probability of the chart indicating no change, when in fact it has, is the probability of making a Type II error. The operation characteristic curves are designed to indicate the probability of making a Type II error. An OC curve for an \bar{X} chart of three-sigma limits is illustrated in Figure 24.5.

24.6 CONTROL CHARTS FOR ATTRIBUTES

Testing may yield only one of two defined classes: within or outside certain limits, acceptable or defective, working or idle. In such a classification system, the proportion of units falling in one class may be monitored with a p chart.

In other cases, observation may yield a multivalued, but still discrete, classification system. In such case, the number of discrete observations, such as events, objects, states, or occurrences, may be monitored by a c chart.

24.6.1 The p and np Charts

When sampled items are tested and placed into one of two defined classes, the proportion of units falling into one class p is described by the binomial distribution. The mean and standard deviation are given as

$$\begin{aligned} \mu &= np \\ \sigma &= \sqrt{np(1-p)} \end{aligned}$$

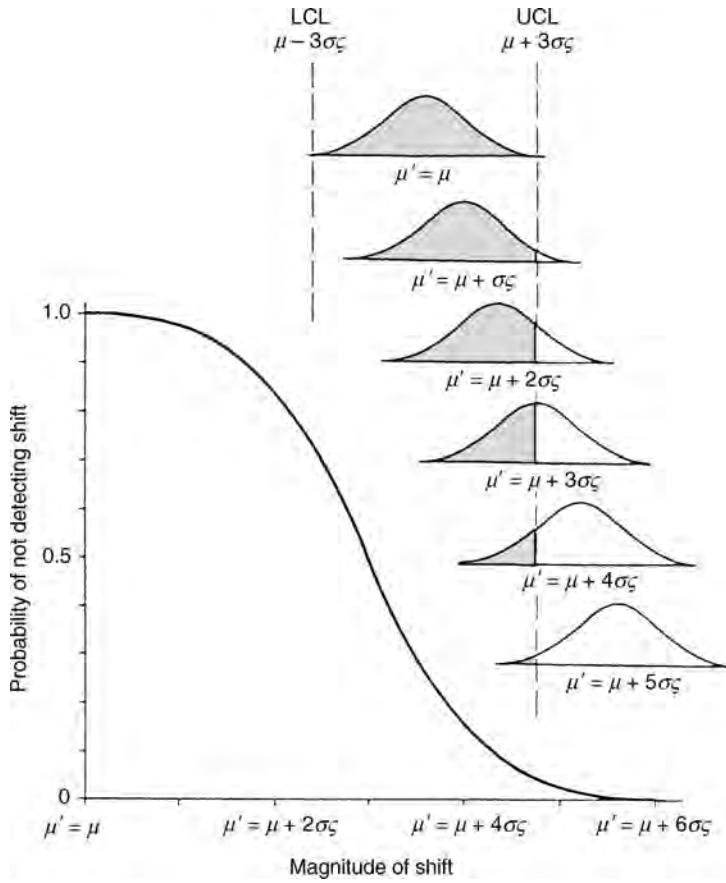


FIGURE 24.5 Operating characteristic curve for 3σ limit.

Dividing by the sample size n , the parameters are expressing as proportions. These statistics can be expressed as

$$\bar{p} = \frac{\text{Total number in the class}}{\text{Total number of observations}} \quad (24.23)$$

$$s_p = \sqrt{\frac{\bar{p}(1 - \bar{p})}{n}} \quad (24.24)$$

The control limits are set either at two-sigma limits with Type I error as 0.0456 or at three-sigma limits with Type I error as 0.0027. The control limits for the p chart with two-sigma limits (Figure 24.6) are defined as

$$CL(p) = \bar{p} \pm 2S_p \quad (24.25)$$

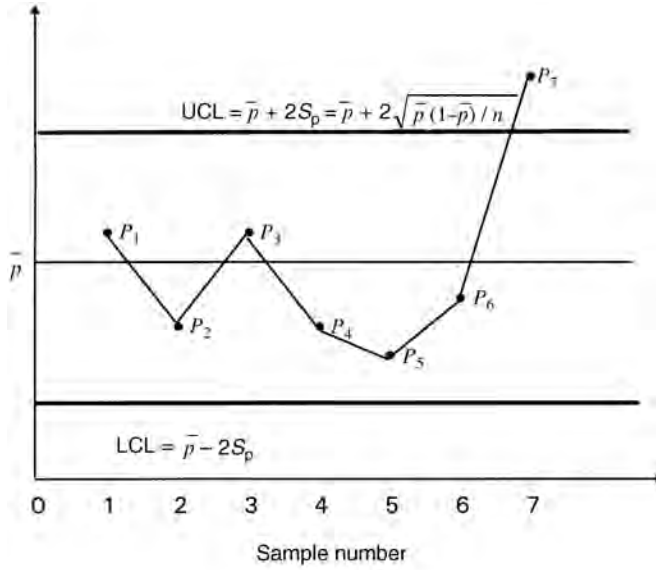


FIGURE 24.6 P charts.

However, if subgroup size is constant, the chart for actual numbers of rejects np or pn may be used. The appropriate model for three-sigma control limits on an np chart is

$$CL(np) = n\bar{p} \pm 3\sqrt{n\bar{p}(1-\bar{p})} \quad (24.26)$$

24.6.2 The c and u Charts

The random variable process that provides numerical data that are recorded as a number c rather than a proportion p is described by the Poisson distribution. The mean and the variance of the Poisson distribution are equal and expressed as $\mu = \sigma^2 = np$. The Poisson distribution is applicable in any situation when n and p cannot be determined separately, but their product np can be established. The mean and variable can be estimated as

$$\bar{c} = S_c^2 = \frac{\sum_l^m C_i}{m} = \frac{\sum_1^m (np)_i}{m} \quad (24.27)$$

The control limits (Figure 24.7) are defined as

$$CL(c) = \bar{C} \pm 3S_c \quad (24.28)$$

If there is change in the area of opportunity for occurrence of a nonconformity from subgroup to subgroup, such as number of units inspected or the lengths of wires checked, the conventional c chart showing only the total number of nonconformities is not applicable. To create some standard measure of the area of opportunity, the

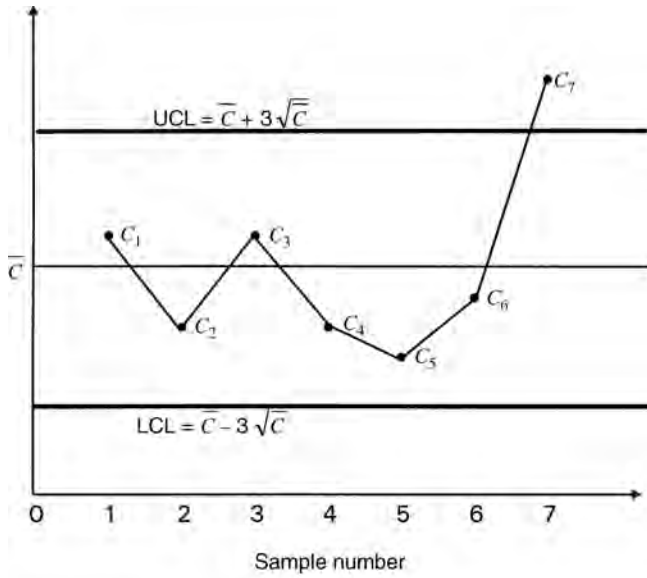


FIGURE 24.7 C charts.

nonconformities per unit (c/n) or u is used as the control statistic. The control limits are

$$CL(u)\bar{u} \pm 3 \frac{\sqrt{\bar{u}}}{\sqrt{n_i}} \quad (24.29)$$

where

$$\bar{u} = \frac{\sum C_i}{\sum n_i} = \frac{\text{Total nonconformities found}}{\text{Total units inspected}}$$

$c = nu$ is Poisson-distributed, u is not

24.7 ACCEPTANCE SAMPLING

The objective of acceptance sampling is to determine whether a quantity of the output of a process is acceptable according to some criterion of quality. A sample from the lot is inspected, and the lot is accepted or rejected in accordance with the findings of the sample.

Acceptance sampling plans call for the random selection of sample of size n from a lot containing N items. The lot is accepted if the number of defectives found in the sample are $\leq c$, the acceptance number. A rejected lot can either be returned to the producer, non-rectifying inspections, or it can be retained and subjected to a 100% screening process, rectifying inspection plan improves the outgoing quality. A second attribute-inspection

plan might use two samples before requiring the acceptance or rejection of a lot. A third plan might use multiple samples or a sequential sampling process in evaluating a lot. Under rectifying inspection programs, the average outgoing quality level (AOQ), the average inspection lot (I), and the average outgoing quality limit (AOQL) can be predicted for varying levels of incoming fraction defective p .

Assuming that all lots arriving contain the same proportion of defectives p , and that rejected lots will be subjected to 100% inspection, AOQ and I are given below:

$$\text{AOQ} = \frac{P_a p (N - n)}{N - pn - (1 - P_a)p(Nn)} \quad (24.30)$$

$$I = n + (1 - P_a)(N - n) \quad (24.31)$$

The average outgoing quality (AOQ) increases as the proportion defective in the incoming lots increases until it reaches a maximum value and then starts to decrease. This maximum value is referred to as the average outgoing quality limit. The hypergeometric distribution is the appropriate distribution to calculate the probability of acceptance P_a ; however, the Poisson distribution is used as an approximation.

Nonrectifying inspection program does not significantly improve the quality level of the lots inspected.

24.7.1 Double Sampling

Double sampling involves the possibility of putting off the decision on the lot until a second sample has been taken. A lot may be accepted at once if the first sample is good enough or rejected at once if the first sample is bad enough. If the first sample is neither, the decision is based on the evidence of the first and second samples combined.

The symbols used in double sampling are

N = lot size

n_1 = first sample

c_1 = acceptance number for first sample

n_2 = second sample

c_2 = acceptance number of the two samples combined.

Computer programs are used to calculate the OC curves: acceptance after the first sample, rejection after the first sample, acceptance after the second sample, and rejection after the second sample. The average sample number (ASN) in double sampling is given by

$$\text{ASN} = [P_a(n_1) + P_r(n_1)]n_1 + [P_a(n_2) + P_r(n_2)](n_1 + n_2) \quad (24.32)$$

24.7.2 Multiple and Sequential Sampling

In multiple sampling, three or more samples of a stated size are permitted, and the decision on acceptance or rejection is revealed after a stated number of samples. In sequential sampling, item-by-item inspection, a decision is possible after each item has been inspected, and when there is no specified limit on the total number of units to be

inspected. OC curves are developed through computer programs. The advantage of using double sampling, multiple sampling, or sequential sampling is to reach the appropriate decision with fewer items inspected.

24.8 DEFENSE DEPARTMENT ACCEPTANCE SAMPLING BY VARIABLES

MIL-STD-105 A, B, C, D, and then ABC-STD-105 are based on the acceptance quality level (AQL) concept. The plans contain single, double, or multiple sampling depending on the lot size and AQL and the probability of acceptance at this level P_a . Criteria for shifting to tightened inspection, requalification for normal inspection, and reduced inspection are listed in the tables associated with plan.

MIL-STD-414 plans were developed to reduce inspection lots by using sample sizes compared with MIL-STD-105. They are similar, as both procedures and tables are based on the concept of AQL; lot-by-lot acceptance inspection; both provide for normal, tightened, or reduced inspection; sample sizes are greatly influenced by lot size; several inspection levels are available; and all plans are identified by sample size code letter. MIL-STD-414 could be applied either with a single specification limit, L or U , or with two specification limits. Known-sigma plans included in the standard were designated as having "variability known." Unknown-sigma plans were designated as having "variability unknown." In the latter-type plans, it was possible to use either the standard deviation method or the range method in estimating the lot variability.

FURTHER READINGS

- Aft LS. *Fundamentals of Industrial Quality Control*. 3rd ed. Menlo Park (CA): Addison-Wesley; 1988.
- ASTM Manual on Presentation of Data and Control Chart Analysis*. Special Technical Pub. 15D, American Society for Testing and Materials, Philadelphia, PA; 1976.
- Bhushan B, editor. *Modern Tribology Handbook*. Boca Raton (FL): CRC Press; 2001.
- Clements R. *Quality ITQM/ISO 9000*. Englewood Cliffs (NJ): Prentice-Hall; 1995.
- Control Chart Method of Controlling Quality During Production, ANSI Standard 21.3-1975*, American National Standards Institute, New York; 1975.
- Devore JL. *Probability and Statistics for Engineering and the Sciences*. New York: Duxburg Press; 1995.
- Dodge HF. *A General Procedure for Sampling Inspection by Attributes—Based on the AQL Concept*. Technical Report No. 10. New Brunswick (NJ): The Statistics Center, Rutgers—The State University; 1959.
- Drake PJ. *Dimension and Tolerancing Handbook*. New York: McGraw-Hill; 1999.
- Duncan AJ. *Quality Control and Industrial Statistics*. 5th ed. Homewood (IL): Richard D. Irwin; 1986.
- Feigenbaum AV. *Total Quality Control—Engineering and Management*. 3rd ed. New York: McGraw-Hill; 1991.
- Grant EL, Leavenworth RS. *Statistical Quality Control*. 7th ed. New York: McGraw-Hill; 1996. (Software included).
- Juran JM, Gryna, FM, Jr. *Quality Control Handbook*. New York: McGraw-Hill; 1988.

- Juran JM, Gryna, Jr. FM. *Quality Planning and Analysis*. 4th ed. New York: McGraw-Hill; 2000.
- Lamprecht J. *Implementing the ISO 9000 Series*. New York: Marcel Dekker; 1995.
- Military Standard 105E, Sampling Procedures and Tables for Inspection by Attributes*, Superintendent of Documents, Government Printing Office, Washington, DC; 1969.
- Military Standard 414, Sampling Procedures and Tables for Inspection by Variables for Percent Defective*, Superintendent of Documents, Government Printing Office, Washington, DC; 1957.
- Military Standard 690-B, Failure Rate Sampling Plans and Procedures*, Superintendent of Documents, Government Printing Office, Washington, DC; 1969.
- Military Standard 781-C, Reliability Design Qualification and Production Acceptance Tests. Exponential Distribution*, Superintendent of Documents, Government Printing Office, Washington, DC; 1977.
- Military Standard 1235B, Single and Multi-Level Continuous Sampling Procedures and Tables for Inspection by Attributes*, Superintendent of Documents, Government Printing Office, Washington, DC; 1981.
- Montgomery DC, Runger GC. *Introduction to Statistical Quality Control*. New York: Wiley; 2001.
- Rabbit J, Bergh P. *The ISO 9000 Book*. Dearborn (MI): ME; 1993.
- Society of Manufacturing Engineers. *Quality Control and Assembly*. 4th ed. Dearborn (MI): Society of Manufacturing Engineers; 1994.
- Supply and Logistic Handbook—Inspection H 105. Administration of Sampling Procedures for Acceptance Inspection*, Superintendent of Documents, Government Printing Office, Washington, DC; 1954.
- Walpole RE, Myers RH, Myers SL, Ye K. *Probability and Statistics for Engineers and Scientists*. Englewood Cliffs (NJ): Prentice Hall; 2002.
- Zohdi ME. *Manufacturing Processes Quality Evaluation and Testing*, International Conference, Operations Research, January; 1976.

25

EVALUATING AND SELECTING TECHNOLOGY-BASED PROJECTS

HANS J. THAMHAIN

- 25.1 Management perspective
- 25.2 Quantitative approaches
 - 25.2.1 Net present value (NPV) comparison
 - 25.2.2 Return on investment (ROI) comparison
 - 25.2.3 Cost–Benefit (CB)
 - 25.2.4 Payback period (PBP) comparison
 - 25.2.5 Pacifico and Sobelman project ratings
 - 25.2.6 Going beyond simple formulas
- 25.3 Qualitative approaches
 - 25.3.1 Collective multifunctional evaluations
- 25.4 Recommendations
- Variables and Abbreviations
- References

Predicting project success is difficult and often unreliable (Prabhakar, 2009; Raz et al., 2002; Shenhar et al., 2002; Thamhain and Skelton, 2007). The long list of prominent project failures, ranging from computers to pharmaceutical and supersonic transport, reminds us of this reality (Cicmil et al., 2006; Lemon et al., 2002). Many projects do not live up to their expectations or outright fail even before their technical completion (El Emam & Koru, 2008). The ability to evaluate project proposals, assessing future success and organizational value, is critical to overall business performance for most enterprises.

25.1 MANAGEMENT PERSPECTIVE

Project success is multifaceted. Typically, it includes not only technical but also financial; marketing; and social, legal, and ethical dimensions. For most projects, the DNA of

success is highly complex, and outcomes are difficult to predict (Shenhar & Dvir, 2007). Evaluating and selecting projects is both an art and a science that, for most cases, has to go beyond a simple cost–benefit analysis. To be sure, few decisions have more impact on business performance than the resource allocations for new projects. Virtually, every organization selects and implements projects, ranging from product developments to organizational improvements, and from customer contracts to R&D activities and bid proposals. Pursuing the “wrong” project not only wastes company resources but also causes the enterprise to (i) miss critical alternatives, (ii) operate less flexible and responsive in the market place, and (iii) miss opportunities for leveraging core competencies. Project opportunities must be analyzed relative to their potential value, strength, and importance to the enterprise. *Four major dimensions* should be considered: (1) added value of the new project, (2) cost of the project, (3) readiness of the enterprise to execute the project, and (4) managerial desire. A well-organized *project evaluation and selection process* provides the framework for systematic data gathering and informed decision-making toward resource allocation (Bstieler, 2005; Ciemil & Marshall, 2005). Typically, these decisions can be broken into four principal categories:

1. *Deciding Initial Feasibility*: screening and filtering, quick decision on the viability of an emerging project for further evaluation;
2. *Deciding Strategic Value to Enterprise*: identifying alternatives and options to proposed project;
3. *Deciding Detailed Feasibility*: determining the chances of success for a proposed project;
4. *Deciding Project Go/No-Go*: committing resources for a project implementation.

While making these decisions looks simple, logical, and straightforward, developing meaningful support data is a complex undertaking. It is also expensive, time-consuming, and often highly eclectic. Typically, decision-making requires the following inputs:

1. specific resource requirements;
2. specific implementation risks;
3. specific benefits (economics, technology, markets, etc.);
4. benchmarking and comparative analysis;
5. strategic perspective, including long- and short-term value assessment.

Although it is challenging to estimate costs, schedules, risks, and benefits, such as those shown in Table 25.1, these measures are relatively straightforward in comparison to *predicting project success*. The difficulty is in defining a meaningful *aggregate indicator for project value and success*. Methods for determining success range from purely intuitive to highly analytical. No method is seen as truly reliable in predicting success, especially for more complex and technologically intensive types of projects. Yet, some companies have a better track record in selecting “winning” projects than others. They seem to have the ability to create a more integrated picture of the potential benefits, costs, and risks for the proposed project relative to the company’s strength and strategic objectives. Producing such a composite is both a science and an art. Traditionally, the management literature suggested by-and-large *rational selection processes* to support project selections (Remer, 1993). However, purely rational-analytical processes apply only to a limited number of

TABLE 25.1 Typical Criteria and Measures used for Project Evaluation

The criteria relevant to the evaluation and selection of a particular project depend on the specific project type and business situation such as for a particular product development, custom project, process development, industry, or market. Typically, evaluation procedures include the following criteria and measures

- Development cost
- Development time
- Technical complexity and feasibility
- Risk
- Return on investment
- Cost–benefit
- Product life cycle
- Sales volume
- Market share
- Project business follow-on
- Organizational readiness and strength
- Consistency with business plan
- Resource availability
- Cash flow, revenue, and profit
- Impact on other business activities

Note: Each criteria is based on a complex set of parameters and variables.

business situations. Many of today's technologically complex business scenarios require the integration of both analytical and judgmental techniques to evaluate projects in a meaningful way (Kavadias and Loch, 2004; Khorramshahgol et al., 1998), predicting success and making the *best choice*.

Yet, in spite of the dynamics involved in the selection process, systematic information gathering and standardized methods are at the heart of any project evaluation process, and provide the best assurance for reliably predicting project outcome and repeatability of the decision process. Approaches to project evaluation and selection fall into one of three principal classes:

1. primarily *quantitative* and *rational* approaches;
2. primarily *qualitative* and *intuitive* approaches;
3. *mixed approaches*, combining both quantitative and qualitative methods.

Because of the interdisciplinary complexities involved, analyzing a new project opportunity is a highly interactive effort among the various resource groups of the enterprise and its partners (Thamhain, 2005). Often, many meetings are needed before (i) a clear picture emerges of potential benefits, costs, and risks involved in the project, and (ii) data emerge that is useful for the project evaluation and selection process, regardless of its quantitative, qualitative, or combined nature.

25.2 QUANTITATIVE APPROACHES

Quantitative approaches are often favored to support project evaluation and selections if the decisions require economic justification. They are also commonly used to support judgment-based project selections. One of the features of quantitative approaches

is the generation of numeric measures for simple and effective comparison, ranking, and selection (Mantel et al., 2011; Remer et al., 1993). These approaches also help to establish quantifiable norms and standards, and lead to repeatable processes. Yet, the ultimate usefulness of these methods depends on the assumption that the decision parameters, such as cash flow; risks; and the underlying economic, social, political, and market factors can actually be quantified and reliably estimated over the project life cycle. Therefore, quantitative techniques are effective and powerful decision support tools, if meaningful estimates of cost-benefits, such as capital expenditures and future revenues, can be obtained, and converted into net present values for comparison. Because of their importance, quantitative methods have been discussed in the literature extensively, ranging from simple return on investment (ROI) calculations to elaborate simulations of project scenarios. Many companies eventually develop their own project evaluation/selection models, customized to their specific needs. However, the backbone for most of these customized models is a set of economic/financial measures that tries to determine the cost–benefit of the proposed venture, usually for some point in the future. Specifically, four measures are especially popular:

1. net present value (NPV),
2. return on investment (ROI),
3. cost–benefit (CB),
4. payback period (PBP).

The calculation and application of these measures to project evaluation/selection will be illustrated by case examples. Specifically, four project proposals (described in Table 25.2) are evaluated in this chapter, using the above measures. The results are summarized in Table 25.3.

25.2.1 Net Present Value (NPV) Comparison

This method uses discounted cash flow as the basis for comparing the relative merit of alternative project opportunities. It assumes that all investment costs and revenues are known and that economic analysis is a valid basis for project selection.

We can determine the *Net Present Value (NPV)* of a single revenue, or stream of future revenues, or costs expected in the future. Two types of presentations are common: (1) present worth and (2) net present value.

TABLE 25.2 Description of Four Project Proposals

<i>Project option P1:</i> Management does not accept any new project proposal. Hence, no investment capital is required, nor is any revenue generated
<i>Project option P2:</i> This opportunity requires a \$1000 investment at the beginning of the first year and generates a \$200 revenue at the end of <i>each</i> of the following 5 years
<i>Project option P3:</i> This opportunity requires a \$2000 investment at the beginning of the first year and generates a variable stream of net revenues at the end of <i>each</i> of the next 5 years as follows \$1500, \$1000, \$800, \$900, and \$1200
<i>Project option P4:</i> This opportunity requires a \$5000 investment at the beginning of the first year and generates a variable stream of net revenues at the end of <i>each</i> of the next 5 years as follows \$1000, \$1500, \$2000, \$3000, and \$4000

TABLE 25.3 Cash Flow and Net Value Calculations of Four Project Options or Proposals, Assuming a MARR of $i = 10\%$

End of Year N	Do-Nothing Option P1	Project Option P2	Project Option P3	Project Option P4
<i>Given cash flow</i>				
0	0	−1,000	−2,000	−5,000
1	0	200	1,500	1,000
2	0	200	1,000	1,500
3	0	200	800	2,000
4	0	200	900	3,000
5	0	200	1,200	4,000
<i>Calculations</i>				
Net cash flow (ΣP)	0	0	+3,400	+6,500
Net present value at the end of year 5 ($NPV _{N=5}$)	0	−242	+2,153	+3,192
Net present value for revenue to continue to ∞ ($NPV _{N=\infty}$)	0	+1,000	+9,904	+28,030
Average annual ROI ($ROI _{N=5}$)	0	20%	54%	46%
Cost–benefit ($CB = ROI_{NPV N=5}$)	0	76%	108%	164%
Payback period for MARR = 10% ($N_{PBP} _{i=10}$)	0	8	1.8	3.8
Payback period for MARR = 0% ($N_{PBP} _{i=0}$)	0	5	1.5	3.3

Note: Given for all four project proposals: (1) a single investment is being made at the beginning of the project life cycle (e.g., at the end of year 0) and (2) the internal rate of return, IRR, or the minimum attractive rate of return, MARR, is 10%.

Present Worth (PW): This is the single revenue or cost (also called annuity A) which occurs at the end of a period n , subject to the *prevailing interest rate* i . Depending on the management philosophy and enterprise policies, this interest rate can be (i) the *internal rate of return (IRR)* realized by the company on similar investments and (ii) the *minimum attractive rate of return (MARR)* acceptable to company management, or the prevailing discount rate. The present worth is calculated as

$$PW(A|i, n) = PW_n = A \frac{1}{(1+i)^n}$$

For the examples used in this chapter, we consider the *internal rate of return, IRR* (defined as the average return realized on similar investments), to be the prevailing interest rate.

Net Present Value (NPV): The *net present value* is defined as a series of revenues or costs, A_n , over N periods of time, at a prevailing interest rate i :

$$NPV(A_n|i, N) = \sum_{n=1}^N A_n \frac{1}{(1+i)^n} = \sum_{n=1}^N PW_n$$

Three special cases exist for the net present value calculation: (1) *for a uniform series of revenues or costs over N periods*: $\text{NPV}(A_n|i, N) = A[(1+i)^N - 1]/i(1+i)^N$; (2) *for an annuity or interest rate i approaching zero*: $\text{NPV} = A \times N$; and (3) *for the revenue or cost series to continue forever*: $\text{NPV} = A/i$.

Table 25.3 applies these formulas to the four project alternatives described in Table 25.2, showing the most favorable 5-year net present value of \$3192 for project option P3.

25.2.2 Return on Investment (ROI) Comparison

Perhaps, one of the most popular measures for project evaluation is the *return on investment, ROI*:

$$\text{ROI} = \frac{\text{Revenue}(R) - \text{Cost}(C)}{\text{Investment}(I)}$$

ROI calculates the ratio of net revenue over investment. In its simplest form, the stream of cash flow is *not* discounted. One can look at the revenue on a year-by-year basis, relative to the initial investment. For example, project option 1 in Table 25.3 would produce a 20% ROI each year, whereas project option 2 would produce a 75% ROI during the first year, 50% during the second year, and so on. In a somewhat more sophisticated way, we can calculate the *average ROI per year* over a given revenue cycle as shown in Table 25.3

$$\overline{\text{ROI}}(A_n, I_n|N) = \left[\sum_{n=1}^N \frac{(\text{Revenue } R)_n - (\text{Cost } C)_n}{(\text{Investment } I)_n} \right] / [N]$$

We can then *compare the average ROI to the minimum attractive rate of return, MARR*. Given a MARR of 10% for our project environment, all three project options P1, P2, and P3 compare favorable, with project P3 yielding the highest average return on investment of 54%. Although this is a popular measure, it does not permit a meaningful comparative analysis of alternative projects with fluctuating costs and revenues. Furthermore, it does not consider the time-value of money.

25.2.3 Cost–Benefit (CB)

Alternatively, we can calculate the *net present value* of the total ROI over the project life cycle. This measure, known as *Cost–Benefit, CB*, is calculated as the present-value stream of net-revenues divided by the present-value stream of investments. It is an effective measure for comparing project alternatives with fluctuating cash flows:

$$\text{CB} = \text{ROI}_{\text{NPV}}(A_n, I_n|i, N) = \left[\sum_{n=1}^N \text{NPV}(A_n|i, N) \right] / \left[\sum_{n=1}^N \text{NPV}(I_n|i, N) \right]$$

In our example of four project options (Table 25.3), project proposal P4 produces the highest cost–benefit of 164% under the given assumption of $i = \text{MARR} = 10\%$.

25.2.4 Payback Period (PBP) Comparison

Another popular figure of merit for comparing project alternatives is the *Payback Period (PBP)*. It indicates the time period of net revenues required to return the capital investment made on the project. For simplicity, *undiscounted* cash flows are often used to calculate a quick figure for comparison, which is quite meaningful if we deal with an initial investment and a steady stream of net revenue. However, for fluctuating revenue and/or cost streams, the net present value must be *calculated for each period individually* and cumulatively added up to the “break-even point” in time, N_{PBP} , when the net present value of revenue equals the investment. Mathematically,

$$N_{\text{PBP}} \text{ occurs when } \sum_{n=1}^N \text{NPV}(A_n|i) \geq \sum_{n=1}^N \text{NPV}(I_n|i)$$

In our example of four project options (Table 25.3), project proposal P3 produces the shortest, most favorable payback period of 1.8 years under the given assumption of $i = \text{MARR} = 10\%$.

25.2.5 Pacifico and Sobelman Project Ratings

The previously discussed methods of evaluating projects rely heavily on the assumption that technical and commercial success is assured, and all costs and revenues are predictable. Because these assumptions do not always hold, many companies have developed their own special procedures and formulas for comparing project alternatives. Two examples illustrate this special category of project evaluation metrics.

The Project Rating Factor (PR): This measure was originally developed by Carl Pacifico for assessing chemical products and predicting commercial success:

$$\text{PR} = \frac{\text{pT} \times \text{pC} \times R}{\text{TC}}$$

Pacifico’s formula is in essence an ROI calculation adjusted for risk. It includes probability of technical success ($0.1 < \text{pT} < 1.0$), probability of commercial success ($0.1 < \text{pC} < 1.0$), total net revenue over project life cycle [R], and total capital investment for product development, manufacturing setup, marketing, and related overheads (TC).

Product Development Figure of Merit: The formula developed by Sobelman

$$z = (P \times T_{\text{LC}}) - (C \times T_{\text{D}})$$

represents a modified cost-benefit measure that takes into account both the development time and the commercial life cycle of the product. It also includes average profit per year (P), estimated product life cycle (T_{LC}), average development cost per year (C), and years of development (T_{D}).

25.2.6 Going Beyond Simple Formulas

While quantitative methods of project evaluation have the benefit of producing relatively quickly a measure of merit for simple comparison and ranking, they also have many

TABLE 25.4 Comparison of Quantitative and Qualitative Approaches to Project Evaluation

Quantitative Methods	Qualitative Methods
<i>Benefits</i> Clear and simple comparison, ranking, selection Repeatable process Encourages data gathering and measurability Benchmarking opportunities Programmable Useful input to sensitivity analysis and simulation Connectable to many analytical and statistical models	<i>Benefits</i> Search for meaningful evaluation metrics Broad-based organizational involvement Understanding of problems, benefits, opportunities Problem solving as part of selection process Broadly distributed knowledge base Multiple solutions and alternatives Multifunctional involvement leading to buy-in and risk sharing
<i>Limitations</i> Many success factors are not quantifiable Probabilities and weights may change True measures do not exist Analyses and conclusions are often misleading Masking of hidden problems and opportunities Stifle innovative decision making Lack people involvement, buy-in, commitment Ineffective in dealing with multifunctional issues, nonlinearities and dynamic situations May mask hidden costs and benefits Temptation for acting too quickly and prematurely	<i>Limitations</i> Complex, time-consuming process Biases introduced via organizational power and politics Difficult to procedurelize or repeat Conflict and disagreement over decision/outcome Does not fit conventional decision processes Intuition and emotion may obscure facts Used for justifying “wants” Lead to more fact-finding than decision making Temptation for unnecessary expansion of fact-finding Process requires effective managerial leadership

limitations, as summarized in Table 25.4. Yet, in spite of the limitations inherent to quantitative evaluation and the increased use of qualitative approaches, *virtually every organization supports its project selections with some form of quantitative measures*—most popular ROI, cost-benefit, and payback period. However, driven by the growing complexity of the business environment, and managers are getting increasingly concerned about these limitations and explore alternatives. They often augment quantitative methods with additional measures for determining the long-range cost-benefits of a project proposal to the enterprise. Many of these contemporary decision-making methods rely to a large degree on *qualitative, judgmental decision-making*. These data gathering methods cast a wide net and consider a broad spectrum of factors that are often difficult to describe or quantify, but are effective in gaining strategic perspective and a more comprehensive picture on potential benefits, risks, and challenges of the proposed project.

25.3 QUALITATIVE APPROACHES

While quantitative methods provide an important toolset for project evaluation and selection, there is also a growing sense of frustration, especially among managers of complex and technologically advanced undertakings, that reliance on strictly quantitative methods, does not always produce the most useful or reliable inputs for decision-making, nor are all methods equally suited for all situations (Kulkarni et al., 2004;

Kumar, 2006). Therefore, it is not surprising that for project evaluations involving complex sets of business criteria, narrowly focused quantitative methods are often supplemented with broad-scanning, intuitive processes and collective, multifunctional decision-making such as *Delphi*, *nominal group technology*, *brainstorming*, *focus groups*, *sensitivity analysis*, and *benchmarking*. Each of these techniques can either be used by itself to determine the “*best, most successful, or most valuable*” option, or integrated into a comprehensive analytical framework for *collective multifunctional decision-making*, which is being discussed in the following section.

25.3.1 Collective Multifunctional Evaluations

This process relies on subject experts from various functional areas for collectively defining and evaluating broad project success criteria, employing both quantitative and qualitative methods (Kumar, 2006). *The first step* is to define the specific organizational areas critical to project success and to assign expert evaluators. For a typical product development project, these organizations may include R&D, engineering, testing, manufacturing, marketing, product assurance, and customer/field services. These function-experts should be given the time necessary for the evaluation. They also should have the commitment from senior management for full organizational support. Ideally, these evaluators should be members of the core team ultimately responsible for project implementation.

Evaluation Factors: Early in the evaluation process, the team defines the factors that appear critical to the ultimate success of the projects under evaluation and arranges them into a list, which includes both quantitative and qualitative factors. A mutually acceptable scale must be worked out for scoring the evaluation criteria. Studies of collective multifunctional assessment practices show that simple scales are most effective for leading to actionable team decisions. The four most popular and robust scales for judging situational outcomes are as follows:

1. *Ten-Point Judgment Scale:* from +5 (most favorable) to –5 (most unfavorable);
2. *Three-Point Judgment Scale:* +1 (favorable), 0 (neutral or cannot judge), and –1 (unfavorable);
3. *Five-Point Judgment Scale:* A (highly favorable), B (favorable), C (marginally favorable), D (most likely unfavorable), and F (definitely unfavorable);
4. *Five-Point Likert Scale:* 1 (strongly agree), 2 (agree), 3 (neutral), 4 (disagree), and 5 (strongly disagree).

Weighing of criteria is not recommended for most applications as it complicates and often distorts the collective evaluation.

The Evaluation Process: Evaluators first assess and then score all of the success factors they feel qualified to judge. Then collective discussions follow. Initial discussions of project alternatives, their markets, business opportunities, and technologies involved, are usually beneficial but not necessary for the first round of the evaluation process. The objective of this first round of expert judgments is to get calibrated on the opportunities and challenges presented. Further, each evaluator has the opportunity to recommend (1) actions that could improve the quality and accuracy of the project evaluation, (2) additional data needed, and (3) suggestions for increasing project success. Before meeting at the next group session, agreed-on action items and activities for improving the decision process should be completed. The evaluation

process is enhanced with each iteration by producing more accurate, refined and comprehensive data. Typically, between three and five iterations are required before a go/no-go decision can be reached for a given project.

25.4 RECOMMENDATIONS

Effective evaluation and selection of project opportunities involves many variables of the organizational and technological environment, reaching often far beyond cost and revenue measures. While economic models provide an important dimension of the project selection process, most situations are too complex to use simple quantitative methods as the sole basis for decision-making. Many of today's project evaluation procedures include a broad spectrum of variables and rely on a combination of rational and intuitive processes for defining the value of a new project venture to the enterprise. The better an organization understands its business processes, markets, customers, and technologies, the better it will be able to evaluate the value, risks, and challenges of a new project venture. Further, manageability of the evaluation process is critical to its results, especially in complex situations. The process must have a certain degree of structure, discipline, and measurability to be conducive to the intricate multivariable analysis. One method of achieving structure and manageability calls for grouping the evaluation variables into four categories: (1) consistency and strength of the project with the business mission, strategy, and plan; (2) multifunctional ability to produce the project deliverables and objectives, including technical, cost, and time factors; (3) success in the customer environment; and (4) economics, including profitability. Modern phase management, such as stage-gate[®] processes provide managers with the tools for organizing and conducting project evaluations in a systematic way. The following section summarizes suggestions that can help managers in effectively evaluating and selecting projects toward successful implementation:

Seek-Out Relevant Information: Meaningful project evaluations require relevant quality information. The four sets of variables, related to the strategy, results, customer and economics, as identified above, can provide a framework for establishing the proper metrics and detailed data gathering.

Ensure Competence and Relevancy: Ensure that the right people become involved in the data collection and judgmental processes.

Take Top-Down Look first, Detail Comes Later: Detail is less important than information relevancy and evaluator expertise. Do not get hung-up on missing data during the early phases of the project evaluation. Evaluation processes should be iterative. It does not make sense to spend a lot of time and resources on gathering perfect data, to justify a "no-go" decision.

Select and Match the Right People: Whether the project evaluation consists of a simple economic analysis or a complex multifunctional assessment, competent people from functions critical to the overall success of the project should be involved.

Define Success Criteria: Whether deciding on a single project or choosing among alternatives, evaluation criteria must be defined. They can be quantitative, such as ROI, or qualitative, such as the chances of winning a contract. In either case, these evaluation criteria should cover the true spectrum of factors affecting success and failure of the project(s). The success criteria should be identified by seasoned

enterprise personnel. In addition, people from outside of the company, such as vendors, subcontractors, and customers, are often included in this expert group and critical to the development of meaningful success criteria.

Strictly Quantitative Criteria can be Misleading: Be aware of evaluation procedures based on quantitative criteria only (ROI, cost, market share, MARR, etc.). The input data used to calculate these criteria are likely based on rough estimates and are often unreliable. Furthermore, a reliance on strictly quantitative data, considers only a narrow spectrum of factors affecting project success or failure, thus ignoring many other important factors, especially those that influence project success in a dynamic or nonlinear way, typical for many complex technologically sophisticated undertakings. Evaluations based on predominately quantitative criteria should at least be augmented with some expert judgment as a “sanity check.”

Condense Criteria List: Combine evaluation criteria, especially among the judgmental categories, to keep the list manageable. As a goal, try to stay within 12 criteria for each category.

Gain Broad Perspective: The inputs to the project selection process should include the broadest possible spectrum of data from the business environment that affect success, failure, and limitations of the new project opportunity. Assumptions should be carefully examined.

Communicate Across the Enterprise: Facilitate communications among evaluators and functional support groups. Define the process for organizing the team and conducting the evaluation and selection process.

Ensure Cross-Functional Representation and Cooperation: People on the evaluation team must share a strategic vision across organizational lines. They also must have the desire to support the project if selected for implementation. The purpose, goals, objectives, and relationships of the project to the business mission should be clear to all parties involved in the evaluation/selection process.

Do not Lose the Big Picture: As discussions go into detail during the evaluation, the team should maintain a broad perspective. Two global judgment factors can help to focus on the big picture of project success: (1) overall cost-benefit perspective and (2) overall risk of failure assessment. These factors can be recorded on a 10-point scale, -5 to $+5$. This also leads to an effective two-dimensional graphic display for comparing competing project proposals.

Do Your Homework between Iterations: Project evaluations are usually conducted progressively in iterative cycles. Therefore, the need for more information, clarification, and further analysis surfaces between each cycle. Necessary action items should be properly assigned and followed up to enhance the evaluation quality with each consecutive iteration.

Take a Project-Oriented Approach: Plan, organize, and manage your project evaluation/selection process as a *project*. Proposal evaluation and selection processes require valuable resources that must be justified and carefully managed.

Resource Availability and Timing: Do not forget to include in your selection criteria the availability and timing of resources. Many otherwise successful projects fail because they cannot be completed within a required time period.

Use Red-Team Reviews: Set up a special review team of senior personnel. This is especially useful for large and complex projects with major impact on overall

business performance. This review team examines the decision parameters, qualitative measures, and assumption used in the evaluation process. Limitations, biases, and misinterpretations that may otherwise remain hidden can often be identified and dealt with.

Stimulate Creativity and Candor: Senior management should foster an innovative risk-shared ambience for the evaluation team. Especially, the evaluation of complex project situations involves intricate sets of variables. Criteria for success and failure are linked among many subsystems, such as organization, technology and business, associated with a great deal of risks, and uncertainty. Innovative approaches are required to evaluate the true potential of success for these projects. Risk-sharing by senior management, recognition, visibility and a favorable image in terms of high priority, interesting work, and importance of the project to the organization, have been found strong drivers toward attracting and holding quality people on the evaluation team, and toward gaining their active and innovative participation in the process.

Manage and Lead: The evaluation team should be chaired by someone who has the trust, respect, and leadership credibility with the team members. Senior management can positively influence the work environment and the process by providing guidelines, charters, visibility, resources, and active support to the project evaluation team.

In summary, effective project evaluation and selection requires a broad-scanning process across all segments of the enterprise, and its environment to deal with the risks, uncertainties, ambiguities, and imperfections of data available for assessing the value of a new project venture relative to other opportunities. No single set of broad guidelines exist that guarantees the selection of successful projects. However, the process is not random! A better understanding of the organizational dynamics that affects project performance, and the factors that drive cost, revenue, and other benefits, can help in gaining a better, more meaningful insight into the future value of a prospective new project. Seeking out both quantitative and qualitative measures incorporated into a combined rational-judgmental evaluation process often yield the most reliable predictor of future project value and desirability. As equally important, the process requires managerial leadership and skills in planning, organizing, and communicating. Above all, the leader of the project evaluation team must be a social architect, who can unify the multi-functional process and its people. The leader must be able to foster an environment professionally stimulating and conducive to risk sharing. It also must be effectively linked to the functional support groups needed for project implementation. Finally, organizational strategy must be aligned and integrated with the evaluation/selection process, early and throughout its evaluation cycle. Senior management has an important role in unifying the evaluation team behind the mission objectives and in facilitating the linkages to the stakeholders and ultimate user community. Senior management should further help in providing overall leadership, and in building mutual trust, respect and credibility among the members of the proposal evaluation team, all critical drivers toward a strong partnership of all team members and the basis for an effective enterprise-wide decision-making system. Taken together, this is the environment conducive to cross-functional communication, cooperation, and integration of the intricate variables needed for effective engineering project evaluation and selection.

Terms

Cross-Functional: Actions that span organizational boundaries.

Phase Management: Projects are broken into natural implementation phases, such as development, production and marketing, as a basis for project planning, integration, and control. Phase management also provides the framework for *concurrent engineering* and *stage-gate processes*.

Project Success: A comprehensive measure, defined in both quantitative and qualitative terms, that includes economic, market, and strategic objectives.

Stage-Gate Process: Framework originally developed by R Cooper and S Edgett for executing projects within predefined stages (See also *Phase Management*) with measurable deliverables (*at gates*) at the end of each stage. These gates also provide the review metrics for ensuring successful transition and integration of the project into the next stage.

Weighing of Criteria: A multiplier associated with specific evaluation criteria.

VARIABLES AND ABBREVIATIONS

<i>A</i>	<i>Annuity</i> is the present worth of a revenue or cost at the end of a period n
CB	<i>Cost–benefit</i> , net present value of all ROIs in dollars
i	Prevailing <i>interest rate</i>
I	<i>Investment</i>
IRR	<i>Internal rate of return</i> , the average return on investment realized by a firm on its investment capital
MARR	<i>Minimum attractive rate of return</i> on new investments acceptable to an organization
NPV	<i>Net present value</i> of a stream of future revenues or costs
PBP	<i>Payback period</i> , the time period needed to recover the original investment
PR	<i>Project rating factor</i> , a measure developed by Carlo Pacifico for predicting project success
PW	<i>Present worth</i> (also called annuity), the present value of a revenue or cost at the end of a period n
ROI	<i>Return on investment</i>
z	<i>Project rating factor</i> , a measure developed by Sobelman for predicting project success.

REFERENCES

- Bstieler L. The moderating effects of environmental uncertainty on new product development and time efficiency. *Journal of Product Innovation Management* 2005;22(3):267–284.
- Cicmil S, Williams T, Thomas J, Hodgson D. Rethinking project management: researching the actuality of projects. *International Journal of Project Management* 2006;24(8):5–686.
- El Eman K, Koru A. A replicated survey of it software project failures. *Software (IEEE)* 2008;25(5):84–90.

- Kavadias S, Loch CH. *Project Selection Under Uncertainty: Dynamically Allocating Resources to Maximize Value*. Norwood (MA): Kluwer Academic Publishers; 2004.
- Khorramshahgol R, Azani H, Gousty Y. An integrated approach to project evaluation and selection. *IEEE Transactions on Engineering Management* 1998;35(4):265–270.
- Kulkarni RB, Miller D, Ingram RM, Wong C-W, Lorenz J. Need-Based Project Prioritization: Alternative to Cost-Benefit Analysis. *Journal of Engineering Transportation* 2004;130(2):150–158.
- Kumar PD. Integrated project evaluation and selection using multiple-attribute decision-making technique. *International Journal of Production Economics* 2006;103(1):87.
- Lemon WF, Bowitz J, Burn J, Hackney R. Information systems project failure: a comparative study of two countries. *Journal of Global Information Management* 2002;10(2):28–39.
- Mantel S, Meredith J, Shafer S, Sutton M. *Selecting projects to meet organizational objectives. Project Management Practice. Chapter 1.5*. Hoboken (NJ): Wiley & Sons; 2011. p. 10–22.
- Prabhakar G. What is project success: a literature review. *International Journal of Business and Management* 2009;3(9):3–10.
- Raz T, Shenhar A, Dvir D. Risk management, project success and technological uncertainty. *R&D Management* 2002;32(2):101–109.
- Remer DS, Stokdyk SB, Van Driel M. Survey of project evaluation techniques currently used in industry. *International Journal of Production Economics* 1993;32(1):103–115.
- Shenhar A, Dvir D, Ofer L, Maltz A. Project success: a multidimensional strategic concept. *Long Range Planning* 2001;34(6):699–725.
- Thamhain H, Skelton T. *Success factors for effective R&D risk management. International Journal of Technology Intelligence and Planning (IJTIP)* 2007;3(4):376–386.
- Thamhain Hans. Project evaluation and selection. *Management of Technology*. Chapter 8 New York: Wiley & Sons; 2005.

26

MANUFACTURING SYSTEMS EVALUATION

WALTER W. OLSON

- 26.1 Introduction
- 26.2 Components of environmentally conscious manufacturing
- 26.3 Manufacturing systems
 - 26.3.1 Levels of manufacturing systems
 - 26.3.2 The plan, do, check, and act cycle
- 26.4 System effects on ECM
- 26.5 Assessment
 - 26.5.1 Assessment planning
 - 26.5.2 Data collection
 - 26.5.3 Site visit and inspection
 - 26.5.4 Reporting and project formulation
- 26.6 Summary
- References

26.1 INTRODUCTION

Environmentally conscious manufacturing (ECM) is the production of products using processes and techniques selected to have the least impact on the environment while still being economically viable. Process selection criteria include minimum waste, minimum use of hazardous materials, and minimum use of energy in addition to the production goals. Products produced in this manner are often more competitive in the marketplace because these criteria result in cost reduction, cost avoidance, and increased appeal to the consumer.

This chapter discusses several techniques for assessing and evaluating environmentally conscious manufacturing performance of manufacturing systems. These techniques begin with an analysis of the paperwork/data followed by a visit of the plant floor and

conducting interviews. This provides the basis for making evaluations of improvement areas and performing the tasks necessary to formulate improvement projects. The emphasis here is on improving the manufacturing systems; however, the benefits occur in production.

This chapter focuses on very practical and fundamental issues in manufacturing. As a result, it is not readily applicable to a firm seeking ISO 14001 accreditation. Nor is it meant to be. Whereas ISO 14001 is an environmental management system, this chapter is about manufacturing systems and their evaluation. Ghisellinia and Thurston identified several decision traps that ISO 14001 imposes “the ‘management’ nature of the standard, failure to identify a rigorous environmental baseline, misconception of pollution prevention, inordinate emphasis on short-term goals, focusing on regulatory compliance, and diversion of EMS resources to the documentation system (Ghisellinia and Thurston, 2005).” The processes here seek to avoid these traps. Therefore, there will be little reference to ISO 14001 or life cycle assessment (LCA) in the following sections.

Manufacturing systems are the planning, communications, coordination, monitoring, and management aspects of manufacturing. Industrial engineers often perform these tasks within the overall manufacturing system. The goal of this chapter is to provide guidance to assist the industrial engineer in finding better methods and techniques to make the overall manufacturing system more responsive to the goals of environmentally conscious manufacturing.

26.2 COMPONENTS OF ENVIRONMENTALLY CONSCIOUS MANUFACTURING

Although one can reduce ECM in a number of specific details, all of which seem independent of each other, the engineer would do well to remember three major points:

1. reduce waste,
2. reduce hazardous materials and processes,
3. reduce energy.

Almost every other ECM detail relates to these simple three objectives.

Waste is probably the single most important element. *Wastes are expenditures of resources that are not incorporated into the product.* Resources used in manufacturing are time, money, capital, labor, and materials. Wastes are characterized by emissions, machining offal, cutting fluids, person hours expended in waiting, buffering of items waiting for the next process, and machine downtime due to change over or maintenance, for example. When waste occurs, the product is more expensive.

A major division of manufacturing processes results in material separations. It is important to realize that it may be necessary to have waste in order to produce a product (Figure 26.1). However, the material lost should be minimized. A systematic approach to identifying and eliminating wastes results in increased ECM while also increasing profitability. Often trade-offs must be made. For example, to be economically viable, a process may require the use of a more wasteful material than one where material waste can essentially be eliminated but at a cost that prevents economic production of the product. Although this is an extreme example that is rarely encountered, effective ECM requires

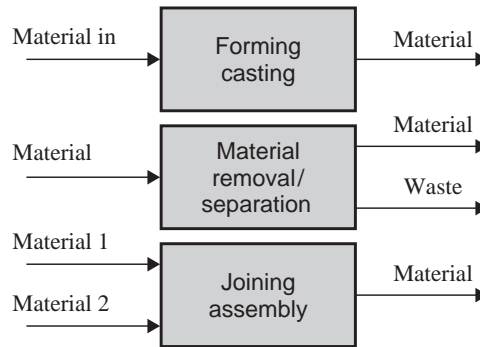


FIGURE 26.1 Classification of manufacturing processes.

that several objectives be met while producing a product that can be marketed successfully.

The greatest waste occurs when a product was produced in quantity that failed to meet market requirements. Thus, it is important to identify wastes and know exactly where wastes are occurring, but it is even more important to know how to reduce waste without incurring greater costs.

Hazardous materials are defined as materials that by their nature, chemistry, or conversion, result in reduced health of those exposed to them, degraded safety, or environmental risks produced by their uncontrolled release. These include toxic, mutagenic, radioactive, inflammable, and explosive substances. In addition to substances, hazards can include such things as noise, compressed fluids, and dusts. In many cases, specific controls and permissible levels are associated to these substances imposed by governments. Because of the additional controls and the additional increased handling awareness, use of these materials increases manufacturing costs.

Every industrial engineer in a manufacturing plant is intimately familiar with the cost of energy. However, they are not necessarily familiar with the environmental damage that energy production creates. Regardless of the form of energy or the source of energy, all energy has associated with environmental damage. For example, direct solar conversion, which many vaunt as clean energy, requires the use of cadmium, selenium, and tellurium—all of which are harmful to the environment. Therefore, the reduction of energy requirements not only reduces manufacturing costs but also reduces environmental damage. Thus, reduction of energy is also an objective of ECM.

26.3 MANUFACTURING SYSTEMS

Systems are required to produce products. A system is an organized whole consisting of subsystems that receive inputs from the external environment, transforming the inputs to outputs (Olson, 2006). Many manufacturing activities required well-defined systematic approaches to ensure efficiency. Manufacturing requires close coordination of labor, tools, materials, and information. If the systems that ensure this coordination are missing or flawed in their performance, waste is produced. These systems must start with the initial concept of the product and extend through the warranty of the product after it is in use. In some cases, this also includes the disposition of the product after the usage phase.

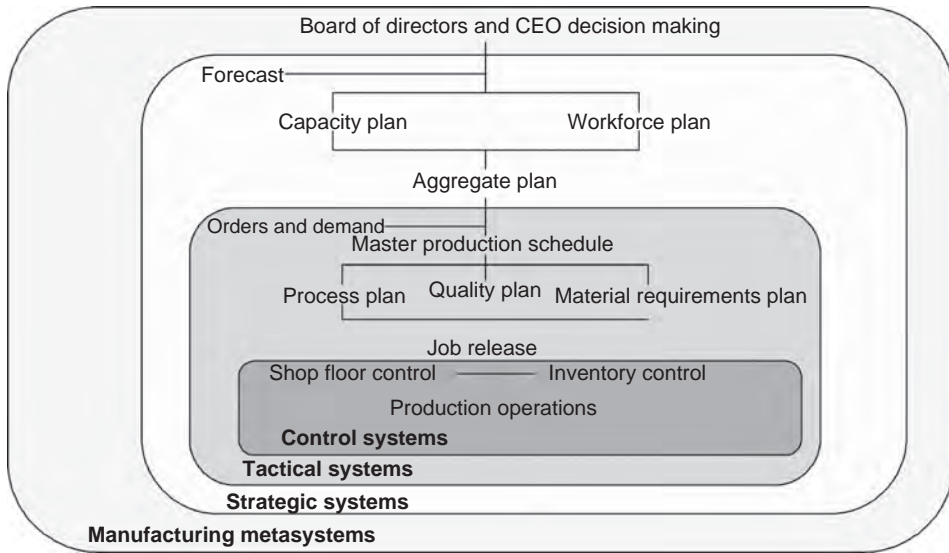


FIGURE 26.2 Manufacturing system levels.

26.3.1 Levels of Manufacturing Systems

Manufacturing systems can be considered at four different levels of refinement (Figure 26.2). At the top level, loosely called the *metasystem*, are decisions about the organization, the subcomponents, and the interactions of the components within the overall manufacturing systems. Factors that influence the metasystem are the top decisions as to what the foci of the businesses are whether to produce to order or produce to stock, the complexity of the product, and the economic responsiveness needed. These will determine the components of the manufacturing system and establish the outputs of the components. The planning horizon for these decisions is the expected life of the corporation. It ranges from years to decades.

At the *strategic* level, forecasting and capacity planning (including facility planning and workforce planning) are the major subsystems. They are taken together with an output of the aggregate plan. The aggregate plan is based on the capacity of the manufacturing firm and includes the labor policies to drive the resources at the firm level. Aggregate planning horizons typically range from 3 months to five years, depending on the size of the firm, the product complexity, and the commitment levels of resources needed. Since it might take 2 years or more to bring a new factory on-line, aggregate plans need to incorporate the long-term economic outlook of the corporation and be consonant with the corporation strategic plan. In fact, one could argue that the aggregate plan is the operational statement of the strategic plan.

The next level down is the *tactical* level of systems. Here, process plans, quality plans, and production scheduling need to be developed. Actual customer orders enter through a demand management system. Depending on decisions made at the metasystem level, orders are met from stock or by introduction of new production orders to the operations. Critical decisions at the tactical level determine what tools, what level of quality, and what response times are needed to meet the aggregate plan. The major output of the tactical

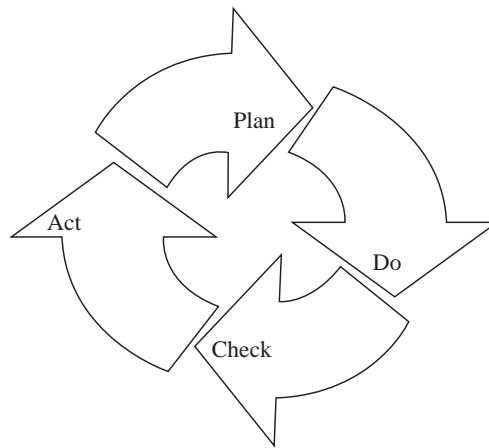


FIGURE 26.3 Plan, do, check, and act cycle.

systems is the production master schedule. At the tactical level, planning horizons range from 1 week to more than a year.

At the lowest system level are *control* systems. Shop floor control, inventory control, quality control, and maintenance control systems are the major components. These controls are driven by the production master schedule and direct and monitor resources to meet the production master schedule. Planning horizons may involve activities of a few seconds up to 3 months.

Decisions that have the most far-reaching effects occur at the metasystem level. As decisions are made at lower levels, they decrease in scope and reach but become more detailed. Thus, decisions at high levels involve more movement of resources and have greater potential for waste reduction than do decisions at the lowest levels.

Although various forms of manufacturing systems may have different names for the systems above and stipulate set of interactions between the systems, any complete manufacturing system will have subsystems performing the functions just described. Essentially, these systems relate real-world factors to models that can be used to direct future manufacturing activities. In this context, it is essential that all systems implement a *plan, do, check, and act (PDCA) cycle*. This is illustrated in Figure 26.3.

26.3.2 The Plan, Do, Check, and Act Cycle

The PDCA cycle, sometimes called the Deming cycle, expresses a relationship between planning, operations/production, evaluations of outcome, and management that are essential in a well-organized and profitable manufacturing system. The “Plan” phase of the cycle prepares the system to meet its operational objectives and goals. In this phase, the events necessary for performance of the system are organized, and the resources needed for production are scheduled. Planning is fed by the action phase that expresses management intent. The “Do” phase is the expression of the system performance. In most cases, the system is named after the “Do” phase as this is where the system produces its product and outputs. The “Check” phase is the assessment and evaluation of the do and plan phases. In the “Check” phase, it important to observe “what went right” as well as “what went wrong.” This is often the role of quality control within the system. Finally, the “Act”

phase is the role of management in the system. Leaders and managers set goals and objectives based on the reports from the check phase and their observations of the do and plan phases. Furthermore, they provide the directive action to correct and improve the system for the next cycle of PDCA.

Within manufacturing systems, most methods to detect system wastes are based on measuring the system variability. In general, if a system is highly variable, the system is not performing well. The major causes of system variability are insufficient resources to perform the assigned tasks, insufficient capacity to handle the workload and insufficient time for planning and execution. Often times, these problems are exacerbated by failure of higher-level systems that provide inputs to the system in question. For example, aggregate plan changes frequently raises havoc with production scheduling and material ordering systems in the attempt to keep the master production schedule responsive. As results, long lead-time materials may be ordered, labor hired, and plant resources designated that are either too much (hence, wasteful) or too little (again, wasteful of time) when actual production is executed.

Overall, the manufacturing systems organize production in an orderly way from high-level goals to the very detailed and mundane activities. Inefficient manufacturing systems results in inefficient production operations. Thus, observations of inefficiency at the operational level must be tracked to the level of system where the inefficiency first was manifested. This process reveals improvement opportunities. Thus far, these systems have been considered independently of environmentally conscious manufacturing. In the next section, it will be shown how these systems affect ECM.

26.4 SYSTEM EFFECTS ON ECM

Manufacturing systems are used to plan manufacturing, develop all of the subsystems, and then control the activities of manufacturing. They have a great impact on ECM. In fact, it is doubtful that ECM could be accomplished without active manufacturing-system participation. At the metasystem levels, there must be an overt commitment to ECM. A company is driven by risk reduction, corporate image, and economic objectives. The leaders of the corporation must understand and accept that ECM can reduce risk, improve corporate image, and improve the economics of the firm. Without such an understanding at the top decision-making levels, ECM stands little chance of being fully implemented. A company pursuing ECM should state that ECM is part of the core business structure. This has the influence of directing the subordinate-level activities toward choices that support ECM.

In addition, the *make to order* and the *make to stock* decisions influence the amount of material that a firm must hold to support its operations. Because of external constraints on certain materials and lead times for certain components, there will be differences in what manufacturing strategy a corporation will take. However, anytime that a company must store raw materials, components, or finished product, there is an increased opportunity for waste to occur. If these materials also include hazardous materials, handling awareness also increases costs.

At the strategic level, ECM is most affected by capacity and facility planning. Facilities determine the processes available to the firm. They also determine the capacity. Process selection for a facility has the greatest influence on whether ECM can be achieved in

a given facility. Any process that requires auxiliary materials to achieve an objective is subject for consideration for improvement under ECM. For example, a process that requires hydrochloric acid to clean a metal surface of rust and preservatives must be carefully examined. There are alternative processes that do not require the acid. However, once the process is selected and built into a facility, cost of replacement, space limitations, and a host of other factors may render it impossible to replace the process and therefore limit the effectiveness of subsequent ECM attempts.

Additionally, at the strategic level, choices are made on the layout of a given facility. Traditionally, these are transfer line, cell or flexible machine layouts, colony or job shop process layout, and (in the most extreme case) stationary product layout. This choice must satisfy the given production requirements that include product, quality, and process flow. Choice of the wrong layout strategy results in increased energy consumption, increased work in progress, increased internal plant transportation, and increased waste.

The second greatest influence on ECM occurs at the tactical level in process planning. Process planning is the conversion of the engineering design to definition of the detailed processes that will enable creation of the product. Process planning can be considered at two levels:

1. The *macro level* where decisions are made as to what process of the facility will be used, in what order.
2. The *micro level* where all of the details of machines, individual tools, and instructions are created.

Because of the influence of process planning on the individual details of creating a product, choices here directly influence waste, use of hazardous materials, and energy consumption. As the old cliché goes, “The devil is in the details!” Process planners intimate with the facility can greatly reduce waste and energy consumption by choice of energy saving and near net shape processes. An additional benefit for this process is that the cost of production is also less.

Another tactical decision that can have an influence on ECM is production scheduling. Long batch-run schedules tend to waste less and consume less energy. Process setups and changeovers are notoriously wasteful and costly. In addition, much of the quality control wastage occurs during the early part of new batch runs. However, the same benefits can often be achieved through planning the facility for flexibility at the strategic level.

At the lowest levels, controls have the least impact on ECM, although significant improvements are still possible. For example, early detection of a faulty process can prevent excess material and energy use. This is even more critical if the process in question is near the beginning of the product’s manufacturing.

The controls are often the best indicators of inadequate ECM decisions made at higher levels, even though they have the least impact. For example, while monitoring the production process, shop floor foremen should be aware of places where waste collects and should report this to the process planners and process engineers. Machine maintenance is often an indicator: frequent stoppages and unplanned maintenance are an indication of an improper process and of waste. Thus, assessment activities focus on data and observations at these levels. But where problems exist, often the improvement will be in a higher-level system.

This section discussed illustrations of systematic effects on ECM. The next sections focuses on assessment and evaluation and improvement methods for achieving ECM.

26.5 ASSESSMENT

Assessment as used here is the examination and evaluation of the effectiveness of a manufacturing facility and its systems in meeting the objectives of ECM. Assessment is properly the *check* phase of PDCA. In assessment, we examine the processes that are occurring and determine how well they function. The purpose of assessment is to benchmark the current state of operations against an ideal state of operations. Plants will have both exemplary operations and operations that do not meet desired standards. It is essential to recognize the exemplary operations for several purposes: exemplary practices should be copied where feasible across the corporation. In addition, recognition of what is being performed right is important to encouraging and rewarding the plant management for making improvements.

There are three parts to assessment: data collection, evaluation, and reporting. During data collection, information is gathered and analyzed to determine how the plant is performing over an extended period, usually a year or more. The evaluation is conducted on-site and is a snapshot of where the plant is at the time of assessment. Reporting is essential in documenting the assessment procedures, the assessment recommendations, and the plant exemplary practices so that they can be communicated appropriately.

26.5.1 Assessment Planning

The time and effort required for assessment will vary, depending on the size of the organization being evaluated. Typically, an assessment team will consist of a team leader and three to five team members. The team leader should have familiarity with the plant and its operations. However, the team leader and its members must be independent of the plant in the corporation organization. It is most common to select members of an industrial staff from another plant to perform the assessment if no formal organization exists in the corporate structure. For small firms with only one plant, these requirements cannot be met; however, assessments can still be performed using either outside resources or by subdividing the plant and limiting the assessments to subunits. An excellent program to assist small companies can be referenced at <http://www1.eere.energy.gov/industry/best-practices/iacs.html> (U.S. Department of Energy Industrial Assessment Program). Often the team leader is a senior industrial engineer. The team members are normally engineers and planners for the organization.

A typical assessment will require about a week to perform, provided the team is trained and has performed previous assessments. This will be far less for small plants and firms having only one plant, but the proportion of time planned will be approximately the same. The first 3 days of the assessment concentrates on data collection and analysis. The fourth day is the on-site plant visit. The fifth day is the on-site evaluation and report writing.

It is not uncommon and in some cases is desired that the first 3 days be separated from the plant visit by a time period of up to a month. This is to allow more time for analysis and to ensure that the correct data have been provided. It is desirable to have actual data, not summaries. However, the manufacturing unit will often try to provide summary only: this must be resisted. The actual data will have timing information and anomalies that are usually absent from summarized data. Where summaries have been furnished, time will

be needed to make requests for the actual data and to receive it in time for the analysis to be completed before the site visit.

The first part of the assessment is analysis of the documentation of a plant. Depending on the plant, this may be performed at the plant location or at another corporate location. Off-site evaluation is preferred to on-site evaluation to reduce potential interference with plant operations, reduce assessment cost, and provide the team with the most time to analyze the data. On-site visits are usually less productive because of the need for the plant management to perform briefings, setup work areas and preinterview team members from the concerned staff. During the first 2 h of the first day, the team leader assigns team members to analyze specific data. One team member should focus on plant processes, quality, and scheduling. Another team member should focus on plant purchases and inventory. Yet, another team member should concentrate on plant emissions and hazardous substances. A team member should be assigned to studying energy, water, and solid wastes for the plant. This analysis will focus on specific manufacturing systems improvement opportunities.

At the end of the third day, the team meets for approximately 2 h to discuss findings, compare notes, and establish what areas are of concern for the on-site visit. Team members should report what they have analyzed, what were the results, and any major omissions that may have occurred. The team leader may ask certain team members to concentrate on certain parts of the facility, based on these findings.

In addition to accessing documents, data collection also includes visiting the processes or facilities being assessed and interviewing managers and employees. The visit is organized by arrival on the site at the beginning of the workday. A management briefing on the plant will review the layout, processes, and safety requirements for the plant. This is usually followed by a management-guided walkthrough of the plant. The management briefings and the management-guided walkthrough should be performed in approximately 2 h. The team leader then assigns members areas of closer inspections and interviews in areas of interest.

At midday, the team should meet in a private location for the midday meal and should discuss any significant findings. Following this meal, the team leader may make additional assignments. The next 3 h should be used to complete the site inspection and to perform any more interviews needed. In multishift operations, assessment team members may be required to attend those shifts for interviews and inspections. Often in a multishift operation, certain operations are only performed in the night shift, or the graveyard shift.

The first hour of the fifth day should be used to formulate any major findings that will be in the team report. At the end of this hour, the plant management is presented with these findings for their further comments. After hearing the plant management, the next 2 h are used to discuss the formulation of recommended improvement projects. The remainder of the fifth day is used to formalize the team report. The formal report should be completed within 2 weeks of the site visit.

26.5.2 Data Collection

During the data collection phase of assessment, an attempt is made to collect all of the documents relevant to the operating state. Usually, a year's worth of consecutive data is needed. Information needed relates to operations, material storage and use, identified hazardous materials, and energy usage. This would include production data in units of production, aggregate plans, production schedules, downtime reports, material safety data sheets (MSDS), quality defect reports (QDR), purchase orders for materials, shipping orders, inventory turnover reports, emission reports, and energy bills. In some cases, the

records will be too voluminous to copy, and the data may have to be examined at the plant location for the usage of the data. Most facilities now keep electronic data rather than paper records. If not, this may be a big area for improving ECM. In addition, if previous assessment reports exist, these should be incorporated into the data analysis and used for comparative purposes.

When electronic records exist, usually the ability to perform statistical analysis is much easier. Basic statistics such as means and variances should be computed and compared to benchmarking statistics if available. In addition, trend analysis may be useful to find processes that may be approaching a critical situation. Special care should be given to data that seem unusual when taken in the context of the complete data stream. Outliers are often improper measurements; however, they may also flag conditions that require further action.

Production schedules indicate how often the production is being changed. In addition, they will indicate by the plan date how much lead time is given to changing the schedule. Short planning lead times are an indicator of poor planning and waste. Downtime reports will provide information on the reliability of a process. Planned schedules should be compared with actual performance to find discrepancies. Where these exist, they should be investigated to determine the reason. QDRs are particularly important, as these will highlight problematic areas with the production line. Most firms today use some form of statistical quality control. If this is not in use, this will flag a very important area for improvement. Excessive QDRs often highlight process planning problems.

Material purchase orders are measures of the inputs that the system is actually consuming. This needs to be compared with shipping orders to determine whether the material ordered is being converted into product. It is particularly important in the collecting and analyzing of purchase orders and shipping orders to observe waste collection volume, type, and frequency. Inventory turnover reports will indicate materials that are stagnant within the system. A high percentage of nonrotating stockages indicates poor production scheduling, poor material planning, or an inability of managers to release one-time stocks once acquired. The result is waste in time, storage, and capital to support maintenance of this inventory.

MSDS will indicate the hazardous materials in use, as well as the handling requirements. The person(s) doing the assessment should observe in the site visit where the materials are and how they are being handled. The emission reports from the facility indicate what substances are being released into the atmosphere. Because of the cost of licensing and compliance, any savings here can have a big impact on ECM.

The energy bills will indicate how energy is used and what types of loads are being billed against. Computation of power (the ratio of real power to total power) and load factors (overall percentage use of electrical equipment) should be performed and graphed. Power factors below 95% should be flagged as opportunities for improvement for reducing energy costs. Load factors below 70% indicate that equipment is not being used near its peak operating requirements. When these data are taken in their entirety, a picture will form of the state of current operations. In addition, inappropriate load factors will be highlighted that are out of place and will become opportunities for improvements. More extensive energy auditing information is located at the Energy Star Web site, <http://www.energystar.gov/>, or the more specific link, <http://www.energystar.gov/index.cfm?c=industry> bus industry plant energy auditing (U.S. Energy Star).

26.5.3 Site Visit and Inspection

The site visit begins with a plant management briefing and plant management-guided walk through. During the management briefing, often in the question and answer period

after the formal briefing, the assessment team should relate any significant findings that may have been noticed in the data and should ask questions regarding these. This discussion will prepare the management staff for the interviews to follow later in the day so they can gather the relevant information.

Information that was found in the documentation should be confirmed on the floor of the facility wherever it possible. During the visit, carefully note the state of the operations. Is the facility adequately lighted? Is the facility operating in an orderly manner? Are the floors and the machines clean? What smells exist, and from where do they originate? Is work in progress entering the areas reserved for transportation and passage? Are you receiving the same messages from the employees and the managers when you interview them and discuss their operations?

In addition, the assessment team member should walk the outside of the facility. It is often surprising what is found on the perimeter of the property of a plant. The team member might spot excess materials, defective product, liquid wastes, and other indicators of ECM defects. In the past, these were often unreported, and they hid errors. These need to be investigated and cleaned up.

26.5.4 Reporting and Project Formulation

The report is formulated in the following outline:

1. Title page
 - a. Name of the plant assessed
 - b. Name and contact details of who the report was prepared for
 - c. Assessment report number
 - d. Date of the report
 - e. Name and contact details of the assessment team leader
2. Executive summary
3. Assessment recommendations
 - a. Title
 - b. Observed problem
 - c. Recommendation
 - d. Estimated costs and benefits
4. Exemplary plant practices
 - a. Title
 - b. Observed practice
 - c. Estimated benefits
 - d. Contact information for more details
5. Synopsis of assessment
 - a. Plant background
 - b. General plant information
 - c. Plant leadership
 - d. Plant processes
 - e. Description of wastes
 - f. Description of energy usage
 - g. Hazardous materials in use

6. Data analysis—synopsis only (details are in appendices)
7. Site visit observations
 - a. What was observed
 - b. Who observed it
 - c. When was it observed
 - d. Why is it significant
 - e. Corrective action or plant response, if any
8. Findings and conclusions
9. Appendices as needed

The most important parts of the assessment report are the assessment recommendations and the exemplary plant practices. Each assessment recommendation (AR) should be itemized. The ARs provide the foundation for creating an improvement project. To do this, an AR should be titled in an active manner, starting with an action verb. An example is “Meter water use to reduce sewage charges.” This then should be followed by a description of anticipated benefits and savings by performing the action. Where possible, this should cite both dollars in savings, as well as fundamental units such as kWh/year or pounds CO₂. This is followed by who observed the operation and what exactly was observed. The team member should take or obtain pictures to illustrate the process recommended for change. Cost of the improvement project should be estimated if possible, with a short financial analysis of payback.

A key component of assessment is two-way communication between the assessment team and the plant management. In the process, there are ample opportunities for this to occur. On the fourth day, the plant management team provides information to the team and may wish certain areas to be emphasized in the study and may make certain recommendations for improvements. The team has discussed with the management staff the findings from the documentation. On the fifth day, the team reports its significant findings to the plant management team for comments.

Exemplary plant practices—or *best practices*, as they often titled—have a similar format. The title should indicate an action: “XYZ Plant uses ultrasonic tank cleaning.” Benefits achieved are itemized. Then the report should give short description of the practice, followed by a contact person in the plant who can provide more information.

The final report must be timely. The report should be ready within 2 weeks of the site visit and should be provided to the plant management team and the corporate vice president for the division. This may result in an order for a follow-up assessment to ensure that the assessment recommendations are being implemented and corrective actions were taken where needed.

26.6 SUMMARY

Manufacturing system evaluation for environmentally conscious manufacturing is based on three overriding goals:

1. reduce waste,
2. reduce hazardous materials and processes,
3. reduce energy.

These goals not only improve the responsiveness of the manufacturing unit to the environment, but they also result in significant cost savings and a better product when performed properly. Evaluation of manufacturing systems for environmentally conscious design begins with an assessment that highlights significant opportunities for improvement. Although the assessment focuses on actual operations, the solutions must be applied to the systems that created, support, and monitor the operations. Whereas ISO 14001 is a worthy goal to pursue and may be essential to the business practices of the firm, it is largely a paperwork exercise that does not solve common problems in manufacturing and does not necessarily lead to continuous improvement of the plant, the product, or the manufacturing systems. The evaluation process described in this chapter has been applied to several hundred different plants in the United States by a large number of auditors and assessors and has been shown to lead to real results in a timely fashion. It can also assist the ISO 14001 aspirations of the firm, if used properly.

REFERENCES

- Ghisellinia A, Thurston DL. Decision Traps in ISO 14001 Implementation Process: Case Study Results from Illinois Certified Companies. *Journal of Cleaner Production* 2005;13:763–777.
- Olson WW. Systems Thinking. In: Abraham MA, editor. *Sustainability Science and Engineering*. Amsterdam: Elsevier; 2006. p. 93.

MEASURING PERFORMANCE OF CHEMICAL PROCESS EQUIPMENT

ALAN CROSS

- 27.1 Introduction
- 27.2 Direct fired heater measurement and process control instrumentation
- 27.3 Crushing and grinding equipment measurements
 - 27.3.1 Basic selection considerations
 - 27.3.2 Jaw Crusher performance
 - 27.3.3 Hammer Mill performance
 - 27.3.4 Ball Pulverizer performance

References

27.1 INTRODUCTION

Engineers are often involved in the design of direct fired process heaters and other types of equipment for processing a variety of liquid, gaseous, or solid feed materials. The petroleum refining and chemical process industries make use of such processing to produce a variety of products, each product material being more useful and valuable than the feed material. In the case of direct fired heaters the design may require relatively simple processing, involving only heat transfer, and consists of raising the temperature of the given feed stream from a low to higher temperature, as in the conversion of crude oil to produce gasoline, diesel fuel, and fuel oil for home heating. In many cases, more complex processing may be required; however, such processing involves heating a process, consisting of a gaseous or vaporized hydrocarbon liquid, to a relatively high temperature, so that thermal decomposition of the feed occurs, producing a mixture of

lower molecular weight components at the heater outlet. The latter process is known as thermal cracking and is used to produce materials such as ethylene used in the manufacture of plastics and other products. Other types of complex processing are used, which involve chemically reacting two or more components at high temperature, to again create a product that is more valuable than the feed material. One such process, steam-gas reforming, converts a gaseous feed material, such as methane, natural gas or a vaporized liquid material such as naphtha, to a product consisting of a mixture of hydrogen and carbon monoxide, known as syngas. With further processing, the syngas is converted into valuable products such as hydrogen, which can be used as a nonpolluting fuel for automobiles, or into ammonia for fertilizer use.

A typical direct fired heater will consist of a multiplicity of tubular heat transfer elements containing the flow of pressurized feed to be processed. The elements are surrounded by a refractory-lined containment structure, which is also provided with a multiplicity of burners, distributed uniformly, and oriented so as to avoid overheating of the heat transfer surfaces.

Because of the many processing factors involved in the latter type of service and because heating of the process streams is accomplished through use of high temperature combustion products, generated by burners, continuous monitoring and controlling of the various processing variables responsible for proper heater performance and extended life of its mechanical components is necessary. The variables to be considered are production rate; product quality; overall thermal efficiency; concentration of undesirable emissions in the combustion product effluent; process stream temperature, pressure and flow rate; fuel composition used for burner firing; fuel temperature, pressure, and flow rate; combustion air pressure, temperature, and flow rate; tubular heating component temperature and pressure; and refractory temperature.

There are other types of equipment used in chemical processing that do not necessarily require the complex process variable monitoring required of fired heating. Some of the equipments used in the processing of solids are crushers and grinders. This type of equipment does, however, require monitoring of a different sort, namely, periodic examination of the feed materials processed, to determine size distribution, physical, and mechanical properties. The latter information allows for the determination of performance, in terms of numerical data, including product size distribution, production rate, and driving motor horsepower. Equations will be provided, as necessary, to indicate the relationship between pertinent variables, as will the manner in which pertinent instrumentation elements are used to measure and control process variables.

27.2 DIRECT FIRED HEATER MEASUREMENT AND PROCESS CONTROL INSTRUMENTATION

Figure 27.1 indicates the arrangement of the measurement and control instruments used for a direct fired tubular heater, processing a hydrocarbon feed at a high temperature, a temperature high enough so that it would be expected that fouling of the internal heat transfer surfaces with coke would occur and eventually progress to a point in time, varying from months to years, to an end of run condition. At this time, production would have to be curtailed or completely halted to avoid tube damage. Coke deposits would then be removed, so that the heating surfaces might be restored to their start of run condition. With reference to Figure 27.1, the components for a typical direct fired

Item No.	Description
4	Decoking system: Oxidation products and coke particles, generated by the decoking reaction, are disposed by the decoking system.
5	Radiant section outlet process product is conveyed to downstream equipment for further processing.
6	The burner fuel transfer line transports a controlled flow of pressurized burner fuel to the burners. The fuel may be gas or oil, and the control system used for this purpose is different for each. The control system for gaseous fuel is indicated.
7	Burner air transfer line conveys a controlled flow of combustion air to the burners.
8	Two or more burner combustion air manifolds allow use of single entry point for a relatively large number of burners.
9	Direct gas fired heater burners.
10	Refractory lined radiant section enclosure of fired heater.
11	Horizontal tube radiant section tubes: Tubes located at the walls of the enclosure and fired on only one side or centrally located between the refractory walls and fired from both sides may be used. Because of the high temperature combustion products at this location, heat is transferred from combustion products to tubes by radiation. The tubes are arranged so as to provide for four parallel streams in both the radiant and convection sections.
12	Refractory lined convection section enclosure of fired heater.
13	Horizontal convection section tubes: These are closely spaced, on two diameter centers, so that combustion products move at relatively high velocity between tubes, transferring heat from combustion products to tubes by convection.
14	Stack discharging combustion products to atmosphere or to equipment designed to separate carbon dioxide from the other combustion products and store or sequester the carbon dioxide underground in an effort to lessen global warming.
15	Stack combustion products flow control damper operator. This device actuated by a controller, automatically positions the stack damper to maintain a negative pressure between the radiant and convection section enclosure, a point corresponding to minimum negative pressure in the enclosure. Pressure at this location must be maintained at a level low enough to prevent an outflow of hot combustion products that would otherwise result in damage to various external heater components and pose a safety threat, as well. Pressure must also be controlled at a level such that excessive cold air infiltration does not occur, as this would result in a significant overall reduction in heater thermal efficiency.
16	Pressure indicator–controller, an instrument that senses pressure at the location previously described and maintains the desired pressure level at this location by appropriate positioning of the stack damper.
17	Oxygen analyzer, an instrument that samples combustion products, at a point above the radiant section, senses oxygen concentration at this point, and controls the flow of combustion air to maintain an excess air concentration of 10–20% in excess of that required for complete combustion by providing a set point signal to a temperature flow indicator—controller, item 18.
18	Combustion airflow indicator–controller, an instrument that senses the burner combustion airflow rate by means of a venturi tube, or orifice meter, which generates a differential pressure, or a comparable device which generates an electrical signal that is proportional to the airflow. The controller maintains the airflow at the proper level required, based on a set point signal from oxygen analyzer, item 17, and positions air stack damper, item 21, accordingly.
19	Burner fuel gas pressure indicator–controller controls pressure at a constant predetermined level, by properly positioning control valve (item 28), thereby allowing proper functioning of item 20.

Item No.	Description
20	Burner fuel gas flow controller–recorder senses flow rate of burner fuel by means of a venturi tube or like flow measuring device, receives an outlet set point signal from item 22 indicating any temperature deviation from that required, and actuates control valve (item 28) to modify fuel flow rate. The recording function permits determination of heater thermal performance for a finite time period.
21	Burner combustion airflow control damper actuator is properly positioned by means of appropriate signal from item 18.
22	Process product outlet temperature sensor–controller provides set point signal that allows flow controller item 20 to maintain burner heat input at a level coinciding with desired product outlet temperature.
23	Outlet tube steam–air decoking temperature indicator allows operator modulation of decoking air–steam ratio so as to maintain decoking temperature at a level, coinciding with a rapid decoking rate, yet one that will not cause damage to the tubular heating heat transfer elements.
24	Decoking airflow indicator–controller assists decoking operator in maintaining a proper decoking temperature.
25	Decoking steam flow indicator–controller also assists in maintaining an effective decoking procedure.
26	Process feed flow recorder–controller operates in conjunction with item 27 to maintain a desired feed inlet temperature. A recording feature permits analysis of overall heater performance for a lengthy period of time.
27	Process feed flow temperature sensor–controller provides set point signal for item 26. Temperature sensors usually consist of thermocouples contained in a protective metal sheath.
28	Process feed flow control valve modulates feed flow by means of signal generated by item 26.
29	Stack combustion product temperature, in conjunction with data from item 20, permits determination of overall heater thermal efficiency.
30	Burner feed gas sample withdrawal connection permits determination of fuel gas composition, which can then be used to obtain overall heat input and thermal efficiency.
31	Process inlet feed flow pressure indicator.
32	Process outlet product flow pressure indicator, at the location shown and at several other intermediate locations, can be used in conjunction with tube metal temperature measurements to determine the adequacy of existing tube wall thicknesses, or to predict tube life.
33	Radiant section outlet tube metal temperature, sensed by a thermocouple welded to this tube and to each of the other three parallel streams, in order to determine whether decoking operations need be commenced, because of tube metal temperatures that exceed the design temperature. Several such thermocouples, uniformly spaced, are also to be located at tubes located at the midpoint of the coil, for all of the four parallel streams.
34	Radiant section refractory temperature measurement is useful in determining whether refractory damage is likely to occur when the heater is operated at through-puts in excess of that for which the heater was designed.
35	Burner combustion air temperature data are required to determine overall heater thermal efficiency.

27.3 CRUSHING AND GRINDING EQUIPMENT MEASUREMENTS

Selection and evaluation of crushing and grinding equipment, as with many other types of processing equipment, is very often made the responsibility of the manufacturer. Thus, the would-be user will supply such information as the type of feed to be processed, the feed to product size reduction ratio, and the desired feed flow rate. The manufacturer, will on the

basis of their experience with similar material and the calculation procedures at their disposal, will then be able to select equipment that will most likely perform in accordance with the user's specifications. In addition, the manufacturer will want to perform tests on the type of equipment that is to be provided, ideally using the same type and size of feed material that is to be processed, so as to demonstrate that the equipment will perform as specified. This procedure, although costly and time consuming, will usually result in equipment that will, in service, perform as intended.

An alternate selection procedure would be one in which the user assumes a greater role in equipment selection, in conjunction with the manufacturer, and would involve having the manufacturer provide sufficient mechanical design data regarding the equipment, and the user having a working knowledge of how one might go about calculating what the performance would be on the basis of the mechanical design of the equipment and the mechanical and physical properties of the feed material. It is the alternate selection procedure, which is the subject of this document. The advantages of having as full an understanding as possible, regarding the workings of the equipment being purchased, usually at great cost, is perhaps obvious, but would, for example, allow for determination of the effects on operational variables such as power requirements, product flow rate capability, and feed/product size reduction ratio when such feed mechanical and physical properties, as feed material size, crush strength, tensile strength, bulk density, and true density are varied.

27.3.1 Basic Selection Considerations

Initial selection will depend upon the initial size of the feed, the final size of the product, and the characteristics of the feed material. For a relatively high strength brittle material such as coal, it must be reduced from a very large size (12 in. or greater) to a very small size (-200 mesh or 0.0029 in.), for use in pulverized coal-fired power boilers, it is not uncommon for processing such a feed in three stages; Thus, for Stage 1, a jaw crusher, as depicted in Figure 27.2 (Perry and Green, 1984) could be used. For Stage 2, a hammer mill, as depicted in Figure 27.3 (Perry and Green, 1997), could be used, and for Stage 3, a

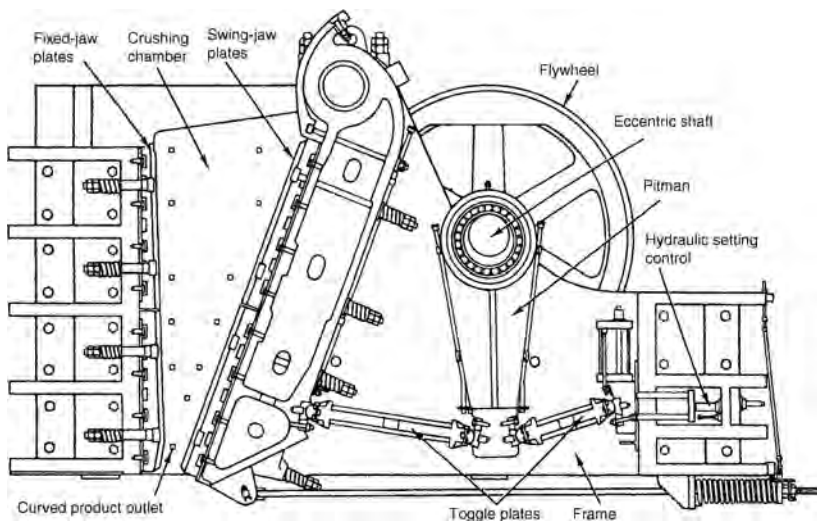


FIGURE 27.2 Primary crusher.

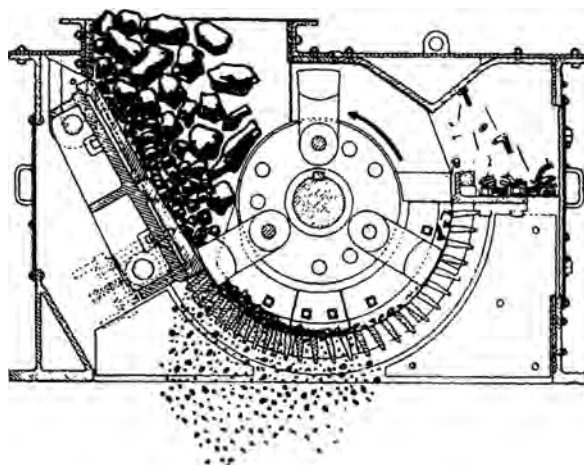


FIGURE 27.3 Hammer mill.

ball pulverizer, or the equivalent, as depicted in Figure 27.4, could be used (Perry and Green, 1984). Realizing that there are many different types, capacities, and operational characteristics of size reduction equipment to choose from, the procedure described herein is just but one of many that can be chosen for purposes of illustrating the selection and calculation procedures that can be used for overall performance evaluation. However,

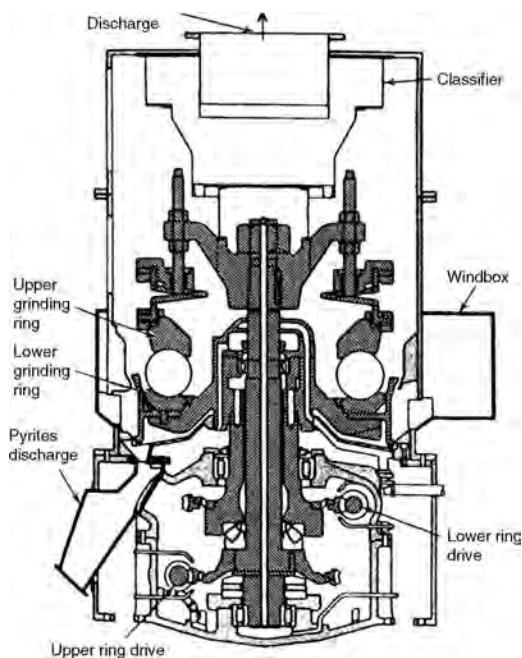


FIGURE 27.4 Ball crusher.

TABLE 27.1 Nordberg Jaw Crusher Model C-80

	Specified Design Data	Calculated Data
Crusher blade operating width, inches	32	32
Crusher blade operating depth, inches	20	20
Crusher blade height, inches, maximum	48	48
HP	100	104
Speed, RPM	350	350
Capacity, thousand tons per hour	60–80	74
Feed size, inches	Not specified	16.5
Product size, inches	2	1.5–2
Setting, inches	2	1.5–2
Number of crushing movements to reduce feed size from 16.5 in. to 1.5 in., see Equation (27.6)	Not specified	6
Bulk density, Lbs/cubic ft	Not specified	47
True density, Lbs/cubic ft	Not specified	95

only three commonly used types of equipment, as referred to above, will be considered herein. Similar methods can nevertheless be used for other types of equipment, feed materials, feed mechanical properties, and feed/product size ranges that may be required.

27.3.2 Jaw Crusher Performance

For illustrative purposes a Nordberg Jaw Crusher Model Number C-80, as depicted generically in Figure 27.2, has been selected for preliminary Stage 1 feed size reduction. Design data are as shown in Table 27.1 considering coal as the feed material.

Assuming coal, the feed material, has a crush strength of 4000 PSI, a tensile strength of 350 PSI, corresponding to about 10% of its crushing strength, and negligible elongation in tension, one may calculate crusher HP as follows:

$$\text{HP} = (\text{Pc})(\text{Acr})(\text{e}) \text{ per rev.} (\text{RPM}) / (33,000)(\text{Eff.}) \quad (27.1)$$

$$(\text{e}) \text{ per rev.} = (\text{Sc})(\text{Ec})(L_{\text{avg.}}) \quad (27.2)$$

where (Acr) is the crusher blade face area, square inches; (RPM) the crusher drive rotational speed, revolutions per minute; (e) per rev. the average compressive deformation of feed particles, feet/rev.; (Sc) the compressive strength, PSI; (Ec) the feed modulus of elasticity, million PSI; ($L_{\text{avg.}}$) the arithmetic average of feed thickness between crusher blades, inches; and (Eff.) the fractional overall mechanical efficiency of crusher.

Maximum coal feed rate is calculated as follows:

$$W_{\text{product}} = \left(\frac{L_{\text{cw}}}{L_{\text{product}}} \right) \left(\frac{L_{\text{product}}^3}{1728} \right) (\text{Dens.}) (\text{RPM}) \left(\frac{60}{2000} \right) \quad (27.3)$$

where (W_{product}) is the product flow rate, Tons/h, noting that downward flow of solids occurs when blades are open and stops when blades are closed; (L_{cw}) the crusher blade width, inches; (L_{product}) the cubical product size, inches; (Dens.) the true density of product, Lbs/cubic foot; and (RPM) the crusher operating speed, revolutions per minute.

Observation of particles of brittle material, such as coal, undergoing compression reveal that the force generated by compression causes the faces of the particles to bulge outward, resulting in multipoint cracking at the faces of the feed particles and ending with fragmentation of the initial particles to produce many smaller size particles. It is concluded that when particles subjected to compression fragment, the fragmentation is actually due to tensile failure of the material, at the many small cracks that form at the faces of the material, which then propagate to the central portions of the material. The reduction in particle size may therefore be predictable from knowledge of the compressive force acting on the material and the mechanical properties of the material. In the case of a jaw crusher, feed size reduction occurs with each forward movement of the crusher blades. In other types of size reduction equipment components of a different type perform the same function. The following mass balance provides a relationship between feed particle size and product particle size, assuming the particles to be cubical in shape.

$$(L_p)^3(N_p)(\text{Dens.}) = (L_f)^3(\text{Dens.}) \text{ and } (N_p) = (L_f)^3/(L_p)^3 \quad (27.4)$$

Furthermore, the compressive force on the feed articles must be equal to the force required to fragment the particles by tensile failure, or;

$$(S_c)(L_f)^2 = (N_p)(S_t)(6)(L_p)^2 \quad (27.5)$$

where (S_c) is the compressive strength of particles, PSI; (L_f) the feed particle size, assumed cubical in shape, inches; (N_p) the number of particles; (S_t) the tensile strength of particles, PSI; (6) the number of particle surface faces; and (L_p) the product particle size, also assumed cubical in shape, inches.

The number of forward crusher blade movements required to reach the desired final particle size is calculated as follows:

$$N_{\text{comp.}} = L_1 + L_2 + L_3 + \cdots + L_n = L_{\text{crusher blade}} \quad (27.6)$$

where (N_{comp}) is the number of compressions required by one forward movement of the crusher blade, with the resultant conversion the feed to product size; $(L_{\text{crusher blade}})$ the total crusher blade height, equals 48 in.; and (L_1) , (L_2) , (L_3) , \dots , (L_n) the sum of the particle lengths from initial to final size, as determined from Equation (27.5), inches.

27.3.3 Hammer Mill Performance

For illustrative purpose, a Meadows Hammermill, Model 90, as depicted generically in Figure 27.3, was selected for Stage 2 size reduction. Design data are as shown in Table 27.2.

The hammer mill is to be of the bottom discharge type wherein the feed free falls from the top of the mill and encounters the rotating hammers. These hammers initially impact and fracture the feed particles, followed by outward and downward flow of the particles at the walls of the mill enclosure, where secondary impact of the particles occurs. The initial impact reduces the particles from their initial size to an intermediate size. The secondary impact reduces the particles from their intermediate size to a final product size. A screen below the hammers, perforated with openings equal to the final product size desired is provided, so that the particles from the first and second impact, which are larger than these openings, are impacted again by the hammers. Thus, the final product collected does not contain oversize material. For much finer product sizes, the mill is provided with

TABLE 27.2 Meadows Hammer Mill, Model 90

	Specified Data	Calculated Data
Horsepower	75–200	204
Operating rotational speed RPM	3,600	3,600
Grinding capacity, pounds per hour	6,000–12,000	12,000
Screen size, width × length, inches	30 × 37 – 1/4	30 × 36 – 1/4
Screen size openings selected, inches	1/2	1/2
Number of hammers per rotor revolution	42	42
Weight per hammer, pounds	1.52	1.52
Feed size, inches	Not specified	1 – 1/2
Intermediate product size, inches	Not specified	0.9
Final product size, inches	Not specified	1/2
Size of hammers, inches	7 × 2 × 3/8	7 × 2 × 3/8
Diameter of rotor discs	12 in.	12 in.
Swing of extended rotor hammers, inches	22	22
Fan provided	No	No
Feed opening, width × height, inches	30 × 11	30 × 11
Bottom screen openings, inches	Not specified	1/2
Number of crushing passes for product to pass through screen openings	Not specified	2

an exhaust fan that extracts the final product by creating a flow of entrainment air in an upward direction through the mill enclosure.

A momentum balance, wherein the change in hammer velocity on impact multiplied by the mass flow rate of the hammers, is used to determine the force necessary to fracture the feed to the final product size desired. Thus,

$$((N)_h ((\text{RPM})/60))(W)_h((V)_{hi} - (V)_{hf}))/ (g) = (F)_p \quad (27.7)$$

$$(V)_{hi} = (\pi)(D_h)(\text{RPM})/(60) \quad (27.8)$$

$$(F)_p = (S)_c(L)_{p\text{-avg.}}(N)_{p\text{-avg.}} \quad (27.9)$$

$$(N)_p = (t)_p(W)_p/(w)_p(3600) \quad (27.10)$$

$$(w)_p = (L_p)^3/(1728)(\text{Sp.vol.}) \quad (27.11)$$

Regarding Equation (27.7), the hammer motor drive is to be such as to maintain a reasonably constant initial hammer velocity at a level which ensures proper feed size reduction, since it has been determined from Equation (27.7), that the initial velocity of the hammers decreases considerably after impacting the feed particles. Proper motor selection will, however, quickly return hammer velocity to its initial value.

Drive motor horsepower is determinable from Equation (27.12) and rated HP, so calculated, is in reasonably good agreement with the specified HP given in Table 27.2.

$$(\text{HP}) = (S)_c(144)(N)_{p\text{-avg.}}((w)_p/(60))/(33,000)(\text{Eff.})(\text{Sp. vol.}) \quad (27.12)$$

where $(W)_h$ is the hammer mass flow, pounds/second; $(V)_{hi}$ the hammer initial velocity, feet/second; (D_h) the rotor diameter with hammers fully extended, feet; $(V)_{hf}$ the hammer final velocity, feet/second; (g) the 32.2 feet/(second)²; $(F)_p$ the average force on particles, pounds; $(w)_p$ the particle weight, pounds; $(t)_p$ the residence time at hammer face, based on free fall velocity of particles, seconds; $(W)_p$ the feed flow rate, pounds/hour; (HP) the horsepower required; $(S)_c$ the crush strength of particles, PSI; $(L)_p$ the width or length of cubical particle, inches; $(N)_p$ the number of particles, based on free fall velocity; (Eff.) the overall fractional efficiency of Mill; and (Sp. Vol.) the specific volume of feed, Cubic feet/pound.

To evaluate $(L)_{p-avg.}$ and $(N)_{p-avg.}$ in Equations (27.11), calculate $(N)_p$ and w_p using Equations (27.10) and (27.11), with initial and intermediate values of $(L)_p$. $(L)_{p-avg.}$ and $(N)_{p-avg.}$ are then considered equal to the arithmetic average of the latter variables at the initial and intermediate values of $(L)_p$.

27.3.4 Ball Pulverizer Performance

For illustrative purposes a B&W Ball Pulverizer, Model EL 76, as depicted generically in Figure 27.4, was selected for Stage 3 size reduction. Pertinent data for this pulverizer are as shown in Table 27.3.

Feed enters the top of the mill and onto and between the circular arrangement of balls. A ball size of 8 in. size was selected so that cubical feed particles, 0.5–1 in. in size, would not be pushed by, but rather would be over ridden by the balls. This was done by ensuring that the force pushing the particles forward was less than the frictional force between the feed particles and the ball race, that is, less than the downward force on the balls multiplied by the coefficient of sliding, or in this case, rolling friction.

An upward flow of preheated air dries the particles so as to facilitate their breakup and to remove those particles having the desired final product size from the pulverizer.

A total of 12 crushing passes were calculated as being necessary to reduce 0.5 in. cubical feed particles to a –200 mesh product, using the procedure described previously, in the case of the primary jaw crusher. This compares with a total of 24 crushing passes available on the basis of the specified rotational speed of the pulverizer, equal to 90 RPM, and a total of about 50, 0.5 in. feed particles at the base of each ball, occupying an area of 3.5 × 3.5 in. under each ball.

TABLE 27.3 B&W Ball Pulverizer Model EL 76

	Specified	Calculated
Horsepower	300	278
RPM	90	90
Diameter of ball race, inches	76	76
Diameter of balls	Not specified	8
Number of balls	24	24
Feed capacity, tons per hour	20	20
Number of crushing passes	24	12 for 200 mesh, 12 for –200 mesh
Initial particle size, inches	Not specified	0.50
Final particle size	Minus 200 mesh	0–200 mesh

Based on the latter calculations, we may calculate the required HP as being equal to the following:

$$\text{HP} = (F)_b(N)_b(C_{rf})(V)_b/(33,000)(\text{Eff.}) \quad (27.13)$$

$$(F)_b = ((0.785)(D_b)^2/(2))((S_c)/(2)) \quad (27.14)$$

$$(V)_b = (\text{Pi})(D_{br})(\text{RPM}) \quad (27.15)$$

where HP is the mill horsepower; $(F)_b$ the load on ball, pounds; (C_{rf}) the coefficient of rolling friction; $(N)_b$ the number of balls; $(V)_b$ the linear velocity of ball race, FPM; (S_c) the crushing strength, PSI; (D_{br}) the diameter of ball race, feet; (RPM) the ball race rotational speed, revolutions per minute; (C_{rf}) the coefficient of rolling friction; and (Eff) the overall pulverizer fractional mechanical efficiency.

The HP so calculated was almost equal to the specified HP and consistent with a coefficient of rolling friction equal to 0.003, a value typical for that of ball bearings.

The calculation of the total load on the balls was based on the assumption that the maximum compressive stress on the balls, based on their spherical shape, could be approximated by the average of the compressive stress immediately below the ball and 0 stress at the ball diameter. Similarly, the contact area of the ball with the feed particles would be 0 immediately below the ball and equal to the maximum cross-sectional area of the ball at the ball diameter.

It has been demonstrated that specifications normally provided by size reduction equipment manufacturers, perhaps at the time a prepurchase quotation is provided, can be used to predict performance of the equipment, should the need arise. For the purchaser to perform the calculations needed to do this, using the calculations provided herein, the basic data required, namely, the crushing strength, tensile strength, bulk density, and true density of the feed material, are needed and are obtainable from an appropriate laboratory.

REFERENCES

- Perry RH, Green DW. *Perry's Chemical Engineers Handbook*. 6th ed. McGraw Hill; 1984.
 Perry RH, Green DW. *Perry's Chemical Engineers Handbook*. 7th ed. McGraw Hill; 1997.

INDUSTRIAL ENERGY EFFICIENCY

B. GOPALAKRISHNAN, D. P. GUPTA, Y. MARDIKAR, AND S. CHAUDHARI

- 28.1 Introduction
 - 28.1.1 Energy usage in manufacturing industry
 - 28.1.2 Benchmarking industrial energy consumption
 - 28.1.3 Potential for industrial energy conservation
 - 28.1.4 Initiatives of the U.S. department of energy for energy efficiency
 - 28.1.5 Economics of energy efficiency and importance to company's bottom-line
 - 28.1.6 Environmental benefits of energy efficiency
- 28.2 Literature review
- 28.3 Data analysis of energy efficiency measures
 - 28.3.1 Utility bill analysis
 - 28.3.2 Analysis of IAC database
- 28.4 Energy efficiency measures in major energy consuming equipment
 - 28.4.1 Furnaces, ovens, boilers, and steam systems
 - 28.4.2 Motors
 - 28.4.3 Lighting
 - 28.4.4 Chillers
 - 28.4.5 Heating, ventilation, and air-conditioning equipment
 - 28.4.6 Air compressors
 - 28.4.7 Combined heat and power (CHP)
- 28.5 Case studies of development of energy-efficiency measures
 - 28.5.1 Glass manufacturing
 - 28.5.2 Plastics manufacturing
 - 28.5.3 Metals manufacturing
 - 28.5.4 Chemicals manufacturing
- 28.6 Conclusion
- Acknowledgments
- References

28.1 INTRODUCTION

Environmentally conscious manufacturing (ECM) has made significant gains on account of advances in energy efficiency. Energy efficiency relates to reducing wasted energy, hence reducing energy consumption. Utilization of fossil fuels adversely affects the greenhouse gases released into the atmosphere and results in undesirable quantities of emissions. Increasing energy efficiency will reduce the unwanted environmental effects produced by manufacturing industry.

28.1.1 Energy Usage in Manufacturing Industry

U.S. manufacturing plants, mines, farms, and constructions firms currently consume about 25 quads (quadrillion British thermal units, or Btu) of energy each year, which is about 33% of the nation's total consumption of energy (Gopalakrishnan et al., 2005b). A few of the cost-effective methods for improving energy efficiency are general housekeeping and maintenance programs, energy management and accounting programs, improved methods, and procedures for existing production methods and product changes. There are several technologies associated with heat recovery, high efficiency motors, variable speed drives, and cogeneration that have applications in many industries. Potential areas can be identified for energy conservation measures, namely lighting, boilers, motors and pumps, destratification of air, insulation of steam lines and process equipment, air and steam leaks, and compressors.

An important factor for industrial energy utilization and efficiency is the corporation's internal culture and external relationships. Capital investment in modern equipment usually enhances energy efficiency even when efficiency is not the primary purpose of the investment. An operations manager's key priorities are retaining efficiency of the production line. In this context, reducing energy costs tend to be a low priority.

28.1.2 Benchmarking Industrial Energy Consumption

The energy costs at a facility can be controlled by effective energy analysis, diagnostics, and management. Energy management is an important aspect for various manufacturing and service operations across U.S. industries to reduce operating costs (Gopalakrishnan et al., 2005a). The management in manufacturing companies is often interested in focusing on efficiency targets and comparing their performance with respect to established industrial practices. In this process, *benchmarking* is often used to identify performance levels with respect to the competition or the industry on an average. In terms of energy consumption in a manufacturing facility, benchmarking can be used effectively to judge whether there is an opportunity to implement significant energy efficiency measures. As an example, the energy use for a particular mill can be compared with that for similar mills or with that for a model mill representing the current best practice (Francis et al., 2004).

28.1.3 Potential for Industrial Energy Conservation

A significant potential indeed exists for energy reduction in the manufacturing community. Energy efficiency and subsequent reduction can be achieved by focusing on the manufacturing processes, utilities that support these processes, and total productive maintenance

procedures. Achieving energy efficiency depends on the technical aspects of operations, as well as the necessity for a cultural change in the viewpoint of management, in day-to-day operations (Sundarajan, 1999).

Industry's options for reducing energy costs were summarized in a study (National Association of Manufacturers, 2005). This study primarily identified energy efficiency measures that yield energy and environmental benefits for large volume, commodity and process industries. Prioritization of these opportunities based on various criteria incorporated in the study suggested 5.2 quadrillion Btu—21% of primary energy consumed by the manufacturing sector. These savings equate to almost \$19 billion for manufacturers, based on 2004 energy prices and consumption volumes (National Association of Manufacturers, 2005).

28.1.4 Initiatives of the U.S. Department of Energy for Energy Efficiency

The major initiatives taken by the U.S. DOE toward industrial energy efficiency include the establishment of industrial assessment centers, BestPractices, plant-wide energy assessments, industrial systems initiatives, and Industries of Future (IOF) program. In order to help industries identify and then capture EEM (energy efficiency measures) and EEO (energy efficiency opportunities) as of 2006, the USDOE EERE/ITP has supported the establishment of 26 industrial assessment centers (IAC) throughout the country. The IAC are located in universities and avail themselves of engineering students under the guidance and supervision of professors to conduct facility resource assessments. The assessments then help industries identify measures they can take to harness EEM and EEO. The IAC have been funded by the U.S. Department of Energy (USDOE), Office of EERE, Industrial Technologies Program (ITP) since 1976.

The focus of the IAC has been toward reducing industrial energy consumption, mainly for small and medium-sized facilities. The no-cost industrial assessments performed by the IAC are available to manufacturing facilities in the SIC 20–39, provided that they have no in-house professional staff to perform the assessment, have gross annual sales below \$100 million, have less than 500 employees at the plant site, have annual utility bills more than \$100,000 and less than \$2.5 million, and be located within 150 miles of the IAC. The energy assessment process includes the analysis of the utility costs and on-site data gathering using effective instrumentation. The discussion with the plant personnel adds to the practicality of the recommendations and the use of the proper support data. One additional benefit from the program is that the data generated by the assessments provides a unique opportunity to quantify the state of energy, waste and productivity management in small and medium-sized manufacturing facilities and the potential of the assessment process to improve efficiency. Since 1981, the data have been compiled from the assessment performed under this program (Department of Energy, 2005a).

Another initiative started by U.S. DOE is BestPractices, a program area within the ITP that supports ITP's mission to improve the energy intensity of the U.S. industrial sector through a coordinated program of research and development, validation, and dissemination of energy-efficient technologies and practices. BestPractices develops, implements, and disseminates best practices in energy management. The manufacturers' participation in BestPractices program can earn the organization technical, financial, and engineering personnel support for achieving energy efficiency. The BestPractices program has provided invaluable software for industrial use known as decision tools for industry.

These decision tools include tools such as process heating assessment tool (PHAST), steam system assessment tool (SSAT), steam system scoping tool (SSST), AirMaster+, MotorMaster+, PEP, and 3Eplus.

Plant-wide energy assessments help large-sized industries to save energy. Plants are selected on a competitive basis and agree to share costs of energy assessment spaced throughout a year. A maximum award of \$100,000 can be granted to a plant in this type of assessment, and equivalent or greater cost is shared by the selected company (Department of Energy, 2006a). Since this assessment is spaced throughout a year, it is a comprehensive energy consumption study and leads to significant EEM.

Industrial energy systems account for 80% of energy consumed by industry. Industrial energy systems are an initiative by which research and development efforts are focused on improvements in basic systems such as steam, process heating, compressed air, motors, pumps, fans, and combustion. This area provides potential of 10–20% reduction in energy usage, which converts approximately 1.6–3.2 quads (Department of Energy, 2005b).

The U.S. DOE has developed the Industries of Future (IOF) program to develop energy conservation opportunities in energy-intensive industries such as aluminum, steel, chemicals, metals casting, glass, mining, forest products, and petroleum and refining. Each project under this category meets rigorous criteria of industry participation, cost sharing, and relevance to industry-defined priorities. To date, the savings from this initiative have totaled to 1.6 quadrillion Btu, or \$6.5 billion (Department of Energy, 2006b).

28.1.5 Economics of Energy Efficiency and Importance to Company's Bottom-Line

Most manufacturing facilities seek a 2-year payback on investment in terms of any projects. It is important that the projects that promote energy efficiency be practical and that they afford economic desirability in terms of implementation by the manufacturing facilities. U.S. manufacturing must extract the highest percentage of productive energy from each Btu of energy produced, especially since it has a high per capita consumption of energy in the industrial sector. This is necessary for competing in the global marketplace. A significant portion of the operating costs of any manufacturing or service industry is in the form of energy costs. Hence, energy conservation and management is an important activity within any manufacturing or service organization, as it can reduce operating costs. The ability of a company to compete for business in an international scale is very important, especially in today's global business environment. Energy management is an important function within any manufacturing or service organization as it pertains to operating costs. Any reduction in operating costs is bound to increase the competitive edge of the industry.

28.1.6 Environmental Benefits of Energy Efficiency

Burning fossil fuels generate CO₂, NO_x, SO_x, and other gases that create an environmental problem. For each kWh of electricity that is reduced, the CO₂ emissions are reduced by as much as 2.19 lb. Similarly, for every MCF (1000 cubic feet) of natural gas that is not burned, the CO₂ emissions are reduced by approximately 113 lb. This indicates that significant emissions reductions are possible due to energy efficiency measures implemented in manufacturing facilities.

28.2 LITERATURE REVIEW

Since the mid-1990s, manufacturing plants in the United States have been undergoing a major shift with measures to reduce energy consumption. The U.S. Department of Energy's Office of energy efficiency and renewable energy (EERE) has encouraged the efforts through its various programs.

A key feature of any energy conservation program is the energy assessment (U.S. Department of Energy, 1978). Management commitment is the most important thing in an overall sense (Capehart et al., 1997). The methodology for conducting an energy assessment that attempts to obtain a comprehensive view of the energy utilization in all forms is important (Warfel, 1993). It addresses the issue that many of the energy assessments conducted by companies have the potential to overlook substantial energy conservation opportunities. It describes how plant performance is estimated based on limited data. According to the *Energy Auditor's Handbook* (Thumann, 1979), the energy audit is a brief, on-site survey and analysis of a plant and its energy usage patterns, identification of opportunities to save energy through the implementation of operational and maintenance changes, and an assessment of its need for implementation of energy conservation measures. Energy auditing is the first phase in improving the energy efficiency of a facility (Hoshide, 1995). Preparation of the energy management report, implementing, and monitoring are the other three phases. The implementation of energy-efficient technologies among manufacturing firms can lead to short-term as well as long-term benefits (Brown et al., 1994). There is also an environmental benefit from projects, the reduction of fossil fuel consumption. For example, the Pacific-Northwest region pursues energy conservation as a source of energy (Spanner et al., 1993). When energy problems caused by the rapid increase in demand in the face of dwindling fuel supply first became apparent, the immediate response was to seek new supplies and alternative fuels (Mate, 2002).

A rational approach to targeting energy costs and outlining the key strategies in energy assessment has been specifically determined for iron foundries (Meffert, 1999). Large metal-casting foundries are excellent targets for energy saving opportunities and should be targeted for continuous technological improvements. The nature of energy conservation opportunities that exist in glass manufacturing facilities are varied (Gopalakrishnan et al., 2001a). The energy management program pursued by a specialty chemical manufacturer has shown significant benefits (Fendt, 2001). The strategy used in a chemical plant to reduce energy consumption is unique in terms of process and utility parameters (Weber, 2001). Effective instrumentation is needed to facilitate a chiller plant energy efficiency monitoring as a prerequisite to any efficiency improvement initiatives for centrally air-conditioned buildings (Yik and Burnett, 1995).

As mentioned earlier, the U.S. DOE Industries of the Future (IOF) program includes industries such as agriculture, aluminum, chemical, forest products, glass, metal casting, petroleum refining, and steel, and these industries consume around 33% of all energy used in the United States¹ (OIT Clearinghouse, 2001). These industries have developed strategies to forecast growth and have identified the tools that are needed to increase their competitiveness. The plant-wide assessment at a paper mill identified five opportunities resulting in overall savings of over \$1.5 million (Augusta Newsprint: Paper Mill Pursues Five Projects Following Plant-Wide Energy Efficiency Assessment 2003).

¹Energy Information Administration. United States Background. Available at <http://www.eia.doe.gov/emeu/cabs/usa.html>.

Assessments at the world's leading aluminum manufacturing plant identified \$60 million in annual savings opportunities companywide.² Energy-saving opportunities included improved heat recovery, better furnace operations, and development of process-energy use targets.

Since DOE began the plant-wide energy assessments program, several plants have submitted successful proposals (Department of Energy, 2006a). The total energy savings that have been realized through this program has been significant and has saved millions of dollars (Department of Energy, 2006a). Only when the energy problems faced in the plant are reviewed in a total perspective do the solutions and the various conservation opportunities become visible. This is called the *systems engineering approach*. In a systems approach, the systems are evaluated macroscopically first and then micro studies are initiated only within the framework of the macro study to ensure the effectiveness of the assessment.

The effectiveness with which energy resources are converted into work is often termed as *energy efficiency* (National Association of Manufacturers, 2005). One example of a metric used often in evaluating heat-producing systems is thermal efficiency. Factors that have an effect on thermal efficiency include completeness of combustion, heat balance, and operating system characteristics (Role of Energy Efficiency: Understand the Value of Energy System Efficiency, 2004). The manufacturing facilities have varied overall energy efficiency that depends on operating system characteristics, production aspects, and technological considerations. Efficiency can be influenced by the design aspects of the manufacturing equipment, as well as by the extent of use. In the manufacturing sector, these energy losses amount to several quadrillion Btus (quadrillion British Thermal Units, or quads) and billions of dollars in lost revenues every year (Department of Energy, Industrial Technologies Program, 2004).

28.3 DATA ANALYSIS OF ENERGY EFFICIENCY MEASURES

This section describes the energy cost components for the energy bills and the analysis of the IAC database. Every manufacturing facility needs to understand how it is being billed for electricity and natural gas. Many facilities do not understand their rate structure, especially when it comes to charges for electrical billing demand. In terms of natural gas, many facilities may be placed in a position to pay high rates based on the type of contract they negotiate.

The Industrial Assessment Center database is extensive and contains elaborate information on all energy assessments conducted by the various IAC since 1981. The analysis of the database can reveal interesting energy profile and savings potential for various types of manufacturing facilities. The database is also a treasure house of knowledge in terms of energy conservation opportunities and the economic impact potential of these opportunities in terms of industrial implementation.

28.3.1 Utility Bill Analysis

The utility bill for a facility may have different components based on the need of the processes. Typically, the facilities have electricity and natural gas as major energy

² Alcoa teams with DOE to reduce energy consumption. Available at <http://www1.eere.energy.gov/industry/bestpractices/pdfs/36152.pdf>.

sources, and some of them may have other utilities such as oil, coal, wood, and water. Since electricity and natural gas are common components in most utilities, this section will be focused on discussion of these utilities. In general, all the facilities will have a contract with the utility company that will dictate the price per unit of electricity and natural gas. Many a time, the facility may have two or more rate schedules to choose from. Based on the data collected from different facilities, analysts may be able to study monthly or yearly usage pattern to find significant savings in terms of utility charges. In some cases, the facility may have done the analysis and chosen the best possible rate but over time, because of change in the usage pattern, the current rate schedule may not be optimal. In other cases, the utility company may decide the rate based on the maximum need of the facility without considering the usage pattern, which may result in more utility cost to the facility. The following section will discuss some key components in electricity and natural gas energy bills.

28.3.1.1 Components in Electricity Energy Bills The required energy cost mainly has three components (Will, 1993): (1) fixed costs (consumer/customer charges, administrative costs), \$/month; (2) electricity costs—that is, the real cost of electricity that is consumed by the process (variable costs); and (3) demand costs—that is, the cost of maintaining a level of energy to run the operation (investment costs), \$/kW. Fixed costs (consumer/customer charges, administrative costs) have no direct relation to the amount of energy used during the billing period. They cover the expenses in readings, accounting, and billing by the power supplier company, which is fixed in any billing period. In general, this cost component is insignificant in comparison with the energy and demand charges. The major components in the electricity bills are electricity (or usage) and demand cost.

28.3.1.2 Energy Charges Energy charge is based on the direct consumption of the electricity in terms of kilowatt-hours (kWh) during the electricity consumption period. The kilowatt-hours value is multiplied by the energy charges per unit (\$/kWh) for the total bill in the billing cycle. These charges may vary based on the service provider, voltage, and energy consumption during each billing cycle (Industrial Rates, Utilities in Greater Cleveland, 2005; West Penn Power Company, 2005; Monongahela Power Company, 2005a, 2005b; Allegheny Power, 2005). A summary of different rate structures may be given as a flat rate for each kilowatt-hour consumed by the facility, a variable rate based on the time of day during which the electricity is consumed, a variable rate based on the time of the year during which the electricity is consumed, a flat or variable rate with low power factor penalty, or a flat or variable rate based on the total amount of power (kVA or kW) consumed.

28.3.1.3 Demand Charges This charge is to compensate the utility company for the capital investment required to serve peak loads, even if that peak load is used only for partial operating period. The demand is measured in kilowatts (kW) or kilovoltamperes (kVA). These units are related to the energy (kWh) consumed in a given time interval of the billing period. The demand periods vary with the type of energy demand. The high fluctuating demand has a short-demand period, which can be as short as 5 min, but is generally 15, 30, or 60 min long (Monongahela Power Company, 2005a, 2005c; Buffington and Wolf, 2005). The utility companies use the period with the highest average demand for determining the demand charges in any billing cycle.

It should be noted that not all the utility companies charge their customers based on energy and demand. Also, there is no specific ratio or number of utility companies that

TABLE 28.1 Example Demand for a 15-Min Interval

Demand (kW)	Time Units for This Demand
10	2
12	4
2	7
10	2
Total	15

charge based on energy only and do not include demand in their bills. The calculation of the demand can be explained with a simple example. Assume that the demand pattern for any particular process is as given in Table 28.1. The average demand charged to the facility for this 15-min interval can be calculated as follows:

$$\text{Demand charged (kW)} = ((10 \times 2) + (12 \times 4) + (2 \times 7) + (10 \times 2)) / 15 \\ = \text{kW}$$

Similarly, the demand will be calculated for each of the 15-min intervals. Finally, the facility will be charged for the maximum of all these calculated values for each 15-min interval during the billing period. It may be noted that in some cases, the demand rate is also a variable charge based either on the time of the day or time of year (Industrial Rates, Utilities in Greater Cleveland, 2005; West Penn Power Company, 2005; Monongahela Power Company, 2005a, 2005b; Allegheny Power, 2005). Finally, the energy charge may be based on the demand (kW or kVA).

28.3.1.4 Rate Schedules We can compute an energy bill for a sample facility. The energy rate structure for the facility is as follows:

Minimum monthly charges	\$2.75/kVA of demand
<i>Energy charges</i>	
	\$0.10599 for first 30 kWh per kVA of demand
	Next 170 kWh per kVA of demand:
	\$0.06965/kWh for first 2,000 kWh
	\$0.05883/kWh for next 8,000 kWh
	\$0.04867/kWh for next 90,000 kWh
	\$0.04175/kWh for all over 100,000 kWh
	For kWh in excess of 200 times kVA:
	\$0.03341/kWh for first 200,000 kWh
	\$0.02855/kWh for all remaining kWh
Average fuel cost adjustment (FCA)	\$0.01902/kWh
Average tax on kWh	8.27% of the kWh cost

Note: The average FCA and tax rates may be calculated from the energy bills or can be obtained directly from the utility company.

28.3.1.5 Example of the Electricity Cost Calculation As an example, for one of the billed months, the readings are assumed to be

$$\text{Total energy} = 201,800 \text{ kWh} \\ \text{kVA recorded} = 416 \text{ kVA}$$

Therefore, the cost of electricity can be calculated as follows:

$$\begin{aligned}
 &\text{Minimum monthly charges} = \$2.75/\text{kVA} \times 416 \text{ kVA} = \$ \\
 &\text{Energy charges} \\
 &\quad \$0.10599 \text{ for first 30 kWh per kVA of} = \$1,322.76 \\
 &\quad \text{demand} = \$0.10599/\text{kWh} \times 30 \text{ kWh/kVA} \times 416 \text{ kVA} \\
 &\text{Next 170 kWh per kVA of demand (or 170 kWh/kVA} \times 416 \text{ kVA} = 70,720 \text{ kWh}) \\
 &\quad \text{First 2,000 kWh: } \$0.06965/\text{kWh} \times 2000 \text{ kWh} = \$139.30 \\
 &\quad (\text{Note: remaining kWh in this bracket is} \\
 &\quad \quad 70,720 - 2,000 = 68,720 \text{ kWh}) \\
 &\quad \text{Next 8,000 kWh: } \$0.05883/\text{kWh} \times 8,000 \text{ kWh} = \$470.64 \\
 &\quad (\text{Note: remaining kWh in this bracket is} \\
 &\quad \quad 68,720 - 8,000 = 60,720 \text{ kWh}) \\
 &\quad \text{Next 90,000 kWh: } \$0.04867/\text{kWh} \times 60,720 \text{ kWh} = \$2,955.24 \\
 &\quad (\text{Note that only 60,720 kWh was available in this bracket}) \\
 &\quad (\text{Note: remaining kWh in this bracket is } 60,720 - 60,720 = 0 \text{ kWh}) \\
 &\text{All over 100,000 kWh: } \$0.04175/\text{kWh} \times 0 \text{ kWh} = \$0.00 \\
 &\quad \text{For kWh in excess of 200 times kVA (201,800 kWh} - \\
 &\quad \quad (200 \text{ kWh/kVA} \times 416 \text{ kVA}) = 118,600 \text{ kWh}) \\
 &\quad (\text{Note that only 118,600 kWh is available in this bracket}): \\
 &\quad \text{First 200,000 kWh} = \$0.03341/\text{kWh} \times 118,600 \text{ kWh} = \$3,962.43 \\
 &\quad (\text{Note: remaining kWh in this bracket is } 118,600 - 118,600 = 0 \text{ kWh}) \\
 &\quad \text{All remaining kWh} = \$0.02855/\text{kWh} \times 0 \text{ kWh} = \$0.00 \\
 &\text{Sub-total cost} = \$1,322.76 + \$139.30 + \$470.64 \\
 &\quad + \$2,955.24 + \$0 + \$3,962.43 + \$0 \\
 &\quad = \$8,850.37 \\
 &\text{Average fuel cost adjustment (FCA)} = \$0.01902/\text{kWh} \times 201,800 \text{ kWh} = \$3,838.24 \\
 &\text{Average tax on kWh} = 8.27\% \text{ of } \$8,850.36 = \$731.92
 \end{aligned}$$

Therefore, the total cost (TC) for the month is calculated as

$$\begin{aligned}
 \text{TC} &= \text{Subtotal cost} + \text{FCA} + \text{Tax} \\
 &= \$8,850.37 + \$3,838.24 + \$731.91 \\
 &= \$13,420.53
 \end{aligned}$$

It may be noted that the FCA rate may change from month to month. The same calculation method will be used to calculate the cost savings from any recommendation. The energy savings will be calculated in terms of kWh and kVA (derived from the kW savings as ratio of kW to the average power factor of 0.9788).

28.3.1.6 Example of Energy Cost Savings Calculation The energy cost savings calculations are based on the average monthly value of energy and demand usage. As an example, for one of the manufacturing facilities, based on the energy bills for 12 months in 2005, the average energy used was 222,533 kWh and the average demand usage was 475 kVA. Based on the values in the bills, the average FCA rate is assumed as \$0.01902/kWh and the average tax will be taken as 8.27%. As discussed earlier, the total average cost of the electricity per month is calculated as \$14,869.65. To calculate the savings from any recommendation, the savings in the energy and demand usage will be calculated in terms of kWh/year and kW-month/year. As an example, if the proposed savings are

10,000 kWh/year and 50 kW-month/year (or a reduction of 4.26 kW in the peak demand in each month), the new average usage can be calculated as follows:

$$\begin{aligned}
 \text{Proposed average kWh} &= (\text{Current average} - \text{proposed savings})/\text{month} \\
 &= (222,533 \text{ kWh}) - [10,000 \text{ kWh/year}/12 \text{ months/year}] \\
 &= (222,533 - 833.33) \text{ kWh/month} \\
 &= 221,699.7 \text{ kWh/month} \\
 \text{Proposed average kVA} &= (\text{Current average} - \text{proposed savings})/\text{month} \\
 &= (475 \text{ kVA}) - [50 \text{ kW/year}/(12 \text{ months/year} \times \text{PF})] \\
 &= (475 - 4.26) \text{ kVA/month} \\
 &= 470.74 \text{ kVA/month}
 \end{aligned}$$

(Note: PF is the average power factor is used to convert the kW to kVA.)

Based on the proposed average kWh and kVA and the method of energy cost calculation discussed in cost calculation section, the new cost of electricity is calculated as \$14,801.66. Therefore, the savings can be calculated as

$$\begin{aligned}
 \text{Savings} &= (\text{current cost} - \text{proposed cost}) \times 12 \text{ months/year} \\
 &= (14,869.65 - 14,801.66) \times 12 \\
 &= \$816/\text{year}
 \end{aligned}$$

28.3.1.7 Components in Natural Gas Bills The utility bills for the natural gas are rather simple as compared to the electrical energy bills. In general, the units for the measurement of natural gas consumption are based on volume or energy. Some common units in these categories are

Volume

Cf: Cubic foot $\cong 1030$ Btu

Ccf: 100 cubic feet

Mcf: 100 cubic feet

Tcf: Trillion cubic feet $\cong 1$ quad.

Energy

DTH: One deca-therm = 10 TH $\cong 1$ MMBtu

TH: One therm $\cong 100,000$ Btu

MMBtu: 1 million Btu

MBtu: 1000 Btu.

The total cost of the natural gas for a billing period may have one or more components and depends on the supplier and amount of natural gas usage in the facility. Some common billing terms used in the natural gas bills are

- *Gas Usage Rate*: usage rate determined by the utilities (\$/unit of gas),
- *Standby Charges*: charges to maintain a constant supply of gas to the facility (\$),
- *Customer Charge*: covers the cost of billing, meter reading, and equipment (\$),

Account number	Previous balance	Payment received	Current billing	Total due	Tariff MCF	Trans- port MCF	Total MCF
	2,005.83	2,005.83	1,835.00	1,835.00	4.00	0.00	4.00
Account no.:			Invoice detail			Billing month: APR 05	
Contract	Description			Basis		Dollars	
	Commodity charge			4.00	MCF	31.31	
	Customer charge			1.00	MON	18.38	
	Delivery charge			4.00	MCF	12.86	
	Gas cost adjustment			4.00	MCF	5.58	
	Standby service			1.00	MON	1,661.82	
	Tax surcharge			1,729.95	DOL	1.21	
	Transition cost			4.00	MCF	0.00	
	Tax surcharge			4.00	MCF	0.03-	
	Pa sales tax—100% taxable			1,731.13	DOL	103.87	
	Total current bill			4.00	MCF	1,835.00	
Includes	0.00 of late fees in previous balance					0.00	
	Late payment charges					0.00	
	Total due					1,835.00	

FIGURE 28.1 Example of natural gas bill with high standby charges.

- *Purchase Gas Adjustment (PGA)*: to recover the cost (fuel cost) of purchasing natural gas (\$/unit of gas),
- *Transportation Charges*: to recover the cost of transportation of gas to the customer site (\$/unit of gas),
- *Taxes*: applicable taxes on the total bill for the facility (\$/unit of gas).

An example of a natural gas bill is given in Figure 28.1. It may be noted that the standby charges for this facility is much more than the usage cost of the gas. This facility was using the gas company as a standby to their gas wells, from which they were getting the natural gas for their process use. It is evident that the usage pattern of the utilities should be considered properly to minimize the unnecessary charges.

28.3.2 Analysis of IAC Database

The IAC perform industrial assessments for small and medium-sized manufacturing companies to identify opportunities to save energy, reduce waste, and increase overall plant productivity. The program maintains a database of all the recommendations from assessments performed by the IAC and currently has a listing of 12,987 assessments and 97,205 recommendations (IAC Database, 2006). This database is continuously updated after any of the centers upload the report from their assessment. The database has many features to find any recommendation that the user may be interested in. The assessments/recommendations can be filtered based on manufacturer codes (NAICS or SIC), IAC, year of energy assessment, annual sales of the facilities, number of employees, annual energy cost, and state of facilities.

28.3.2.1 Recommendations To facilitate the search for a particular recommendation, the database has assigned special codes, called as Assessment Recommendation Code

(ARC), for each of the recommendations. The ARC contains six numbers with a format as X.YYYY.Z. The first number is designated for the recommendation type, the next four numbers detail the actual recommendation within the main category, and the last number is the application of the recommendation (e.g., building, process, or other applications). All the recommendations are mainly categorized as energy management, waste minimization/pollution prevention, and direct productivity enhancements. Each of these categories has lists of subcategories based on the energy systems. The general categories of these recommendations are as follows.

28.3.2.2 Energy Management The recommendations related to energy management deal with combustion systems, thermal systems, electrical power, motor systems, industrial design, maintenance and equipment control, lighting, space conditioning, administrative and other costs, and alternative energy sources such as wind and solar energy.

28.3.2.3 Waste Minimization/Pollution Prevention This includes the recommendations related to reduction of waste and therefore pollution through operations, equipment, maintenance, raw materials use, water use, and disposal.

28.3.2.4 Direct Productivity Enhancements The example recommendations in this category are manufacturing enhancement, purchasing, resource optimization, reduction of downtime, management practice, and other administrative savings. The top five recommendations in the list are shown in Table 28.2.

28.3.2.5 Implementation The IAC program keeps a track of the recommendations implemented in the facilities, which are called 6–9 months after the assessment date. This information is reported to the IAC database and is accessible to the public. The program has an excellent implementation record of more than 50% implementation rate for its recommendations. Table 28.3 lists the top five recommendations that were implemented

TABLE 28.2 Top Five Recommendations

Recommendation	Number of Times Recommended
Utilize higher-efficiency, lower-wattage lamps, or ballasts	8189
Eliminate leaks in inert gas and compressed air lines	4828
Use most efficient type of electric motors	4372
Install compressor intakes in coolest locations	3860
Utilize energy-efficient belts and other improved mechanisms	3227

TABLE 28.3 Top Five Implemented Recommendations

Recommendation	Number of Times Implemented
Utilize higher-efficiency, lower-wattage lamps, or ballasts	4700
Eliminate leaks in inert gas and compressed air lines	4020
Use most efficient type of electric motors	2877
Install compressor intakes in coolest locations	1911
Utilize energy-efficient belts and other improved mechanisms	1815

the most by the manufacturing facilities. It may be noted that the same recommendations that were recommended the most have been implemented the most as well.

28.3.2.6 Energy Savings Based on the analysis of the database, the total amount of energy savings from all the recommendations in terms of electricity is over 2 billion kWh and natural gas is over 38 million MMBtu. It is worth mentioning that the database has the records for many other fuel sources such as #2 oil, coal, and wood, but have not been considered here because electricity and natural gas are main energy sources in most of the industries visited through the IAC program.

28.3.2.7 Cost Savings The IAC program has recommended savings over \$1.5 billion. The average savings per assessment is estimated as \$55,000. The database lists the energy savings for each recommendation in terms of the utilities (electricity, natural gas, oil, coal, water, etc.) related to the process. The total annual cost savings from recommendations related to electrical energy exceed \$120 million, and natural gas savings recommendations result in an annual savings of more than \$180 million. Top 10 recommendations and their cost savings are shown in Table 28.4.

28.3.2.8 Payback on Investment Based on the total cost savings and the implementation cost of the recommendations, the average payback on the implementation cost for all the recommendations is estimated as 4 months. It may be noted that some of the recommendations do not require any implementation cost and thus result in an immediate payback, whereas others may have some implementation cost in terms of capital or labor cost investment. For most of the recommendations, the average payback is within 2 years.

28.3.2.9 Reduction in Emissions The recommendations during the assessments not only save money for the facilities but result in reduction in emissions. Based on the analysis performed in the energy savings section, the reduction in CO₂ emissions through electricity and natural gas recommendations are estimated as approximately 9 billion pounds.

TABLE 28.4 Top 10 Cost Savings Recommendations

Recommendation	Total Savings (\$)
Add equipment/operators to reduce production bottleneck	88,846,940
Replace existing equipment with more suitable substitutes	44,651,442
Install automatic packing equipment	40,672,093
Use a fossil fuel engine to cogenerate electricity or motive power and utilize heat	37,876,660
Utilize higher-efficiency lamps and/or ballasts	35,925,118
Use waste heat to produce steam to drive a steam turbine generator	24,659,543
Eliminate leaks in inert gas and compressed air lines/valves	24,504,452
Replace electrically operated equipment with fossil fuel equipment	23,947,776
Change procedures/equipment/operating conditions	23,336,815
Repair and eliminate steam leaks	22,959,756

28.4 ENERGY EFFICIENCY MEASURES IN MAJOR ENERGY CONSUMING EQUIPMENT

This section discusses a sample of the important energy-saving opportunities for major energy-consuming equipment commonly used in facilities (e.g., boilers, furnaces, ovens, electric motors, lighting, chillers, cooling towers, heating ventilation and air-conditioning, air compressors, and combined heat and power systems). Description for assessing equipment, data collection requirements, related theory, necessary modifications in the equipment setup, implementation cost, and benefits of implementation are discussed further. U.S. DOE's BestPractices tools can be used to determine the energy savings in several areas of equipment usage.

28.4.1 Furnaces, Ovens, Boilers, and Steam Systems

Furnaces and ovens are equipment used to heat air, material, or other fluids. Common applications can be melting, annealing, preheating, heat treatment, and baking. Boilers produce hot water or steam under pressure in a closed vessel used in various processes.³ Furnaces/ovens/boilers burn different types of fuel to generate heat, which is then used for the desired purpose. Following are some of the energy-saving opportunities that can be realized for furnaces/ovens/boilers.

28.4.1.1 Insulate the Bare Surfaces of Ovens, Furnaces, Boilers, and Steam Pipes The bare surfaces of crucible furnaces, reverberatory furnaces, electric ovens, pit ovens, and boiler and steam pipes, if not insulated, will radiate significant energy due to lack of insulation. These surfaces can be insulated, thereby reducing the heat losses. Energy savings is a function of the difference in heat loss from the bare and insulated surface, area of the surface that needs insulation, annual operating hours of the equipment, and the efficiency of the heat supply. The DOE BestPractices tool 3E Plus can be used effectively for this energy efficiency measure (BestPractices Software Tools, 2006).

28.4.1.2 Preheat Natural Gas Oven/Furnace/Boiler Combustion Air Using Hot Flue Gas Significant amount of waste heat is available in the oven/furnace/boiler stack. This heat can be recovered by preheating the combustion air. A heat exchanger along with the necessary ductwork can be installed to preheat the intake air of the natural gas oven/furnace/boiler. The combustion air can be preheated and directed into the natural gas-fired oven/furnace/boiler air intake port using galvanized steel insulated ductwork. Warmer combustion air leads to increase in combustion efficiency. An increase in efficiency of approximately 1% is possible for every 40°F increase in combustion air temperature (Dyer and Dyer, 1981). The existing combustion efficiency of the oven/furnace/boiler can be determined by using a stack gas analyzer. The energy savings to be realized by directing preheated air to the oven/furnace/boiler air intake is a function of the rating of the oven/furnace/boiler, the annual operating hours, and the existing and proposed combustion efficiencies of the oven/furnace/boiler (Industrial Assessment Center Reports, 1992–2006). The USDOE's BestPractices software tool PHAST can be used to determine the improvement in efficiency by preheating the combustion air (BestPractices Software Tools, 2006).

³ Wikipedia. The free encyclopedia. Available at <http://en.wikipedia.org/wiki/Motors>.

28.4.1.3 Recover Waste Heat from Gas Ovens/Furnaces/Boilers Air-to-air type heat exchangers can be used to capture heat from the exhaust gases from the gas oven/furnace/boiler to heat the plant area during winter. By recovering the waste heat from the oven/furnace/boiler stack, the plant space heating system will be used for less time. The recovered heat would be used to heat the plant area during the heating season. The outlet temperature for the hot side of the gas oven/furnace/boiler can be determined using a stack gas analyzer or an infrared temperature gun in case the stack is not insulated. Outside fresh air or plant air can be used on the cold side of the heat exchanger. The potential annual space heating energy savings is a function of the exhaust gases' mass flow rate from the stack, specific heat of exhaust gases, temperature of exhaust gases from stack before heat exchanger, minimum practical exhaust temperature after heat exchanger (Dyer and Dyer, 1981), operating hours during the heating season, and efficiency of the heating system (Industrial Assessment Center Reports, 1992–2006).

28.4.1.4 Adjust Air-Fuel Ratio on the Natural Gas Oven/Furnace/Boiler Adjust the combustion system air-fuel ratio for the oven/furnace/boiler to reduce the amount of excess air passing through the oven/furnace/boiler and to improve the combustion efficiency of the oven/furnace/boiler. Based on the field combustion test conducted on the oven/furnace/boiler, it can be determined whether they need air-fuel ratio adjustment. Combustion test on the oven/furnace/boiler can be done using a stack gas analyzer. The optimum amount of O₂ in the flue gas of a natural gas-fired oven/furnace/boiler is 2.2%, corresponding to 10% excess air. In practice, a reduction in excess O₂ to about 3.5% (about 18% excess air) can be achieved. This estimate is based on discussions with several local oven/furnace/boiler adjustment contractors, who have indicated that a reduction in excess O₂ to less than about 3% is difficult for most oven/furnace/boiler. The corresponding CO₂ value for an O₂ value of 3.5% should be about 9.8%, as seen in Table 28.5, assuming no change in stack gas temperature before and after the adjustment. Combustion efficiency will increase due to the adjustments on the air-fuel ratio (Industrial Assessment Center Reports, 1992–2006).

The energy savings is a function of the rating of the oven/furnace/boiler, the annual operating hours of the oven/furnace/boiler, the firing factor of the oven/furnace/boiler, and the existing and proposed combustion efficiencies of the oven/furnace/boiler (Industrial Assessment Center Reports, 1992–2006). The USDOE's BestPractices tool PHAST can be used to determine the improvement in efficiency by adjusting the air-fuel ratio (BestPractices Software Tools, 2006).

28.4.1.5 Inspect, Repair, and Maintain Steam Traps Steam traps are used to separate the steam from condensate, air, and other noncondensable gases (Industrial Assessment Center Reports, 1992–2006). Institute a permanent steam trap management program. The program should include inspecting the steam traps, cleaning the traps, replacing the trap

TABLE 28.5 Optimal Percentages of O₂, CO₂, and Excess Air in the Exhaust Gases

Fuel	O ₂ (%)	CO ₂ (%)	Excess Air (%)
Natural gas	2.2	10.5	10
Liquid petroleum fuel	4.0	12.5	20
Coal	4.5	14.5	25
Wood	5.0	15.5	30

disc or lever if not working properly, and replacing the trap itself if it is defective. This program could be implemented by the plant maintenance staff. Steam traps are the key to an efficient steam system. The objectives of the steam traps are to remove condensate, air, and other noncondensable gases, and prevent steam loss. Efficient operation of any steam system requires well-designed trapping, which is periodically inspected and properly maintained. It is only in this way that condensate and air will be removed automatically as fast as they accumulate without wasting steam (Energy Conservation Program Guide for Industry and Commerce, 1974). The traps can be tested using an ultrasonic testing device. Common outcomes of steam trap testing are continuous blowing, partial or complete blockage, or the trap functions properly. Energy savings is thus a function of the total steam loss from steam traps, heat content of saturated steam, specific heat of water at constant pressure, feed water temperature, boiler overall efficiency, and annual operating hours for which the steam source operates (Industrial Assessment Center Reports, 1992–2006). The USDOE's BestPractices tool SSAT can be used to determine the losses through steam traps (BestPractices Software Tools, 2006). The inputs to the tool are mainly the steam pressure, total number of steam traps in use, number of steam trap failures, boiler fuel source, and the number of steam trap failures after repairs. Based on these inputs to the tool, savings in fuel and makeup water cost can be determined.

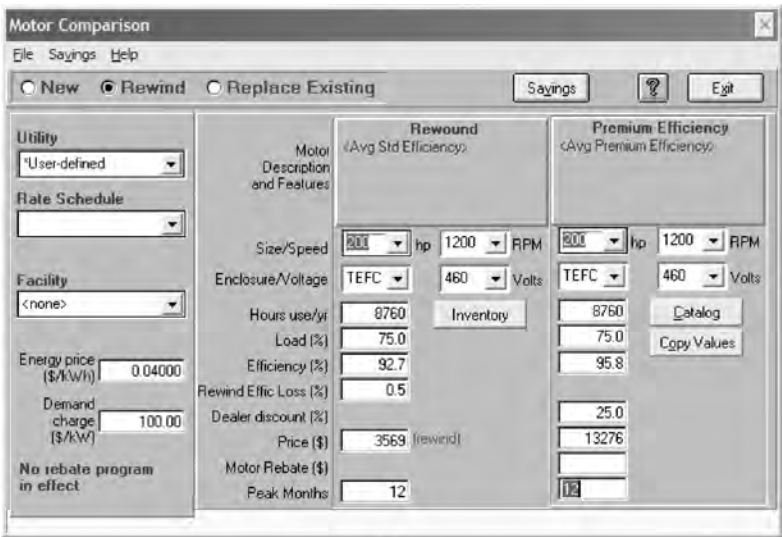
28.4.1.6 Other Steam-Related Energy-Efficiency Measures The other steam-related energy-efficiency measures can be listed as follows: (1) install reverse osmosis unit to reduce boiler blow-down rate; (2) inspect and repair steam leaks; (3) implement a condensate recovery system; (4) install waste heat boiler to produce steam; (5) install deaerator unit to reduce steam system corrosion; and (6) install turbulators on fire tube boilers (BestPractices Software Tools, 2006).

28.4.2 Motors

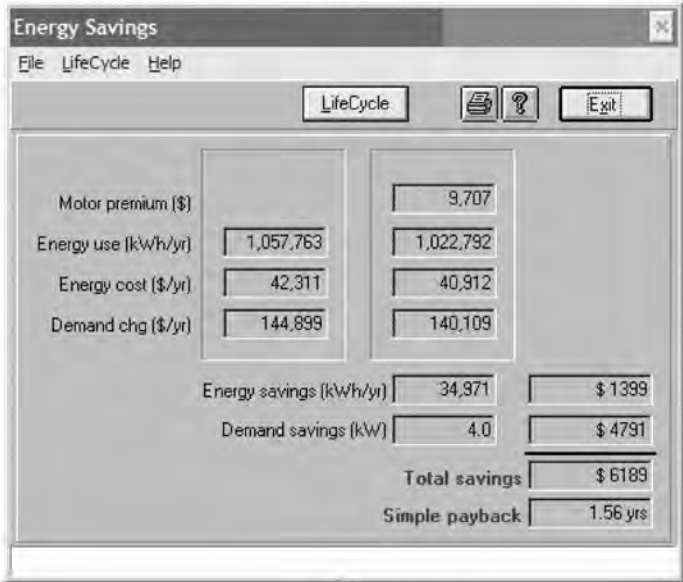
Electric motors are used to convert electrical energy into mechanical work (Industrial Assessment Center Reports, 1992–2006). There is a drop in motor efficiency with its use. The following energy-saving opportunities discuss ways to reduce transmission losses and improve motor efficiency.

28.4.2.1 Implement a Motor Management System Implement a motor management system (MMS) such as MotorMasterTM to help document motor inventory and to identify/analyze motor-driven systems for various energy conservation opportunities. Implementing the motor management system requires the facility to first obtain MotorMasterTM software, which can be downloaded at no charge from the U.S. Department of Energy (USDOE's) web site (BestPractices Software Tools, 2006).

After obtaining the software, the user can develop an inventory of motors organized by facility, department, and process. Information such as nameplate data, load profile, field measurements, and energy cost can be stored for each motor so that motor performance can be tracked. A motor management system is designed to assist the facility to reduce energy costs through maximizing production efficiency, minimizing energy consumption, correcting for power factor, understanding utility billing statements, and establishing a preventative and predictive maintenance program. MotorMasterTM software (Figure 28.2) is capable of analyzing the energy and cost impacts of various hypothetical situations that occur either before and/or after a motor failure. These situations include downsizing to a



(a)



(b)

FIGURE 28.2 U.S. DOE’s BestPractices Tool MotorMaster™ for determining energy savings through use of energy-efficient motor.

smaller motor, rewinding a failed motor, replacing a failed motor with same size motor, and replacing a failed motor with a larger spare motor.

28.4.2.2 Other Motor-Related Energy-Efficiency Measures The other important energy-efficiency measures as related to motors are: (1) perform vibration analysis on equipment; (2) replace drive belts on motors with energy-efficient cog belts; (3) maintain properly sized motor system; and (4) install variable-speed drive on the motors.

28.4.3 Lighting

Lighting systems use considerable amount of electrical energy to produce desired illumination levels.³ Lighting technology has made significant developments with time and provides state of the art energy solutions (Industrial Assessment Center Reports, 1992–2006).

28.4.3.1 Replace the 400 W Metal Halide Lamps with 360 W Metal Halide Lamps The lighting energy usage can be reduced by replacing the 400 W metal halide lamps with 360 W metal halide lamps. Because of more efficient 360 W bulbs, the illumination level practically remains the same with the reduction in power used. This is an option for reducing the amount of energy used for lighting. The replacement does not require the change of the fixture and the ballast. The lamps can be retrofitted with 360 W metal halide lamp, as the existing 400 W lamps burn out with time. Energy savings is a function of the difference in power used by the 400 W and 360 W lamps, the ballast factor of the lamp, the annual operating hours of the lamps, and number of similar lamps to be replaced in the facility (Industrial Assessment Center Reports, 1992–2006).

28.4.3.2 Replace the Existing T12 Lighting Fixtures with Magnetic Ballasts by T8/T5 Lighting Fixtures with Electronic Ballasts and Reflectors Install T8/T5 lighting fixtures using electronic ballasts with reflectors in place of the T12 fixtures containing magnetic ballasts. This will reduce the lighting energy usage while maintaining the same lighting levels at the work surface. Lighting technology has evolved rapidly in recent years. In commercial buildings, significant reductions in energy use can be achieved by installing energy-efficient bulbs, fixtures, and controls. Retrofits to install new technologies such as electronic ballasts and specular reflectors are often cost-effective, providing a reasonable payback. T8/T5 bulbs provide almost the same light levels as T12 bulbs with the reduction in power consumption. Electronic ballasts consume less (approximately 5–10%) power as compared to the regular magnetic ballasts (Industrial Assessment Center Reports, 1992–2006). Hence, the energy savings is a function of the number of bulbs per fixture, the number of florescent fixtures in the facility, existing and proposed wattage of each fixture, the ballast factor, and the annual operating hours (Industrial Assessment Center Reports, 1992–2006).

28.4.3.3 Install Occupancy Sensors in Designated Areas Occupancy sensors can be installed in some areas to reduce the electrical usage for lighting during unoccupied periods. If certain areas in the facility with less utility are heavily lit, then by wiring occupancy sensors into these areas, the lighting usage could be reduced during the unoccupied periods. Energy savings will result from reduced electrical usage for lighting. Some lights can be left on in the room for emergency purposes. Energy savings is a function of the total power consumption by the lights and the annual hours for which the lights can be controlled (Industrial Assessment Center Reports, 1992–2006).

28.4.4 Chillers

Chillers are used to produce chilled water for plant and process cooling. Mechanical chillers and absorption chillers are the two commonly used types.³ Mechanical chillers use electric motors to drive the refrigerant compressors, whereas heat source is used in absorption chillers.

Increasing the chiller set point temperature would save electrical energy. Generally, the efficiency of chillers increases as the chilled water temperature increases. This is because in order to obtain lower-temperature chilled water, the refrigerant must be compressed at a higher rate, which in turn increases the compressor power requirements and decreases the efficiency of the chiller. There is approximately 1% increase in efficiency for each degree Fahrenheit increase in the chilled water set point temperature (Modern Industrial Assessments, 2004). Before implementing this recommendation, it is advised to first check any effect of changing the set point temperature on the process. Thus, the energy savings can be realized by increasing the chiller set point temperature and is the function of the chiller capacity, increase in chiller efficiency due to increased set point temperature, chiller coefficient of performance, annual operating hours for the chiller, and operating load on the chiller (Industrial Assessment Center Reports, 1992–2006). The USDOE's BestPractices software tool CWSAT can be used to determine the energy savings from this recommendation (BestPractices Software Tools, 2006).

28.4.5 Heating, Ventilation, and Air-Conditioning Equipment

Heating, ventilation, and air-conditioning (HVAC) units are used to condition air (i.e., control temperature, humidity, and air flow as per the space requirements). Heating is provided to heat the air in the heating season, whereas air-conditioning is used to cool the air in the cooling season. Ventilation involves air makeup units to maintain positive air pressure in the plant.

Older air-conditioning units used for cooling were not made with installed economizers. Space cooling energy usage can be reduced by adding an economizer. Economizers are essentially a duct and damper system that allows fresh outside air to be used directly for space cooling whenever outdoor air has a lower total heat content, *enthalpy*, than return air. Using cool outside air whenever possible, the energy usage by the mechanical cooling units can be reduced significantly. To determine possible energy savings, the enthalpy of outside air should be known. The enthalpy is a function of the dry bulb temperature and the humidity. The higher the dry bulb temperatures and humidity levels, the larger the enthalpy value is. Using the ASHRAE Handbook of fundamentals, enthalpy values can be determined directly from wet bulb temperatures. Using TMY data and a bin analysis, 5°F wet bulb temperatures can be determined with their associated monthly hours. The enthalpy can also be determined directly from a psychometric chart. In order to use this, the user needs the dry bulb temperature and the humidity or both the wet bulb temperature and the dry bulb temperature. Energy savings is a function of the total Btu \times h per lbm of air for a month during the cooling period, total cubic feet per minute of the cooling capacity, density of air, coefficient of performance of the air-conditioning unit, and the annual usage of the unit (Industrial Assessment Center Reports, 1992–2006).

28.4.6 Air Compressors

Air compressors are electric motor driven systems used to increase the air pressure by reducing its volume. Compressed air finds wide range of application in industries.

28.4.6.1 Use Outside Air for Air Compressor Intake Duct outside air to the air compressor intake. This can be done by connecting one end of the pipe to the compressor intake. The other end is routed through the wall to the outside. The existing intake

temperature at the compressor intake can be measured using an infrared temperature gun. On average, the outside air is cooler and therefore denser than the indoor air. The compressor's work for the usual operating conditions in manufacturing plants is proportional to the absolute temperature of the intake air. Being denser, cooler air has higher mass and thus compresses more mass of air in one compression cycle (BestPractices 2006). Therefore, by utilizing the outside air as supply to the air compressor intake, it can be compressed easily with reduced energy requirement. Energy savings is a function of the compressor motor capacity (horse power rating); fractional reduction in the compressor's work is due to lowering the compressor's intake air temperature; the horsepower reduction factor, based on operating pressure and power consumption at maximum pressure; operating load on the compressor; annual operating hours; and efficiency of the compressor motor (Industrial Assessment Center Reports, 1992–2006).

28.4.6.2 Repair Compressed Air Leaks Repair compressed air leaks on a regular basis. Air leaks in the facility can be detected using an ultrasonic leak detector. The leak detector detects noises caused by air leak that are not audible to human ears in a noisy production setup. Output from the detector can be read from a visual analog (or digital) display or can be heard using earphones. Air is compressed from atmospheric pressure to the compressor discharge pressure at the expenditure of compressor work. Considerable volume of high-pressure air is lost through air leaks. At discharge pressure conditions, power is lost due to loss of compressed air volume through air leaks. Repairing air leaks result in reduced compressed air wastage and thus reduced artificial demand on the compressor. Air leaks are typically present near pipe joints, fittings, valves, and regulators. Energy savings is a function of the power loss due to leak, which is a function of the volumetric flow rate of free air exiting the leak and annual hours during which the leak occurs (Industrial Assessment Center Reports, 1992–2006). The USDOE's BestPractices tool AIR-Master+ can be used to determine the amount of energy lost through air leaks (BestPractices Software Tools, 2006).

28.4.7 Combined Heat and Power (CHP)

Combined heat and power systems generate heat and electricity as against traditional power plants only producing electricity. A CHP system typically involves a compressor, combustion chamber, gas turbine, and mechanical equipment or a generator coupled to the turbine shaft. The exhaust gases leave the turbine at an elevated temperature. Considerable heat is lost from the turbine exhaust. This heat energy is recovered and serves as a heat source for process heating, space heating, and boiler feedwater heating (Oland, 2004).

A back pressure turbine can be installed at the boiler house to use the potential pressure difference between the turbine's inlet and exit steam as a prime mover or power source to any equipment using electricity. The high pressure steam produced by the boiler is often throttled at reducing stations to a lower pressure before it is used in the production process. There exists heat energy potential in this steam, which can be used by replacing the reducing station with a back pressure turbine. Therefore, the back pressure turbine will use the energy between the high-pressure steam at boiler operating conditions and the low-pressure steam being supplied to the plant. The back pressure turbine can, therefore, serve as a prime mover to any rotating equipment such as an electric motor. Also, a generator can be coupled to the turbine shaft to generate electricity, which can serve as a source of power for any equipment (electric motors) using electricity. The energy savings

is a function of the mass flow rate of the steam, difference in enthalpies of the high pressure and low pressure steam, overall efficiency of the back pressure turbine, and the efficiency of the mechanical coupling between the back pressure turbine and the rotating equipment (Industrial Assessment Center Reports, 1992–2006).

28.5 CASE STUDIES OF DEVELOPMENT OF ENERGY-EFFICIENCY MEASURES

Various types of industries have implemented energy-efficiency measures to improve efficiency of their existing operations. This is especially true for manufacturing industries that have high energy intensity in their operations. These include manufacturing plants in glass manufacturing, plastics manufacturing, primary metals manufacturing, and chemical manufacturing industries. Some of the case studies that reveal typical energy-saving opportunities that existed in this industry segment and savings realized as a result of a particular measure are discussed here.

28.5.1 Glass Manufacturing

The manufacturing activities carried out in the glass-manufacturing facilities range from glass decorating using purchased glass to manufacturing glass from basic raw materials using the hand-blown or machine-based processes. Four factors that characterize the manufacturing facilities are energy consumption, the plant size, the number of employees, and the annual sales (Gopalakrishnan et al., 2001b). The industrial assessments conducted at the glass-manufacturing facilities revealed a variety of Energy Conservation Opportunities (ECOs). The manufacturing variety for glass was as follows: hand-blown glass, machine-made glass, pressed glass, flat glass, glass bending, cut glass, glass fabric production, window manufacturing, decorative glass manufacture, hermetic seal production, and marbles. Significant relationship exists between the annual energy usage and the plant size (Gopalakrishnan et al., 2001b). A list of ECOs was recommended for each plant assessed. The energy conserved was reported for each ECO. The ECOs were selected for each plant based on the percentage of energy savings they contributed with respect to the total energy savings recommended for that plant. In general, for each plant, ECOs that had energy savings of 20% or higher with respect to total recommended energy savings were selected for this list, as they were considered to be the most significant. The ECOs found in the glass-manufacturing facilities were adjusting air to fuel ratios, preheating combustion intake air, heat recovery and distribution, insulating hot surfaces, replacing motors with energy efficiency motors, modifying chiller operation, modifying heat output in process, improving compressor operation, and reducing infiltration and space-heating loads. A total of 280,000 MMBtu of energy savings were recommended at the 32 glass-manufacturing facilities as a result of these energy-efficiency measures.

28.5.2 Plastics Manufacturing

Energy assessment at a leading plastics manufacturer done by Industrial Assessment Center at West Virginia University has helped the company save nearly \$100,000 per year in energy costs. The company is one of the largest plastic-packaging specialists in Europe and is expanding into the U.S. market. The plant manufactures injection-molded, rigid

containers with open tops. The facility measures 187,000 square feet in size and operates continuously, 7 days per week. Energy costs before the implementation of energy-efficiency measures at the plant totaled approximately \$760,000 per year, most of which was for electricity and the remainder for natural gas. Of the total 12 opportunities found feasible after study by assessment team, 6 were implemented at the manufacturing plant within first year. The implemented energy efficiency measures are briefly listed here. It was found that some bare surfaces of molding machines were radiating significant energy due to lack of insulation. However, by insulating these surfaces, the machines' heaters will operate less frequently, which will reduce energy consumption. USDOE-developed MotorMaster+ software was suggested to be used for motor management system. This software helped the company in evaluating economic decisions about buying new premium efficiency motors and rewinding or replacing a motor. Compressed air is used significantly in the plant. To help the company improve compressed air system energy efficiency, the assessment team suggested that the company reduce the compressor pressure to a safe minimum required by the system, repair compressed air leaks, and install engineered nozzles to reduce air consumption. As a result of the energy efficiency study and recommendations made by assessment team, the company is able to save 2.3 million kWh or 7950 MMBtu annually. This translates into significant amount of energy costs savings of nearly \$100,000 (Department of Energy, 2005c).

28.5.3 Metals Manufacturing

An energy-efficiency study carried out at a leading steel manufacturer in West Virginia has helped the company survive the market slowdown. The company recycles scrap steel from adjacent areas to produce steel sections, channels, and structural beams that are used in fabrication of frames, grousers, and welded frames for heavy equipments. The plant is approximately 40 acres in size and has been in operation since 1906. The plant operates approximately 8,160 h annually. The energy consumption of the plant before energy assessment was 1,216,753 MMBtu/year, or \$10,029,682/year. The assessment team found the following feasible energy-efficiency measures at the time of assessment: "ladling preheat energy reduction, hand-held cutting flame torches energy reduction, reduce infiltration in doors on furnace openings, preheat combustion air in ceramic recuperator, motor management, downsize motors, variable speed drives in several applications, cogged belt drives, improve power factor, insulate the furnaces and tundish dryer, synthetic lubricants, repair air leaks, outside air, lighting efficiency." The assessment recommendations made by assessment team helped the company save an estimated 139,843 MMBtu/year. The total energy cost savings were approximately \$1,034,547/year (Gopalakrishnan and Plummer, 2003).

28.5.4 Chemicals Manufacturing

One of the leading chemical manufacturers in the world undertook a plant-wide energy assessment with West Virginia University. The objectives were to identify energy-saving projects in the company's utilities area, including boilers and associated steam systems, compressor systems, and electrical motor-driven pump systems. The project evaluation process was unique in that the company has obtained very favorable rates for electricity. Even so, the company found strong economic justification for several projects that would reduce either electricity consumption or fossil fuel consumption. The plant produces more than 1.2 billion pounds of chemical products each year. These materials are used in cars,

appliances, furniture, home construction, steel manufacturing, food preparation, and many consumer products. The group operates around 350 companies worldwide and employs 117,000 workers. The energy-efficiency measures that were determined were replacing burner, condensate return, using VSDs on pumps, optimizing compressed air system, insulating steam system components, repairing steam leaks, adjusting the fuel/air ratio in one of the boilers, and repairing compressed gas leaks (air, NG, and nitrogen). The total cost savings were identified as 1,427,800 with an estimated implementation cost of \$937,200, which results in a simple payback of less than 8 months (Department of Energy, 2003).

28.6 CONCLUSION

Industrial energy efficiency is necessary for several reasons. The energy efficiency will reduce operating costs and hence lead to manufacturing industry becoming globally competitive and contribute to increased profitability and growth in terms of jobs and market share. Energy efficiency will make the nation more secure and economically strong as it reduces the consumption of fossil fuels, hence playing a role in controlling the cost of energy to have a positive impact on every facet of the society. Energy efficiency will also enable the protection of the environment in terms of reduction in emissions. Pure and simply put, energy conservation makes sense.

Manufacturing facilities have the goal of putting the product “out the door” on a daily basis to survive and be profitable and they often do not have enough time and resources to ascertain whether they are using the least amount of energy possible per unit of product manufactured. The programs of the U.S. DOE’s EERE and Industrial Technologies Program have gone a long way in helping industry and hence the nation to become energy efficient through a myriad of programs. Energy costs may or may not be a significant part of the product costs, hence not appearing in the top list of priorities of many companies. Rising fuel costs have changed some of that attitude. Aiding industry to save energy is a smart option that cannot simply be viewed as corporate welfare because the benefits are significant to the nation as a whole.

ACKNOWLEDGMENTS

The authors wish to thank the U.S. DOE’s Office of EERE and ITP program for supporting the Industrial Assessment Center (IAC) at West Virginia University since 1992. The development of this chapter by the director and the students of the IAC is a direct result of their experience gained at the IAC. We also thank the West Virginia Development Office’s Office of Energy Efficiency for providing support to conduct energy assessments within the state of West Virginia for manufacturing facilities that do not qualify for an IAC assessment.

REFERENCES

Allegheny Power. *Schedule “C”, General and commercial service*. Available at <http://www.allegheny-power.com/Tariffs/MD/Mdc.pdf>. Accessed November 22, 2005.

- Augusta newsprint: paper mill pursues five projects following plant-wide energy efficiency assessment*. Available at http://eereweb.ee.doe.gov/industry/bestpractices/pdfs/fp_cs_augusta_newsprint.pdf. 2003.
- BestPractices Software Tools. Available at <http://www1.eere.energy.gov/industry/bestpractices/software.html>. Accessed April, 2006.
- BestPractices. *Compressed air challenge*. Available at http://www.focusonenergy.com/data/common/dmsFilesStaging/B_GL_MKFS_CompressedAirBestPra.pdf. Accessed April, 2006.
- Brown M, Smid D, Matthews B, McKeon M, Numminen S. Increasing the implementation of energy-efficient technologies among U. S. manufacturers. *Energy Engineering* 1994;91(6):42–70.
- Buffington, Wolf. *Deregulation of electricity generation in Pennsylvania*. Available at <http://www.age.psu.edu/extension/factsheets/h/H77.pdf>. Accessed November 22, 2005.
- Capehart BL, Turner WC, Kennedy WJ. *Guide to Energy Management*, 2nd ed. Atlanta: Fairmont Press; 1997.
- Department of Energy. Plant-Wide Energy Assessment Case Study, “Bayer Polymers: Plant Identifies Numerous Projects Following Plant-Wide Energy-Efficiency Assessment”. DOE/GO-102003-1677, Aug, 2003.
- Department of Energy. Industrial technologies, program. “Energy use, loss, and opportunities analysis, U.S. manufacturing & mining”. Available at http://www.eere.energy.gov/industry/pdfs/energy_opps_analysis.pdf. Accessed December, 2004.
- Department of Energy. *BestPractices Program*. Available at http://www1.eere.energy.gov/industry/bestpractices/about_iac.html. Accessed September, 2005a
- Department of Energy. *Energy systems program*. Available at http://www.eere.energy.gov/industry/energy_systems/. Accessed December, 2005b.
- Department of Energy. Industrial Technologies Program Factsheet. Superfos Packaging: Plastics Manufacturer Saves \$100,000 by Implementing Industrial Energy Assessment Recommendations. DOE/GO-102005-2169, Sept, 2005c.
- Department of Energy. *Plant-wide assessments program (PWA)*. Available at http://www1.eere.energy.gov/industry/bestpractices/plant_wide_assessments.html. Accessed February, 2006a.
- Department of Energy. *Industries of future program*. Available at <http://www.eere.energy.gov/industry/technologies/industries.html>. Accessed April, 2006b.
- Energy Conservation Program Guide for Industry and Commerce. NBS Handbook 115, U.S. Department of Commerce/National Bureau of Standards, Sept, 1974.
- Francis DW, Towers MT, Browne TC. *Energy Cost Reduction in the Pulp and Paper Industry—An Energy Benchmarking Perspective*. Pulp and Paper Research Institute of Canada (Paprican), Sept, 2004.
- Fendt F. A highly successful holistic approach to a plant-wide energy management system. *Steam Digest*. Office of Industrial Technologies; 2001.
- Gopalakrishnan B, Plummer RW, Alkadi N. Analysis of energy conservation opportunities in glass manufacturing facilities. *Energy Engineering Journal* 2001a;98(6):27–49.
- Gopalakrishnan B, Plummer RW, Alkadi NM. Comparison of glass manufacturing facilities based on energy consumption and plant characteristics. *Journal of Energy and Development* 2001b;27(1):(Autumn).
- Gopalakrishnan B, Plummer RW, et al. *Industries of Future, “Energy Assessment Report—Steel of West Virginia”*. West Virginia University, Sept, 2003.
- Gopalakrishnan B, Mate A, Mardikar Y, Gupta D, Plummer R, Anderson B. Energy efficiency measures in the wood manufacturing industry. Proceedings of the 2005 ACEEE (American Council for an Energy Efficient Economy) Summer Study on Energy Efficiency in Industry on CD ROM. ISBN 0-918249-54-6, Jul; West Point, New York; 2005a.

- Gopalakrishnan B, Selvaraj R, Turton R, Plummer RW, Sukumar S. A systems approach to plant-wide energy assessment. *Energy Engineering Journal* 2005b;102(5):49–80.
- Hoshide R. Effective Energy Audits. *Energy Engineering* 1995;92(6):6–17.
- IAC Database. Available at <http://iac.rutgers.edu/database/>. Accessed April 17, 2006.
- Industrial Assessment Center Reports. West Virginia University; 1992–2006.
- Industrial Rates, Utilities in Greater Cleveland. Available at <http://www.cose.org/pdf/FactSheets/Utilities.pdf>. Accessed November 22, 2005.
- Maple D, Maxwell D. *Boiler Efficiency Improvement*. Auburn (AL): Boiler Efficiency Institute; 1981.
- Mate A. *Energy analysis and diagnostics in wood manufacturing industry [Master's thesis]*. West Virginia University; 2002.
- Meffert W. Energy assessments in iron foundries. *Energy Engineering* 1999;96(4):6–18
- Modern Industrial Assessments*. Rutgers University; 2004.
- Monongahela Power Company. *Schedule "B", General service*. Available at <http://www.alleghenypower.com/Tariffs/WV/Wvmontariffs/WVMPPRetailTariff.pdf>. Accessed November 22, 2005a.
- Monongahela Power Company. *Schedule "K," General power service rate*. Available at <http://www.alleghenypower.com/Tariffs/WV/Wvmontariffs/WVMPPRetailTariff.pdf>. Accessed November 22, 2005b.
- Monongahela Power Company. *Schedule LGS, large general service*. Available at <http://www.alleghenypower.com/Tariffs/WV/Wvmontariffs/WVMPlgs.pdf>. Accessed November 22, 2005c.
- National Association of Manufacturers. *Efficiency and innovation in U. S. manufacturing energy use*. Available at <http://www.nam.org/energyefficiency>. Accessed June, 2005.
- OIT Clearinghouse. *Vision: Results for Today. Leadership for Tomorrow*. Office of Industrial Technologies, Feb; 2001.
- Oland CB. *Guide to combined heat and power systems for boiler owners and operators*. Oakridge National Laboratory. Available at http://www1.eere.energy.gov/industry/bestpractices/pdfs/guide_chp_boiler.pdf. 2004.
- Role of energy efficiency: understand the value of energy system efficiency*. Available at http://www.eere.energy.gov/industry/energy_systems/pdfs/role.pdf. Accessed December, 2004.
- Spanner G, Brown D, Sullivan G, Riewer S. Impact evaluations of industrial energy conservation projects in the Pacific Northwest. *Energy Engineering* 1993;90(4):
- Sundarajan N. *Analysis of the trends in energy conservation studies of the IAC program [Master's thesis]*. West Virginia University; 1999.
- Thumann A. *Handbook of Energy Audits*. The Fairmont Press, Inc.; 1979.
- U.S. Department of Energy. *Instructions for Energy Auditors*. National Technical Information Service, Springfield, Vol. I, Virginia, Sept, 1978.
- Warfel C. An energy audit method for utilities and industry. *Energy Engineering* 1993;90(2):35–47
- Weber J. Celanese chemicals clear Lake plant energy projects assessment and implementation. *Steam Digest*. Office of Industrial Technologies; 2001.
- West Penn Power Company. *Schedule 20, General service*. Available at <http://energy.opp.psu.edu/eng/ElecUtil/ElecRate/APS/PA20.pdf>. Accessed November 22, 2005.
- Will CF. Factors that affect your plant power bill. Textile, Fiber and Film Industry Technical Conference. IEEE 1993 Annual, page 10/1-10/4. May 4–6, 1993.
- Yik FWH, Burnett J. An experience of energy auditing on a central air-conditioning plant in Hong Kong. *Energy Engineering* 1995;92(2):6–30.

29

INDUSTRIAL WASTE AUDITING

C. VISVANATHAN

- 29.1 Overview
 - 29.2 Waste-minimization programs
 - 29.3 Waste-minimization cycle
 - 29.4 Waste auditing
 - 29.4.1 Phase I: preparatory work for a waste audit
 - 29.4.2 Phase II: preassessment of target processes
 - 29.4.3 Phase III: assessment
 - 29.4.4 Phase IV: synthesis and preliminary analysis
 - 29.5 Conclusion
- Further readings

29.1 OVERVIEW

In the pursuit of *sustainable production and consumption*—as the true value of natural resources and nonrenewable energy sources are being globally perceived—wastes can no longer be viewed as substances that are spendable. Research shows that wastes traditionally discharged into natural bodies as unwanted substances still possess some economical value. What is useless in one context can be useful in other. Importantly, pollution problems can be significantly reduced if wastes can be reused and recycled instead of being discharged to natural bodies. There is a radical shift in the perception of opportunities with industrial waste, and the current tone is to *conserve and cultivate* rather than *deploy and deplete*.

In many countries, the manufacturing industry is one of the largest polluting sectors, and every year enormous effort and financial resources are spent worldwide to deal with industrial waste. Therefore, from an industry perspective, a global change in the environmental perception has a profound significance. Industrial processes, management, goals, and ethics

are under pressure from a rising environmental awareness. There is an ever-increasing demand to externalize the environmental cost of industrial activities. Gradually, a stage is set to internalize the environmental cost, not by marginalizing environmental concerns but, contrarily, by increased environmental stewardship of products and processes.

In response, the global industry is embracing proactive methods that principally focus on energy conservation and waste minimization by the application of cleaner production techniques. These techniques are progressively conceived to meet the goals of environmental and economical sustainability of industries in a more dependable way. Use of traditional *end-of-pipe* treatment of waste alone is progressively becoming inadequate to satisfy the tightening requisites of modern environmental legislations. Industrial processes are also impacted by progressive phasing out or banning of several chemicals that have been regularly used in industries. These include ozone-depleting substances (ODS) and persistent organic pollutants (POPs), for example. Several national governments have pledged in various global multilateral environmental agreements to gradually eliminate such substances.

Open or global market regimes are bringing additional complications on top of environmental needs by putting up a stiff pricing competition. Over and above, the questions of sustainability are becoming more pressing as several global institutional buyers are including environmental criteria (such as ISO 14000 certification or others) in the procurement specification. This is driving industries to reevaluate their activities and associated costs, which also include waste treatment cost that covers roughly 15–30% of the total operational cost. The most assertive way is to reduce energy consumption and waste generation in the first place, which improves the overall process efficiency. Several case studies proved that such proactive approaches reduce the overall production cost and environmental liabilities. Moreover, being *greener* is also helping companies to market their products with institutional buyers and attract wider public attention.

With increasing popularity and attention to proactive methods, structured methodologies are being developed to systematically explore, analyze, and implement energy conservation and waste minimization or cleaner production programs in industries. In many industries, such techniques are now being used as one of the management tools to monitor and control process efficiency and environmental liabilities.

Traditionally, environmental impacts from industry are mainly assessed based on the type, characteristic, and volume of waste that it generates. However, recent analyses show that higher use of energy create significant environmental impacts when environmental issues related to energy production are taken into account (typical example is GHG emission). Therefore, energy use and waste generation in industries have been recognized as interlinked systems in the way that the more energy is used, the more pollution is produced, and the more waste generated, the more energy is required. As such, this chapter will refer to both waste and energy, beginning with waste minimization.

29.2 WASTE-MINIMIZATION PROGRAMS

All manufacturing processes will require raw materials and energy to produce a product (or an intermediary) and will generate waste in some form. Each manufacturing plant is unique in the type, characteristics, and quantity of waste generation. In other words, the manufacture of specific products creates particular waste quantity and quality. Thus, it is difficult to make generalizations regarding waste.

Since manufacturing is one of the largest single polluting sectors, several driving factors are compelling the manufacturing industries to change their outlook about waste management. Six major factors are described:

1. *Changing Perception of Industrial Pollution*: There has been a tremendous rise in awareness about industrial pollution in general public and institutions. This has forced governments to take steps to control pollution from industries.
2. *Changing Legislations*: Environmental laws and regulations for industries are tougher, and implementation is more rigorous. As a result, waste treatment technologies now require a stringent level of efficiency to meet the discharge standards.
3. *Changing Waste Treatment and Discharge Costs*: As a result of tightening legislation and discharge standards, waste treatment cost is continuously increasing.
4. *Changing Availability and Cost of Raw Material and Energy*: Greater demand of raw material and energy has tightened supply, causing their price to rise. This will definitely affect the cost of the finished products.
5. *Changing Traditional Markets and Trade Barriers*: The concept of protected markets is giving way to more competitive global markets. This is forcing industry to increase efficiency and reduce raw material and energy consumption. Buyers in many developed countries are progressively incorporating environmental specification (ISO 14000 certification, eco-labeling, green productivity) in their procurement processes to screen companies.
6. *Changing International Commitments*: More and more national and local governments and institutions are committing to international bilateral and multilateral agreements (such as UN-mediated multilateral environmental agreements) to curb pollution by reducing and eliminating known harmful substances such as ozone-depleting substances and persistent organic pollutants. Many of these substances are heavily used in manufacturing. This pressures industries to change their manufacturing processes and design new products.

As a result of these driving factors, many industries are taking new perspectives and strategies in handling their waste management issues, and are trying to resolve them in a more sustainable way. The current trend of waste management is to balance proactive methods with traditional reactive methods. The concept of *reactive method* is about treating the waste once it is generated, also called *end-of-pipe treatment*, while the *proactive method* includes energy and waste minimization, waste recycling and reuse, and cleaner production that reduces end-of-pipe waste. In brief, the main advantages of energy conservation and waste minimization are as follows:

- Raw material consumption can be reduced, which in turn reduces the product cost.
- Energy consumption can be reduced, thereby reducing specific energy required for the product.
- Process efficiency can be improved, increasing product yield and quality.
- Waste generation can be reduced, thereby reducing waste treatment and disposal cost.
- Waste materials can be segregated, leading to containment of hazardous and toxic waste, which in turn can improve workers' health and safety.

- By-products can be recovered from waste. In addition to recycle and reuse of waste, waste heat recovery and waste exchange (with other industries) can generate additional income.
- Increased environmental stewardship can lead to higher attention from institutional buyers and marketing of products.
- Investor confidence can be increased.

There are also some known barriers to implementing waste minimization:

- Some waste minimization or cleaner production techniques may involve significant capital investment.
- There may be obvious risk involved in implementing new systems.
- There is a lack of proper manpower and expertise in appropriate technology.
- There is a lack of information and awareness, especially among small and medium scale industries.
- Often there is a hesitation to change traditional ways of doing things.

The response of an industry will largely depend on several factors:

- nature of the industrial process,
- size and structure of the firm,
- technology and information available to the company,
- economics of prevention,
- attitude of the government to control industrial pollution through legislations, incentives, and penalties.

29.3 WASTE-MINIMIZATION CYCLE

Typically, a development cycle of industrial waste minimization programs comprises six phases: inception, audit, analysis, design and development, implementation, and evaluation, as illustrated in Figure 29.1. The overall goal of waste reduction and cleaner production program is to critically investigate, evaluate, design, and implement such environmentally benign processes and process improvements that would minimize consumption of resources, and energy and reduce waste generation in order to reduce adverse environmental impacts and effect in overall economic benefits.

These six phases can be discussed in more detail

1. *Inception Phase:* This phase comprises setting up the goals, commitments, methodologies, task force, time frame, and budget for the project. Such goals should be quantifiable, measurable, achievable, and usable to measure the success or failure of any waste minimization or cleaner production program in real terms. Senior management plays a key role in setting up the project framework, resources, and the project team.
2. *Audit Phase:* In the audit phase, the relevant factory processes include management processes and waste treatment processes. These processes are thoroughly

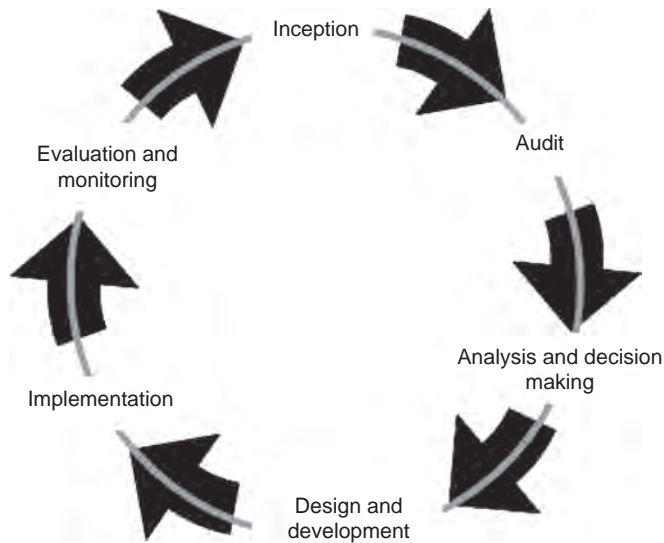


FIGURE 29.1 Typical waste minimization and cleaner production cycle.

investigated to obtain a complete balance sheet of the raw material and energy input and output, including waste. Over and above collecting and compiling all facts and figures, the audit exercise should recommend energy and waste minimization options to attain the desired goals of the program that could be carried to the detail analysis phase. If possible, preliminary technical and economical analysis may also be carried out to prioritize the options.

3. *Analysis Phase:* The analysis phase starts with detailed analysis of the findings and recommendation from the audit phase to explore the various opportunities and risks associated with the various options. It also explores further possibilities. This phase ends with making decisions on which options to be pursued and which to be dropped.
4. *Design and Development Phase:* The design and development phase starts with setting up the framework of design, development, and implementation of selected options. This would require planning of all actions. It is a good idea to implement different waste minimization and cleaner production improvements in stages to reduce impacts from introducing new process and process modifications. At this stage, process changes and new processes are designed and procured, and all preparatory works for implementation are undertaken.
5. *Implementation Phase:* In the implementation phase, processes and process modifications are installed and integrated to the existing system, commissioned, and put into operation. All operators and workers are trained for the changes in the process and new processes.
6. *Evaluation Phase:* This phase continues after the changes or new systems are fully integrated into the normal production processes. In this phase, the actual results from the process modifications are monitored, evaluated, and compared to that originally conceived. The cost of such monitoring is normally included in routine quality assurance/quality control activities.

This chapter deals with the audit phase. It discusses different aspects of waste and energy audit systems. The scope of such discussions has been limited to waste and energy audits within a typical manufacturing industry. For the purpose of this chapter, industry would be treated in general, with specific examples from different manufacturing sectors. The methodologies described here would generally apply to most manufacturing sectors. Some modification to the described methodologies would be necessary, depending on specific activities undertaken in the industry. Such modifications are left for the industry professionals to work out according to the requirements.

29.4 WASTE AUDITING

Once a waste-minimization program is set to be undertaken, the physical work starts with a series of detailed surveys of ongoing activities inside the industry, starting from raw material entering the premises to finished products and by-products (including wastes). These audits can be termed as *waste audits* or *waste-minimization audits*. The principal intent of such audits is to critically assess various inputs, processes and outputs to find methods and practices for minimizing waste and reduce the resource consumption in a more sustainable environmentally benign way. Traditionally, industrial waste audits do not include examination of the design of the product itself but investigate all activities of the production processes and opportunities of waste recycling/reuse, including waste treatment systems. Other terminologies may be used such as pollution prevention audit, eco audit, or green audit that essentially focus on some of the common objectives of preventive approaches.

The phases of a typical waste audit process are illustrated in Figure 29.2. Note that an energy audit process can also be divided into similar phases. The rest of the chapter discusses each of these steps in detail with illustrations, examples, and workouts.

29.4.1 Phase I: Preparatory Work for a Waste Audit

Preparatory work for a waste audit consists of three main steps:

1. getting the management and staff involved in the program,
2. forming an audit team and appointing a team leader,
3. planning the audit exercise.

Once a waste minimization program is begun, its time to provide the program with personnel, technical, and financial resources. The first step is to involve stakeholders in the program to get management and staff involved directly or indirectly in the program. This should be mainly done by one of the core management groups who would ultimately be responsible for managing the overall pollution prevention program. Normally, the production management or the environmental management group has the responsibility to execute such programs. Although all stakeholders can be involved in this process, more emphasis should be given on internal stakeholders such as the management, supervisors, and workers in taking up the initiative.

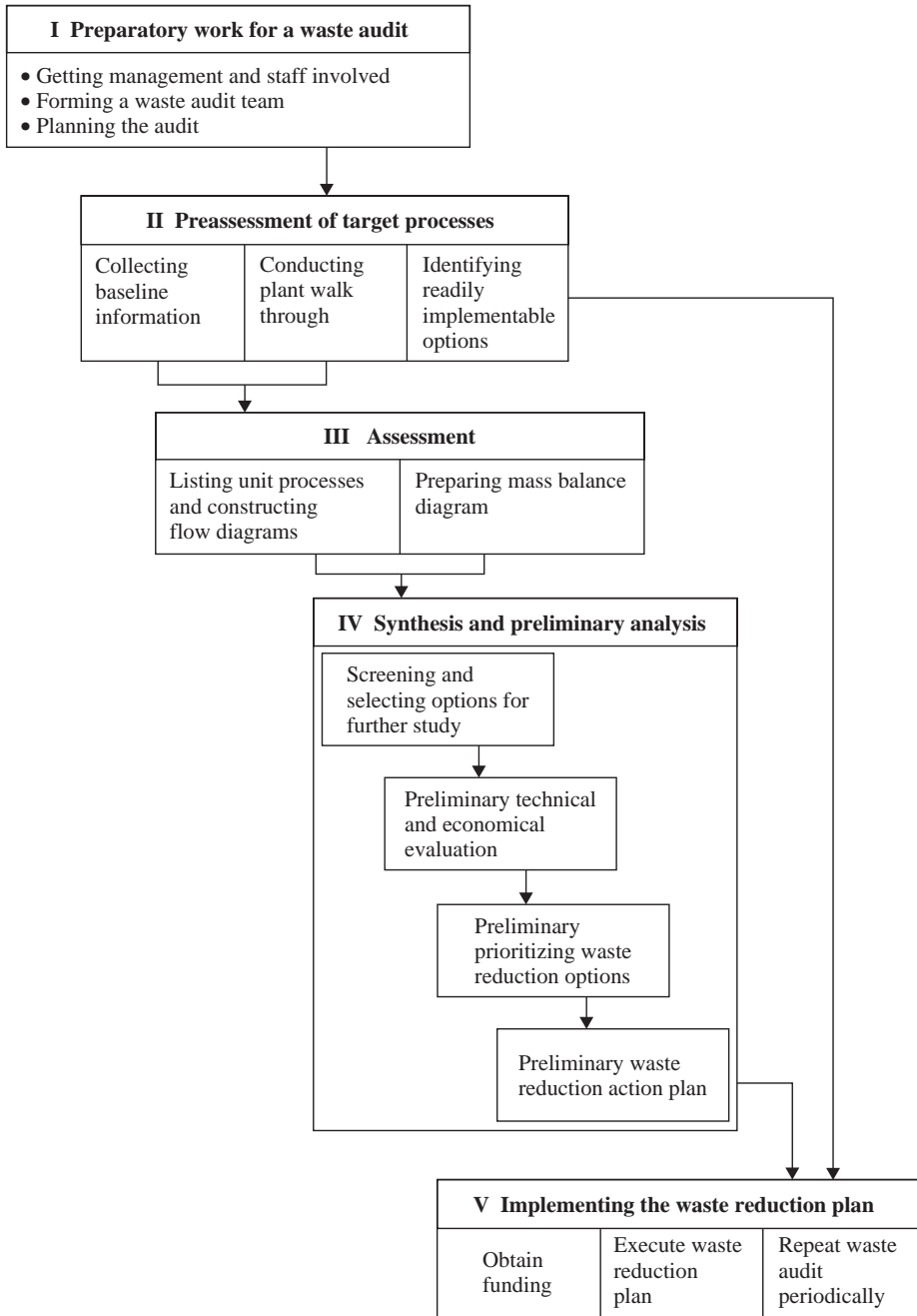


FIGURE 29.2 Steps of waste audit methodology.

29.4.1.1 *Getting Management and Staff Involved* Commitment from different management groups is a decisive factor in the success of a pollution prevention program. Management representatives from all relevant departments should be involved in the program. In a typical manufacturing industry, such departments may include the following:

- executive management (typically CEO or a deputy),
- product development and design (if present),
- production,
- procurement and inventory control,
- operation and maintenance,
- environmental,
- marketing,
- finance,
- quality control.

As the management staff would be supposedly aware of the environmental concerns and directly related to the welfare of the company, less effort would probably be required to motivate them for voluntary participation. However, it has to be confirmed that during the whole tenure of the program there is a high level of cooperation and support, even if such a program may cause short-term disruption to normal activities.

Involvement of the supervisors and workers is another key factor in the success of such programs. They have the hands-on involvement with each of the individual activities of the factory. In order to make the best use of their day-to-day experience with the machineries, processes, and a myriad of issues with the factory processes, audit exercise should be carried out with their full involvement and support. Typically, the barrier to such an involvement is fear on the part of certain supervisors and workers that a waste audit will expose inefficiencies that lead to job cuts. In order to overcome such a barrier, supervisors and workers should be assured of their job security, barring evidence of Fraud or sabotage.

The exercise can also be used as an opportunity to create more employee awareness about the environmental issues and energy and waste-minimization and cleaner-production program. Rewards in the form of bonuses, prizes, or acknowledgment would motivate employees to voluntarily participate in the program. There are several common ways of raising interest and motivation that could benefit the program (Box 29.1).

29.4.1.2 *Formation of an Audit Team* One of the key requirements of an audit exercise is to form a proper and balanced audit team that would be responsible for all subsequent audit works. The team is normally formed by the direction of the responsible management group, in discussion with senior management and all other management groups. If required, expert advice from external consultant may be sought at this stage.

Waste audit is an interdisciplinary activity. Therefore, the team should be formed with diversified expertise from representative groups or departments, which will have major contribution and interest in the program. Usually, an audit team is a subset of the project

BOX 29.1 COMMON WAYS TO RAISE INTEREST IN WASTE AND ENERGY MINIMIZATION PROGRAM

- Use posters or banners to inform the staff about the pollution scenario and the requirements, benefits, objectives, and goals of the upcoming waste minimization and cleaner production program.
- Publicize the upcoming audit exercise.
- Provide some prominent identification mark (such as badges) to the members of the audit team and also provide them some special privileges such as free access to any place or information during the audit program.
- Organize “Environment Week,” with programs such as speeches, videos, skits, or tree planting.
- Offer cash prizes/monetary incentives for the staff members who come up with innovative idea leading to cleaner production.
- Offer incentives for the machine operators/staff for quality production with minimum resources.
- Hold inter-departmental competition for waste minimization.
- Share the financial gains due to waste/energy minimization programs.

team but may include various other personnel whose contribution may be useful. For example, an audit team can be formed from these groups:

- environmental managers,
- plant managers,
- process or operation and maintenance engineers,
- occupational health and safety officers,
- supervisors and operators,
- laboratory technicians.

The team should appoint a team leader who will lead the audit team and coordinate the activities of all other team members. The plant manager or the environmental manager can be prospective candidates for team leader. It is recommended that a balanced audit team should be made up of three to six persons, including the team leader. This recommendation is based on typical industry structures but may be varied. In practice, the selection of the team leader and composition of the audit team will depend on the nature of the processes in the industry, scope of audit, and the scale of the industry.

Each of the audit team members should be aware of the goals and objectives of the overall energy and waste minimization program and should be able to contribute to it. The team leader should delegate duties to each of the members, depending on strength and capabilities, and should monitor them throughout auditing phase. Members of the audit team should be relieved from their routine activities during the audit exercise.

The requirement of external consultants or experts in the team would depend on the objectives of the program and complexity of the industrial processes. If the goals are to explore simple and readily applicable waste minimization or cleaner production

opportunities (e.g., reduction in general water and electricity consumption), plant engineers and supervisors may be sufficient for auditing purposes. However, if the requirement is complex, such as performance of cryogenic units or distillation columns, experts in these technologies may be required. In these cases, hiring an external consultant may be justified.

29.4.1.3 Planning the Audit Exercise Before the actual audit exercise is undertaken, a substantial amount of planning work is to be done in order to carry out the audit within time and budget and least interference to normal activities of the plant. Such planning works would mainly include the following:

- define the scope of the waste audit,
- develop an audit program,
- prepare specific workplan and checklists,
- develop uniform reporting procedures,
- inform all departments about upcoming audit programs.

However, depending on the specific need of the program, additional works may be added to the list.

Defining Scope of the Waste Audit Determining the scope of an audit is one of the fundamental steps before taking up the actual auditing process. Scope should be defined according to the goals and objectives of the waste minimization program. Goals should be quantitative, realistic, and achievable. Compliance with a set of legislative requisites within a fixed time period could be the prime goal in some instances. In other instances, there could be the need to reduce (by a certain percentage) the operation of waste treatment cost by reducing the quantity and strength of some priority pollutants (e.g., chromium, copper, or organic pollutants such as phenols, cyanides). In some other cases, the goal may be to reduce consumption of chemicals (by certain tonnes per unit product) by improving certain processes. Waste audits can also focus on specific objectives such as improving the general environmental management system in environmental management system audits, improving operational health and safety by reducing use of toxic chemicals in operational health and safety audits, or total environmental risk audits, and so forth. In every case, there should be some target sectors or processes, which have to be closely scrutinized. Identification of these target processes is crucial in defining the scope. The most common target processes of waste or energy audits are as follows:

- *Production Processes:* All unit production processes in the plant need to be audited to check for process inputs, outputs, operating conditions and controls, and process efficiency.
- *Inventory Processes:* Inventory processes such as storage and handling of chemicals, machine parts, and other accessories need to be evaluated to check loss of material while handling or storage, loss of chemicals that have expired, and so on.
- *Housekeeping Processes:* Housekeeping practices should be audited in order to check various losses during cleaning, arranging, and various types of maintenance practices. Better housekeeping practices can reduce leaks, spills, dragout, losses in rinsing, cleaning, and so on.

- *Waste Treatment Processes:* All waste treatment processes such as effluent treatment plants, gas treatment, and dust filters should be audited to determine treatment efficiency, operating conditions and controls, and so on.
- *Packaging Processes:* Packaging sometimes requires specialized systems that can generate wastes. Therefore, such processes must also be audited if waste generation or energy consumption is high in such processes.

Unless the scope of the waste audit is positioned to achieve the overall goals of the program, many efforts can go waste. Table 29.1 shows how some of the common objectives are related to the target processes. Defining anything more than what is required or less may result in a poor outcome. For example, if the objective of a waste-minimization project is to reduce water consumption in a factory, only those processes/activities (target processes) that are using significant amount of water need to be checked. In this case, it is not particularly useful to audit inventory processes or final packaging systems that consume no water or very little water. However, when the goals are to explore opportunities of electricity reduction, probably all activities need to be audited, as there are practically no areas in a factory where electricity is not consumed. But in the first instance, it may be better to audit those sections where large amounts of electricity are consumed. The scope can thus be much wider compared with water-reduction programs.

Defining a balanced scope is not generally a one-step procedure. Starting with preliminary scope, the audit team may modify the scope as the audit is undertaken and new information emerges, which may need extension of scope to certain target processes and elimination of certain other sectors that may not significantly contribute to the overall objective. The scope should also be subjected to debate by all team members of the program to arrive at a consensus.

Develop an Audit Program Similar to all other projects, an audit exercise should stick to a program and plan to avoid any overrun. The audit team should develop a series of programs for auditing each of the target processes and timeframe for each of the activities.

TABLE 29.1 Audit of Target Processes as per Objectives of Waste-minimization Program

Main Objective	Target Processes
Legislative compliance	Probably all processes, with more emphasis on production and waste treatment processes
Reduction of toxic and hazardous wastes	Production and inventory processes where hazardous and toxic chemicals are used. Special attention to be given on transport, handling, storage, and use of such chemicals, as well as treatment and disposal of the hazardous wastes
Operational health and safety Improvement	Special attention to production processes, especially with high temperature, high voltage applications, and so on. Attention to be given to handling, storage, and use of hazardous and toxic chemicals. Housekeeping processes need to be checked as well
Compliance audit	Mainly the effluent treatment units only, with emphasis to check if they are meeting the required discharge standards

Generally, for a complete waste and energy reduction audit, the following time frames can be used for manufacturing plants:

- small scale (less than 50 workers): 3–4 weeks,
- medium scale (50–100 workers): 4–6 weeks,
- large scale (more than 100 workers): 6–10 weeks.

This timeframe is generalized; several factors such as the number of steps in the production process, the degree of complexity, the level of automation, the quantity, and characteristics of different feedstock used should be considered for determining the timeframe. In the absence of any specific condition for the program, there are two ways of setting up an audit program:

1. follow the material flow path through the industry,
2. audit target processes according to the significance in terms of waste reduction.

Both systems have some merits. For example, following the material flow helps keep good track of several raw material as these are transformed into products and by-products, which gives a complete account of material inputs and outputs. While auditing according to priority of target processes, allow a thorough investigation of major target processes that can identify more waste-minimization opportunities. Development of program is a dynamic activity and can be modified as more and more audit exercises are undertaken and new clues evolve.

Audit timing and frequency are important factors in planning audit exercises. Timing and frequency can be fixed, depending on the following:

- type (continuous, batch, etc.) of the production process,
- number of parameters to be audited,
- scale of production,
- accuracy of audited information required.

Auditing can be scheduled during peak production hours or uniformly throughout a complete production cycle, depending on the nature of the production process. For example, in a typical batch production process (such as textile dyeing), quantity of wastewater discharged at the end of each cycle need to be audited, whereas in a continuous production process (such as pulp processing in a pulp and paper industry), discharge of wastewater may be monitored at 4-h intervals. Care should be taken not to disrupt the normal factory processes.

It is recommended that the major target processes are audited more than once, typically three times, to obtain good representative results and to ensure repeatability of the information, while less important sectors can be audited once. Quality and reliability of an audit exercise would very much depend on the frequency of audit on each process, as each time a process is studied, there is an increased chance of getting a new clue.

Prepare Specific Workplan and Checklists Each team member should prepare his or her own specific workplan and checklists, depending on the nature of the activities to be audited. Preparing workplan and checklists would enable auditors to focus on key activities to

be audited and collect all necessary information for those processes and activities. Some understanding of the processes would be required to prepare the workplan and checklists.

Develop Uniform Reporting Procedures The audit report is the final product from the audit phase. All subsequent activities would very much depend on the audit report, and its importance in the program is substantial. It is, therefore, recommended that the audit report should be well planned and easily comprehensible, and should contain all the information that may be required during the subsequent phases. The contents of the report should be determined based on the objectives of the program. A typical table of contents for a report is given in Box 29.2. Depending on the requirement, some of the sections can be excluded. However, important sections should be retained in a good audit report.

Inform All Departments about Upcoming Audit Programs Once the audit program is finalized, the audit team should notify all the respective departments about the upcoming audit. This would enable the departments to be prepared for the audit. The program should be advertised at suitable locations throughout the factory to remind the supervisors and workers about the audit schedule.

BOX 29.2 TABLE OF CONTENTS FOR A TYPICAL WASTE AUDIT REPORT

- title page,
- disclaimer, if any (disclaiming responsibility by the publisher for data or views expressed),
- table of contents,
- executive summary (summarizing the complete audit exercise in the industry),
- problem statement, objectives, and priorities,
- adopted approach and reasoning,
- process layout, description, and observations,
- sources and quantities of pollution load generated,
- existing treatment facilities and additional requirements,
- observations on proposed treatment systems,
- integrated pollution prevention strategy,
- generation of options, screening,
- option grouping and prioritization,
- identification of candidate systems,
- system evaluation criteria,
- calculations,
- method and results of system evaluation,
- recommendations,
- implementation strategy,
- suggestions to the management,
- summary and observations,
- appendices (tables, graphs, assumptions, and calculations).

29.4.2 Phase II: Preassessment of Target Processes

The purpose of the preassessment is to collect all information on the target processes that may be required in detailed analysis

- collect baseline information,
- conduct a plant walk-through,
- identify immediate implementable options.

29.4.2.1 Collection of Baseline Information Collection and compiling of information is one of the major tasks and purposes of waste auditing. While information on target processes would be of primary value, practical experience shows that various other information may be useful. As such, the scope of data collection can range over the entire cross-section of the factory:

- *Organizational Data:* Organizational chart, factory layout, and site plan are included. Also collect information about the surrounding area indicating topography, water bodies, hydrology, agricultural areas, and human settlements.
- *Material and Product Data:* Data include specifications of feedstock, process water, product and by-products with composition, instruction on usage and discharge, material safety data sheets, and usable and storage life. Information on quality assurance and quality control of feedstock, process water, products, and by-products must also be collected.
- *Raw Material and Logistic Consumption Data:* Feedstock, energy, and water consumption records are important. Possible source of records of feedstock consumption can be available from stores and accounts department. Water usage can be obtained from water meters or bills, and energy usage from energy bills.
- *Process Data:* Collect process flow diagrams, block diagrams, material balance diagrams, control and operational logic diagrams and instructions, manufacturer's data on each machine and process, and maintenance plans and records.
- *Environmental Data:* Collect data on air emission, solid waste generation and effluent from different machineries, including waste treatment systems, environmental directives, and licenses.
- *Management Data:* It is important to document the number of staff, their position and responsibilities, performance records, administrative instruction, occupational and safety procedures, quality assurance, and quality control procedures.
- *Financial Data:* Product, utility and raw material cost, cost of waste treatment, operating and maintenance cost are available as part of the company's financial statements.
- *Industry Data:* If possible, gather all of the data for industries of similar nature. Such information, though difficult to obtain, would however allow comparison with other plants that may help in defining realistic targets and goals.

All the audit members should be well familiar with the information that would allow them to objectively carry out the audit. To an experienced auditor, even scanning through such information can also give some clue about the opportunities, which areas need to be audited in detail, and room to improve the audit exercise. All information should be properly

preserved and sources of such information should be noted to check the reliability at a later date.

All information should be checked for underlying data quality in terms of correctness and repeatability. If possible, historical data should be gathered over a period of time (suggested 2–3 years' records) and the process of data collection should be continued during the entire audit phase on a monthly or quarterly basis. It is suggested that statistical analysis be performed on the collected data to ensure quality of data before these are used for analysis purpose.

29.4.2.2 Plant Walk-through Survey A thorough walk-through the plant is an essential part of the preassessment phase. It is recommended that all team members be involved in such audit exercise, which should generally be carried out over few days, with one or two sections covered in each day. Moreover, the audit team should be accompanied by the responsible section manager, engineer, or supervisor who is completely aware of each activity in the respective section. All types of management and technical information should be available during such surveys for the audit team to check. There are several benefits of such exercise:

- It is quite likely (especially in the case of large industries) that the members of the audit team may not be very familiar with the different activities carried out in each section other than their own. Therefore, this exercise would give them a chance to get conversant with many more activities and processes.
- It can reveal some obvious waste-minimization or cleaner-production options that may not need a detailed assessment to work out the recommendation. A common sense approach, or simple calculations with a lesser degree of accuracy or thumb rule estimates, may be sufficient for arriving at the conclusion. For instance, in situations where the steam pipes are not insulated or valves are leaking, there is no need to carry out a detailed assessment of energy or steam losses to arrive at the recommendation to insulate the steam pipes or repair the valves.
- By eliminating such obvious options, the auditors can narrow down the scope to those areas that require detailed assessment. This frees up some resources, and the auditors can then concentrate on more intriguing issues.
- With a balanced team of auditors, personnel from other sections or departments can more critically observe activities of another department. This can give rise to lateral thinking, and more avenues for improvement can be explored.

The best strategy for the walk-through is to follow the material flow path through the industry—from the storage of raw material, through various production processes, until it is converted to the final product and stored, in absence of any other plan. During the walk-through exercise, each member of the audit team should take detailed note of all the activities, facts, and figures; and any other information that may be useful at a later stage. It is suggested that even trivial observations are noted, as these can form some clue at a later stage. It is preferred that the auditors should prepare their own sketches, schematic arrangements, material flow diagram, block diagrams, and site plans during this walk-through exercise. Even if some process diagrams are already present, often minor changes are carried out in the plant during operations. Notes of such changes should be taken. At the end, team members should prepare their own report of the plant walk-through.

29.4.2.3 Identification of Readily Implementable Options After carrying out the plant walk-through and making detailed notes, the audit team must discuss the various observations made and should identify a number of simple and obvious measures to reduce waste generation. Such options should be simple, quickly implementable, and inexpensive. Significant waste reductions can often be achieved by such options, which are based on improved operation, better handling, and tightening up of housekeeping practices. For example, simple measures such as attending to leaking hoses or installing automatic level controllers may lead to significant water savings.

Segregation of waste is arguably one of the numerous measures that can effectively lead to waste reduction. It is the most central of such options, and is a universal issue that needs to be addressed. Segregation of waste can offer enhanced opportunities for recycling and reuse with resultant savings in raw material costs, at the same time reducing treatment costs. Concentrated simple wastes are more likely to be of value than dilute or complex wastes. The waste collection and storage facilities should be reviewed to determine if waste segregation is possible.

Such options are implemented as soon as possible without waiting for the final recommendations. Typically, implementation of such options should be completed within 2–3 weeks. If implemented, the performance and impacts of the changes should be closely monitored and included in the audit report. All the modifications made should be noted, and these will have to be considered in the later stage while developing the detailed process flow diagram.

It is expected that at the end of the preassessment phase, the audit team is

- organized and aware of detailed scope,
- aware of all target process layouts for further audit,
- aware of all unit operations in each of the target processes,
- aware of sources of waste and their causes.

At the end of the preassessment, the plant personnel should be well informed of audit purposes. Resources should be secured, and readily implemented waste-reduction measures should be identified and, if possible, implemented.

29.4.3 Phase III: Assessment

This phase can be broadly divided into two steps:

1. list the unit processes and constructing flow diagrams,
2. prepare a mass balance diagram.

29.4.3.1 Listing of Unit Operations and Constructing Process Flow Diagrams This step is essential for an audit program, as it gives a detailed insight into the production operations/process vis-à-vis sources of waste generation and hence enables identification of avenues for better operating practices and waste reduction.

To develop a good representative block process diagram, the audit team should undertake a detailed walk-through in the production units and utility areas, in order to gain understanding of all the processing operations and their interrelationships. The production or plant staff should be interviewed to know about the actual operating controls,

parameters, and issues. Only after conducting a detailed walk-through and interviews with the production staff, should the audit team compile the required baseline data.

By connecting the individual unit operations in the form of a block diagram and highlighting the flow of materials, a process flow diagram can then be prepared. All the information related for example to raw materials, products/by-products, energy, water inputs, waste discharged, material and energy flows, motion, and time should be compiled during this preassessment stage and should be presented on the block diagram.

The input and output information for each unit operation should be summarized in standard units by reference to the process flow diagram. Standardized color coding may be used to represent, say, raw material input by a black line, products by blue line, wastes by red lines, and recycled stream by green lines. Intermittent operations such as cleaning, makeup, or tank dumping may be distinguished by using broken lines to link the boxes. Similar notation may be used to distinguish batch and continuous discharges.

29.4.3.2 Material Balance: Process Inputs and Outputs In the material balance exercise, a detailed account of the process inputs and outputs is made to identify the problem areas and thus the need for improvement. Material balance is important for any waste-minimization project to identify and quantify previously unknown losses or emissions. Material balance is also useful for estimating the costs of additional installations and/or modifications.

By definition, the material balance includes materials entering and leaving a process. Inputs to a process or a unit operation may include raw materials, chemicals, water, air, and energy. Outputs include primary product, by-products, rejects, wastewater, gaseous wastes, liquid, and solid wastes that need to be stored sent off-site for disposal and reusable or recyclable wastes (Fig. 29.3). In its simplest form, a material balance is drawn up according to the mass conservation principle:

$$\text{Mass in} = \text{Mass out} + \text{Generation} - \text{Consumption} - \text{Accumulation}$$

If no chemical reactions occur and the process progresses in a steady state, the material balance gets simplified to

$$\begin{aligned}\text{Mass in} &= \text{Mass out} \\ \text{Water in} &= \text{Water out} + \text{losses (evaporation, spills, etc.)}\end{aligned}$$

Sources of Information for Material Balance There are many sources of information in establishing material balances for the various unit operations within the plant. Data may be obtained from sample analysis and measurements of raw input materials, raw material purchase records, material and emission inventories, equipment cleaning and validation procedures, batch composition records, product specifications, operating logs, standard operating procedures, and manuals.

Material balances are easier, more meaningful, and more accurate when they are done for individual production units, operations, or production processes. For this reason, it is important to define the material balance envelope or boundary limit accurately, in addition to the tie compound. Ideally, a more accurate balance should be established for the unit operation that is more critical from the waste-generation and reduction point of view, and a less accurate balance could be established for other processes.

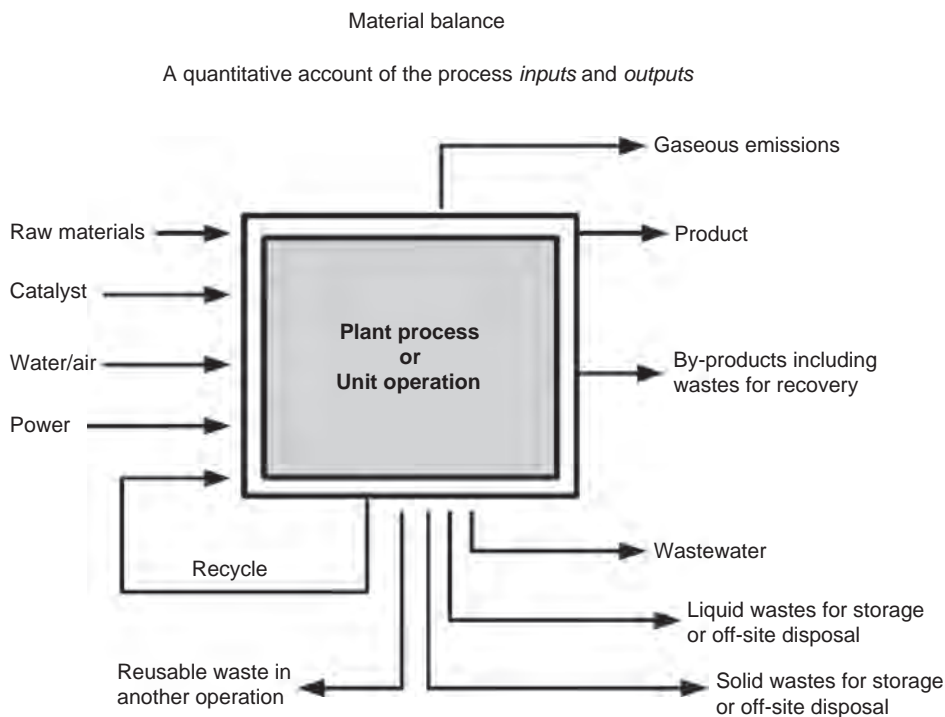


FIGURE 29.3 Schematic representation of a material balance sheet.

Although it is not possible to lay precise and complete guidelines for establishing the material balance, the following guidelines might be useful:

- In the case of an extensive and complex production system, it is better to first draw up the material balance for the whole system (or even the entire production facility as such), and then concentrate on individual operations.
- When splitting up or desegregating the total system, choose the most simple, individual subsystems that are critical from the waste reduction point of view.
- Choose the material balance envelope in such a way that the number of streams entering and leaving the process is the smallest possible.
- Always choose recycle streams within the envelope to start with.

For complex waste-minimization audit, it might be desirable first to make a preliminary or draft material balance and in the second step to evaluate and refine it. However, in case of simple audit for small plants, the steps can be merged into one.

Selection of Priority Unit Operation Although the material balance should be set up for all the unit operations, the unit operation most important from the point of view of waste generation must be identified and efforts are concentrated for that particular unit operation. This can be done by professional judgment and technical know-how of the audit team and specifically the production personnel.

Selection of Tie Compounds A tie compound is the parameter (or substance) for which the material balance is established around a unit operation or a process. It is important to select an appropriate tie compound. Criteria for selecting the tie compound could be

- expensive raw material/intermediate,
- material common in most processing stages,
- substance of hazardous nature,
- substance/compound easy to measure/estimate.

A simple example of a tie compound could be water to account for most wet operations. Establishing water balance for the processes using substantial amount of water can often provide useful clues for cleaner production. In practical situations, more specific tie compounds (e.g., nickel or zinc in electroplating shops or dyestuff in textile processing) would be ideal. Another good example could be that of chromium in leather tanning.

Chemical oxygen demand (COD) is another useful tie parameter that sharpens the material balance exercise, especially to link the production areas with the effluent treatment plant. The audit team can estimate the contribution of each process department in terms of total COD load in kilogram per day, knowing the volume of wastewater and the COD discharged by each department and cross checking it with the COD load observed at the treatment facilities.

One need not be very particular over the accuracy (of the order of 99%) of the material balance. In practice, such high accuracies are rarely achievable. Material balance within the tolerance range of 10% should generally be acceptable. However, if the tie compound for material balance is hazardous, a higher order of accuracy should be targeted.

Steps for Preparing a Material Balance There is a logical series of steps for preparing a constructive material balance:

1. *Determine Inputs:* The inputs to the process and to each unit operation need to be quantified. As a first step toward quantifying raw material usage, purchasing records should be examined; this readily gives an idea of the quantities involved. The raw materials purchases and storage and handling should be recorded in a table format in order to derive the net input to the process.

Water is frequently used in the production process, for cooling, gas scrubbing, washouts, rinsing, and steam cleaning. The water usage needs to be accurately quantified as an input. Also, some unit operations may receive recycled material from other unit operations. These also represent an input. Hence, water and recycled materials need special attention, and therefore steps 1A and 1B describe how to evaluate these two factors.

- 1A. *Record Water Usage:* The use of water, other than for a process reaction, should be covered in all cleaner production programs. The use of water for washing, rinsing, and cooling in process and in utility operations is often overlooked, although it represents an area where effective waste reduction can frequently be achieved simply and cheaply.

- 1B. *Measure Current Levels of Material Recycling:* Some materials may be transferred from one unit to another (e.g., reuse of the final rinse in a soft drink bottle washing plant as the initial rinse); either directly or after some modifications/treatment. If recycled materials are not properly documented,

double counting may occur in the material balance, particularly at the process or complete plant level; that is, a material will be quantified as an output from one process and as an input to another. Proper attention must be paid to this issue, and care must be taken to avoid any discrepancies.

2. *Quantifying Outputs:* To calculate the second half of the material balance, the outputs from unit operations and the process as a whole need to be quantified. Outputs include primary product, by-products, wastewater, gaseous wastes (emissions to atmosphere), and liquid and solid wastes that need to be stored and/or sent off-site for disposal and reusable or recyclable wastes. It is important to identify appropriate units of measurement.

If the product is sent off-site for sale, then the amount produced is likely to be documented in company records. However, if the product is an intermediate to be input to another process or unit operation, then the output may not be so easy to quantify. Production rates will have to be measured over a period of time. Similarly, the quantification of any by-products may require field measurement.

- 2A. *Account for Wastewater:* On many sites, significant quantities of both clean and contaminated water are discharged to sewers or to a watercourse. In many cases, this wastewater has environmental implications and incurs treatment costs. In addition, wastewater may wash out valuable unused raw materials from the process areas. Therefore, it is extremely important to know how much wastewater is going down the drain and what the wastewater contains. The wastewater flow, from each unit operation as well as from the entire process, must be quantified, sampled, and analyzed.

- 2B. *Measure Gaseous Emissions:* To arrive at an accurate material balance some quantification of gaseous emissions associated with the process is necessary. For example, a tea drier exhaust may carry fine particles of tea dust. Measurement in such cases calls for instruments such as thimble probe dry gas meter–vacuum pump assembly. In many instances, gaseous emissions carry some amount of hazardous materials also (such as VOCs). Expert assistance may be needed to determine the material/product loss through gaseous emissions.

3. *Prepare a Preliminary Material Balance:* A material balance is designed to provide better understanding of the inputs and outputs, especially waste, of a unit operation such that areas where information is inaccurate or lacking can be identified. The initial balance should be considered as a rough assessment that must be further refined and improved.

The units of measurement should be standardized (liter, ton or kilogram) on a per day, per year, or per batch basis. The measured values in standard units should be summarized by reference to the process flow diagram. It may be necessary to modify the process flow diagram following the in-depth study of the plant. It is highly desirable to carry out a water balance for all water inputs and outputs to and from unit operations, because water imbalances may indicate underlying problems such as leaks or spills. Similarly, a detailed material balance should be carried out for important tie compound, as agreed upon by the audit team during the planning phase.

4. *Evaluate and Refine Material Balance:* The individual and sum totals making up the material balance should be reviewed to determine information inaccuracies. Ideally, the input should equal the outputs, but in practice, this will rarely be the case. Some judgment will be required to determine what level of accuracy is acceptable. If there is a significant material imbalance, then further investigation is needed.

When constructing material balances, watch for factors that could overstate or understate waste streams. Sometimes, all or at least a few steps of material balance may need to be repeated a few times in order to refine the material balance. These may include quantification of a few input or output streams or even hunting for some material flows that might have been totally missed in the initial stage. Additional field sampling and analysis may also be required to be carried out in certain cases, and thus the data collected should again be organized and represented so as to establish an accurate material balance.

29.4.4 Phase IV: Synthesis and Preliminary Analysis

Phases I to II have covered planning and undertaking waste audit, resulting in the preparation of a material balance for each unit operation. Phase IV represents the interpretation of the material balance to identify process areas or components of concern. Figure 29.4 represents a material balance algorithm for the textile industry in establishing waste reduction options.

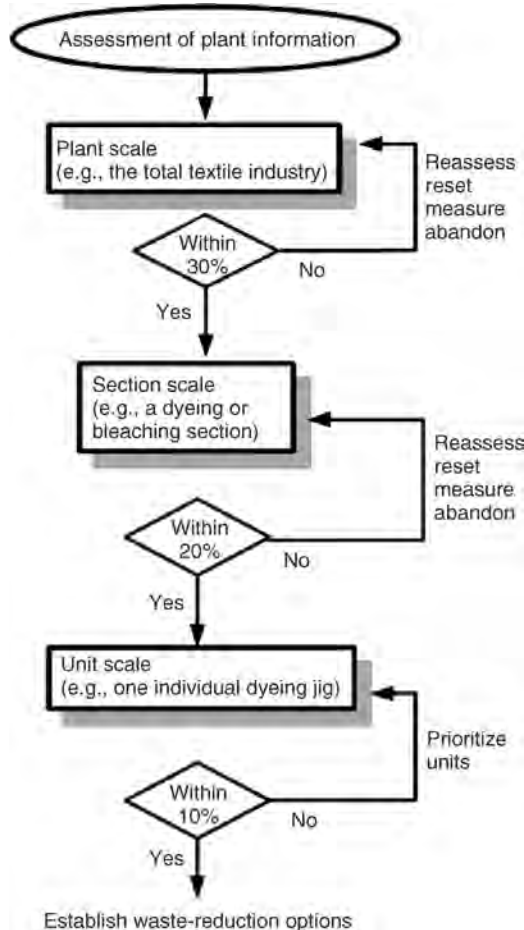


FIGURE 29.4 Material balance algorithm for textile industry.

To interpret a material balance, it is necessary to have an understanding of normal operating performance. Thus, a member of the audit team must have a good working knowledge of the process. To a trained eye, the material balance will indicate areas for concern and help to prioritize problem wastes. Using the material balance, major sources of waste may be identified, deviations from the norm in terms of waste production may be found, areas of unexplained losses may be determined, and operations that contribute to flows that exceed national or site discharge regulations may be pinpointed.

In this phase, several possible waste-reduction measures are identified that can be proceeded to the analysis phase. Different waste-minimization programs may require varying degrees of effort, time, and financial resources:

- obvious waste-reduction measures, including improvements in management techniques and housekeeping procedures that can be implemented cheaply and quickly,
- long-term measures involving process modifications or process substitutions to eliminate problem wastes.

29.4.4.1 Screening and Selecting Options for Further Study (Weighted Sum Method) For options that require numerical evaluation, the most commonly used tool is the *weighted sum method*. Screening and selection is recommended when a large number of options have to be considered. Only those options that have sufficient merits should be carried forward. The weighted sum method provides a means of quantifying the important criteria that affect waste management in a particular industry:

1. *Determine what criteria would be considered in evaluation of the options:* The higher degree of improvements achieve in the given criteria, the better is the result. For example, in a typical waste minimization program, such criteria can be
 - amount of reduction in waste quantity,
 - amount of reduction in raw material consumption,
 - amount of reduction in hazardous or toxic waste,
 - amount of improvement in health and safety condition,
 - ease of implementation,
 - cost of implementation,
 - amount of resource and time required.

These criteria should be determined in terms of meeting the overall waste-minimization objectives and goals. This should also take into consideration for various types of constraints that may be present. Judgment would be required to select the criteria.

2. *Once the Criteria Are Determined, Each Criterion Should Be Given A Weight:* The more important a criterion is, the higher its weight. Again, a fair amount of judgment is required to determine the weight for each criterion. Generally, a relative scale of 0–10 is used to allot weight. For example, if reduction in waste treatment and disposal costs is more important, while time of implementation is relatively less important, then the reduction in waste treatment costs is given a weight of 8, whereas the time of implementation is given a weight of 2 or 3.

3. *Each Option Is Then Rated for All the Criteria:* A scale of 0–10 could be used for rating. Marks are given according to the degree the criteria are satisfied (i.e., higher marks are given when the option fulfils or satisfies the criteria and low marks when the option does not suit the criteria).
4. *Each Rating (for the Option) Is Multiplied By the Corresponding Weight of the Criterion:* An option's overall rating is the sum of all the products of ratings times the weight of the criteria.

The options that carry higher marks would be carried for further analysis. Table 29.2 presents an option evaluation by weighted sum method.

29.4.4.2 Preliminary Technical and Economical Evaluation Once the options are screened and selected for further analysis, preliminary technical and economical evaluation may be required to set the priority of options and develop a preliminary action plan.

Technical Evaluation. The technical evaluation determines whether a proposed waste-minimization option will be technically feasible and achievable within the framework of existing constraints. A technical evaluation often begins with examining the impacts of the proposed option on production processes, production schedule, product quality, extra resource requirement, real estate requirement, operational feasibility, and safety. Several constraints such as disruption to normal production schedule, shutdowns, match of specifications (between the existing equipment and the new one), lack of technical knowledge, and trained manpower may be presented. Moreover, psychological resistance to changes may also be an issue. Therefore, it is recommended that all serious changes are first tested at a laboratory-scale and pilot scale. And, trial runs with the prototypes and test products are undertaken before the change is implemented and integrated to the actual production process. It is also suggested that engineers, supervisors, and operators are suitably trained for the change.

Waste-minimization options can also have some environmental impacts. During technical evaluation, the environmental effect of implementing the option needs to be checked. For example, if an option calls for recycling of rinse water, the effect of disposal of such water needs to be evaluated (as the concentration of solid in the recycled water would increase many times due to recycling), which may impact the receiving body.

Economic Evaluation. Economic feasibility is one of the most important criteria in determining the selection of an option. Unless forced by legislative requirement, there would be no cases where the economic merit of an option would be measured. Normally, each organization has its own economic criteria for selection of projects. However, three main criteria must be evaluated, irrespective of other criteria:

1. The capital, operation, and maintenance cost of the option.
2. The benefit that it would return over and above the existing system.
3. The resulting payback period.

The relationship among capital cost, payback period, and likely acceptance is given in Table 29.3. Payback period can be considered as *long* if it exceeds 2 years; *medium* if it is between 1 and 2 years; and *short* if less than 1 year for common waste minimization options.

TABLE 29.3 Relationship Between Capital Cost, Payback Period, and Likely Acceptance

Capital Cost	Payback Period	Likely Acceptance
High	Long	Low
High	Medium to short	Medium
Medium	Long	Medium
Medium	Medium	Medium
Medium	Short	Medium to high
Low	Long	Medium to high
Low	Medium	High
Low	Short	High

29.4.4.3 Preliminary Prioritizing Waste Reduction Options Once waste-minimization options are evaluated, they would need to be prioritized for further analysis, design, and implementation. A preliminary prioritization at the audit phase will expedite the process of analysis in the next phase. Prioritization can be done by weighted sum method, in absence of any special criteria (e.g., where some waste-minimization measures become mandatory due to legislative requirements, these can then be classed as high priority options, without further analysis). Some of the criteria that can be used to prioritize the options follow:

- technical ease in implementation,
- resource and time requirement,
- impacts on production schedule,
- short-term capital requirement,
- payback period.

29.4.4.4 Developing Preliminary Waste Reduction Action Plan Upon suggesting the priority, the audit report should also delineate a preliminary action plan for implementation. This can be represented as a regular bar chart or in any other form that is acceptable. In normal cases, the implementation sequence should follow the order of priority. It is suggested that implementation of waste minimization options is taken up in stages so that the cumulative impact on production processes, resources, and finance can be kept low. It is probably best to initially implement low-cost options with relatively simple technical requirements followed by progressive implementation of more complex changes that may require higher capital costs. It may also be a good idea to implement an option and test the success of it before the next one is undertaken to reduce the total risk involved in undertaking changes.

29.5 CONCLUSION

Waste auditing is as important as any other step in a waste-minimization or cleaner-production project. A proper waste audit should essentially provide a platform on which the rest of the project would be built. It should be a repository of all data and information that

would be required to carry out the rest of the phases. As information is the strength so is a successful waste audit.

The principal intent of a waste audit is to critically assess various inputs, processes, and outputs to find methods and practices for minimizing waste and reducing the resource consumption in a sustainable and environmentally benign way without compromising the commercial interest of the company. The waste audit phase would typically be a data collection and information synthesis phase that explores the current situation. Traditionally, an industrial waste audit also develops a list of waste minimization options and undertakes some preliminary technical and economic feasibility studies on the identified options to recommend about these options.

In general, the first step of a systematic waste audit is to prepare for the audit by forming a team, defining the scope, and programming the audit timings and budget. Scope should be defined in line with the overall objective of the waste-minimization program. The next step is to collect all baseline information, conduct plant walk-through surveys and identify readily implementable options. This would be followed by a detailed assessment of target sectors and unit processes and performing the mass balance analysis. At the end, the data should be analyzed to buildup an array of information and recommendation that would set the direction of the next phase. All data collected, information derived, and recommendation made should be presented in a form of a comprehensible and well-laid-out audit report.

It should be appreciated that a waste audit forms the vital activity of data collection and investigation of the current condition of the factory and its activities. Therefore, it is the very base of the next series of activities. The waste audit should be undertaken with utmost care and should be as thorough as possible. The more effort is spend at this phase, the better will be the chance of success of the project.

FURTHER READINGS

- Aldokhin NA, Goncharov AI, Grachev MA, Suturen AN. Recycling of wastewater and solid waste at the selenginsk pulp and paper plant. *Industry and Environment* 1990;13(3–4):21–23.
- Bishop, *Pollution Prevention: Fundamentals and Practice*. New York: Mc-Graw Hill; 2000.
- Cushnie GC. *Pollution Prevention and Control Technology for Plating Operations*. Ann Arbor: National Center for Manufacturing Sciences; 1994.
- Modak P. *Waste Minimization: A Practical Guide to Cleaner Production and Enhanced Profitability*. Ahmedabad, India: Center for Environmental Education; 1995.
- Modak PM, Visvanathan C, Parasnis M. Cleaner production audit. *ENSIC Review* 1995; 32.
- Nemerow NL. *Zero Pollution for Industry: Waste Minimization through Industrial Complexes*. New York: John Wiley; 1995.
- Sorrell S, Skea J. *Pollution for Sale: Emission Trading and Joint Implementation*. Cornwall, UK: MPG Books Ltd.; 1999.

30

ORGANIZATIONAL PERFORMANCE MEASUREMENT

JENNIFER A. FARRIS, EILEEN M. VAN AKEN, AND GEERT LETENS

30.1 Introduction

- 30.1.1 What is organizational performance measurement?
- 30.1.2 Why is organizational performance measurement important?
- 30.1.3 What are the key components of a performance measurement system?
- 30.1.4 How should organizations design performance measurement systems?
- 30.1.5 How should organizations design more effective measurement systems?
- 30.1.6 How should organizations identify causal relationships between measures?
- 30.1.7 What process should an organization use to design its performance measurement system?
- 30.1.8 How should an organization implement and use its performance measurement system?

30.2 Summary

References

30.1 INTRODUCTION

Another challenging domain of measurement expertise for engineers is related to organizational performance measurement. An organizational performance measurement system attempts to provide a “dashboard” for the purposes of controlling and improving an organizational unit (referred to here as the “target system” or “target organizational system”), such as a work team, department, division, or the organization as a whole. The following chapter will discuss some key issues in the design, implementation, and use of organizational performance measurement systems. Before this discussion, we start with some basic definitions to lay a foundation for the rest of the chapter.

30.1.1 What is Organizational Performance Measurement?

Organizational performance measurement has been defined in various ways in the existing literature and is closely related to several similar terms—including “performance measures,” “performance management,” and “performance measurement system.” Although there is no one commonly accepted definition for any of these terms, we discuss here some of the alternate definitions proposed in the literature, as well as the definitions that will be used for the purposes of this chapter.

One recent source defined “performance measurement” in simple terms: “Performance measurement is the comparison of actual levels of performance to pre-established target levels of performance” (Slizyte and Bakanauskiene, 2007, p. 317). While perhaps incomplete in some sense, this definition highlights the key fact that the purpose of performance measurement is to be able to ascertain the extent to which the organization’s current performance meets organizational objectives related to maintaining or improving its processes, activities, functions, and/or outcomes. For the purposes of this chapter, *performance measurement (PM)* is defined as “the act of using a set of preidentified indicators to track the way the organization functions over time in one or more key areas related to strategic goals.” Thus, performance measurement is intricately linked to the organization’s strategic planning process, as will be discussed more in the following sections.

Meanwhile, an “organizational performance measure” can be defined as “a metric used to quantify the efficiency and effectiveness of business strategy. It conveys financial or nonfinancial information that influence decision-making and managerial action taking” (Chearskul, 2010, p. 16 adapted from Simons, 2000, and Neely et al., 1995). Other terms used for this concept include “performance metric” or “performance indicator.” In this chapter, *performance measure* is defined as “an individual indicator used to quantify some key aspect, either financial or nonfinancial, of one or more organizational functions related to strategic goals.”

Although sometimes used interchangeably with “performance measurement,” the term “performance management” has been used in the literature to denote the act of actually using the information provided by performance measurement activities for organizational decision-making regarding maintaining or improving performance. Specifically, Slizyte and Bakanauskiene (2007, p. 317) argue that performance management “uses performance measurement information to manage and improve performance and to demonstrate what has been accomplished.” In this chapter, *performance management* is similarly defined as “the process of using performance measurement information to maintain or improve organizational performance.”

Finally, the literature defines a “performance measurement system” as “the formal, information-based routines and procedures managers use to maintain or alter patterns in organizational activities. These systems focus on conveying financial and nonfinancial information that influence decision-making and managerial action taking. The recording, analyzing, and distributing of this information are embedded in the rhythm of the organization and are often based on predetermined practices and at preset times in the business cycle. These systems are designed specifically to be used by the managers (de Waal, 2002, p. 5). Thus, a “performance measurement system” includes the infrastructural elements used to support the performance management process, including methods used to gather and report performance information, although it is often taken to include the performance measures themselves. For the purposes of this chapter, a *performance measurement*

system (PM system) is defined as “the procedures and infrastructural elements—including personnel, computing resources, and equipment—used to gather and report performance measurement information for the purposes of supporting performance management, as well as the set of performance measures and goals.”

30.1.2 Why is Organizational Performance Measurement Important?

As previously mentioned, the goal of a PM system is to support the managerial control (decision-making) process within the target organizational system to maintain or improve performance. We can view PM as occurring within the iconic plan-do-study-act (PDSA) cycle of continuous improvement (Figure 30.1). Managers first use PM system information to evaluate the current situation and determine what course of action to take (*plan*). They must then oversee the implementation of this course of action, and it is also beneficial to track progress on implementation using the PM system and other tools (*act*). Following or potentially during action, managers will then use PM system outputs again to assess the extent to which the action appears to be effective; if the action is not as effective as planned, the organizational will seek to understand why (*study*), generating knowledge for further action. The action can then be adjusted, replaced, or discontinued as needed (*act*). This type of decision-making should be built into the rhythm of the organization through meetings held at regular intervals as part of the performance review process, which will be discussed in more detail later in this chapter.

The types of organizational functions supported through this cycle of decision-making using the PM system include

- strategic planning,
- problem identification,
- process improvement,
- forecasting,



FIGURE 30.1 PDSA cycle of continuous improvement.

- individual performance evaluation,
- organizational learning.

In particular, as indicated, effective use of a well-designed PM system can continuously improve overall organizational performance over time by increasing organizational learning in several different areas, including understanding of

- current performance versus goals;
- areas of competitive advantage;
- opportunities to improve performance;
- priority of performance objectives among various stakeholders;
- stakeholder beliefs about how performance is created (mental models of performance, which can be captured through the strategy mapping process);
- appropriate strategies for improving performance.

The types of performance improvements realized through this cycle of continuous improvement include significant improvements in

- *Overall financial performance* (Lingle and Schiemann, 1996; Scott and Tiessen, 1999; Hoque and James, 2000; Malina and Selto, 2001; Davis and Albright, 2004; Martinez and Kennerley, 2005; Crabtree and Debusk, 2008; Iselin et al., 2008; Gimbert et al., 2010);
- *Industry standing* (Lingle and Schiemann, 1996; Martinez and Kennerley, 2005);
- *Customer satisfaction* (Martinez and Kennerley, 2005);
- *Employee learning* (Martinez and Kennerley, 2005);
- *Employee accountability* (Martinez and Kennerley, 2005);
- *Organizational alignment* (Meekings, 2005);
- *Teamwork* (Martinez and Kennerley, 2005; Meekings, 2005).

Clearly, a well-functioning PM system can provide many benefits for an organization, including the ability to continuously improve itself to maintain or increase performance to survive in today's competitive industry. For this reason, many organizations have recently attempted to redesign their measurement systems to follow state-of-the-art guidelines on design, implementation, and use. Yet, as will be discussed in more detail later in this chapter, the majority of PM system change efforts fail (de Waal and Counet, 2009). Thus, careful upfront planning and a well-thought-out execution strategy are needed before an organization attempts any changes to its existing PM system. Yet, if the organization is willing to put in the work upfront, the benefits can far outweigh the costs.

30.1.3 What are the Key Components of a Performance Measurement System?

Because achieving PM system change is such an inherently difficult task in any case, it is helpful if the organization realizes that there are two key parts to any PM system. Although these parts must be addressed in an integrated fashion, they differ significantly in terms of content and types of organizational issues faced in attempting to change each of them.

First, a PM system contains a prespecified set of performance measures designed to be used to maintain or improve organizational performance. This set of measures must be carefully designed to give a holistic and actionable view of organizational performance. This includes selecting the measures themselves and also their portrayal mechanisms. Sufficient balance across key dimensions of organizational performance must be obtained, while also preserving a manageable set of measures and actionable portrayal design.

Second, in order to achieve its objectives, as previously discussed, a PM system must actually be used to manage organizational performance. Achieving effective PM system use implies the deployment or implementation of the system procedures and infrastructural elements, a commitment to continuously assess organizational performance, and the actual use of the system to make new decisions and follow-up on old ones, particularly through structured and regular performance review processes.

Thus, the discussion of PM systems can be separated into issues related to the *design* of the PM system and issues related to the *implementation and use* of the PM system. We now discuss each of the two components in turn.

30.1.4 How Should Organizations Design Performance Measurement Systems?

Although the focus on implementation and use of PM systems has been growing in recent years, the design of PM systems has still received the most attention to date in the PM community, including the research literature. A major emphasis of this work on PM system design has been on determining a balanced set of the measures organizations can use to measure performance. Until the latter part of the twentieth century, most organizations' PM systems focused on financial measures, such as return on investment (ROI), market growth, and share price, as well as financially based operational measures, for example, labor cost, scrap cost, productivity, and efficiency. While these measures may seem intuitive, as they are the "bottom-line" metrics most of interest to managers and shareholders, they pose several significant problems when the organization attempts to solely use them to maintain or improve organization performance. They only provide a partial view of organizational performance and are almost always lagging in time—that is, the performance level displayed at any point of time is often due to events which happened for weeks, months, or even years earlier. Thus, a focus on financial and operations metrics only often leads to inaccurate assumptions about an organization's competitive position and faulty decision-making due to lack of understanding of cause and effect relationships that link organizational performance with strategic thinking.

To illustrate this truth using a simple example, suppose a manufacturing department manager is given the objective of cutting costs within the department over the next 6 weeks. Due to the tight time frame and the fact that the options appear to be limited, the manager decides to eliminate 9 finished goods inspection operation in one of the manufacturing lines, despite a historical scrap rate of 2%. The initial impact of this decision is a reduction in operating expense and a reduction in scrap costs thus increasing departmental and organizational profit (total income – total expense)—and the department manager is applauded by the superiors. However, several weeks later, the organization begins to see a significant increase in the return of defective products, eroding the profits gained. Even worse, the organization's sales department begins to report that several long-term customers are failing to place new

orders. By the end of the year, the organization's profits have declined below baseline levels, the organization is dealing with a myriad of customer complaints about quality, and the manufacturing manager has been fired.

Although it may appear obvious in this simple example that the decision to cut the inspection operation without any other actions to improve quality is almost certainly a bad decision, this example illustrates the types of problems encountered when the organization focuses only on financial and operational metrics. In particular, the lack of other types of measures, in this case, related to quality or customer satisfaction, makes it difficult to detect poor decisions until they are ultimately felt—sometimes much later—in the financial measures themselves. Even when observed, it can be difficult to detect exactly what chain of events led to the degradation of the measures. Indeed, typical limitations of financial measures include (Neely, 1999)

- They encourage short-term thinking: that is, the tendency to delay capital investments and other decisions that will harm short-term returns, even if they improve the performance of the organization in the long run.
- They are historically focused: as mentioned, financial measures almost always focus on the past. Even as recently as 2007, Brown estimated that 75% of *all* measures used in organizations—including almost all financial measures and the majority of operational and employee metrics—were historically focused. A target is to have only one-third of measures focused on the past (Brown, 2007).
- They encourage suboptimization: financial measures may be more likely than other measure types to encourage performance at the organizational subunit level that decreases overall organizational unit effectiveness; for example, a manufacturing department may build more inventory than is required by the customer to maximize productivity and utilization measures, leading to increased inventory holding cost, increased defects or obsolescence, and possibly decreased selling prices.
- They typically are not linked to organizational strategy: at best, financial indicators serve as end-result metrics indicating when the organizational strategy has been achieved; they typically offer little insight into organizational progress toward achieving strategy and, even in the best case, are not linked to strategic performance dimensions beyond financial viability, for example, quality, flexibility, or responsiveness.
- They fail to provide information on customer needs (typically available through customer satisfaction or loyalty measures) or other key performance dimensions (such as quality): as a lagging indicator, the best financial measures can do is indicate that customers *are* buying from the organization; they do not indicate *why* they are buying or *how well* the organization is meeting customer needs.

Thus, there is clearly a need for a carefully designed set of performance measures that gives a holistic view of organizational performance, and allows the definition of causal relationships between measures to identify leading and lagging indicators of performance and other important relationships, for example, correlations between different lagging indicators which are driven by the same leading indicators. In addition, a structured approach to PM design B needed to ensure that the measurement system is aligned with the organization's strategic goals.

30.1.5 How Should Organizations Design More Effective Measurement Systems?

Perhaps, the most popular modern measurement framework is the balanced scorecard (BSC), introduced by Kaplan and Norton (1992). The BSC has been named one of the top 15 management tools of recent years (Andersen et al., 2001), and it is estimated that 70% of organizations worldwide have attempted to adopt the BSC (Rigby and Bilodeau, 2007). Specifically, the premise of the BSC is that, in order to fully understand its operations, an organization must design measures to balance performance in four key dimensions

- *Financial*: indicators of fiscal success to shareholders and other evaluators. Example metrics include traditional financial measures such as profitability, share price, and market growth.
- *Customer*: measures of the extent to which the organization meets the needs of the people who use its products or services. Example metrics include measures of customer perceptions (such as satisfaction) and behavior (such as loyalty or repeat visits).
- *Internal business process*: metrics capturing operating efficiency and effectiveness in creation of products or services. Example metrics include productivity, quality, and process lead time measures.
- *Learning and growth*: indicators demonstrating how the organization maintains its ability to innovate in a dynamic environment. Example metrics include measures related to employee satisfaction and learning, as well as measures related to process innovation (number of suggestions implemented) and product innovation (new product launch rate, number of patents granted).

Beyond merely measuring performance in the four key dimensions, however, the BSC approach specifies that organization must also seek to understand causal relationships across metrics both within and across the four key dimensions, which is the focus of the next section.

Meanwhile, Table 30.1 summarizes some other important PM frameworks (note, this list does not attempt to be exhaustive, but rather to highlight common themes and differences across existing frameworks). Despite the great variety in some individual elements, a set of common characteristics across existing PM frameworks can be identified. In general existing frameworks: (1) identify multiple key dimensions of business performance; (2) attempt to provide a balanced representation of performance across these dimensions; (3) attempt to provide a comprehensive representation of organizational performance across these dimensions; (4) attempt to enable discernment of the organization's overall performance through examination of the measures included; (5) allow integration of measures both vertically and horizontally across the organization; and (6) assume and allow testing of cause and effect relationships between measures (Kennerley and Neely, 2002).

It should be noted here, even if the organization does adopt a holistic measurement framework, there are a myriad of other problems that might be encountered with the PM system. Several of the most commonly documented problems with PM systems in organizations include

- There are too many measures: organizations may attempt to measure and control everything, leading to information overload and many trivial or nonactionable measures; and/or, organizations may simply neglect to delete measures as they add new

TABLE 30.1 Other PM Frameworks from the Existing Literature

Framework	Performance Dimensions
Tableau de Bord (Lebas, 1994)	Corporate/HQ Divisions Functions/Departments
Performance Measurement Matrix (Keegan et al, 1989)	External cost, for example, competitive cost position, R&D expenditures External noncost, for example, market share, customer loyalty (repeat customers), customer satisfaction (number of complaints) Internal cost—for example, design, material or manufacturing cost Internal noncost, for example, cycle time, service level, new product introductions
Sink and Tuttle Framework (Sink and Tuttle, 1989)	Effectiveness—output produced (“doing the right things”), for example, motors shipped, tons produced, rooms rented Efficiency—input resources consumed (“doing things right”), for example, labor hours, material Productivity—ratio of output/input, for example, tons produced/labor hour Quality—extent to which materials, work in process, finished goods and service support end customer expectations regarding goods and services; measured at several points in supply chain—from suppliers of raw materials to end customer satisfaction Quality of work life—response of employees to working conditions, for example, absenteeism or employee satisfaction Innovation—organizational ability to adapt to the changing environment, for example, number of improvements implemented, number of patents awarded Profitability/Budgetability—ratio of income to expenses; measured through standard financial measures
Five Critical Success Factors (CSF) (Beischel and Smith, 1991)	Quality Customer service Resource management Cost Flexibility
Performance Measures for Time-Based Companies (Azzone et al., 1991)	Cost Quality Innovation Time
Performance Pyramid (Lynch and Cross, 1991)	Overall corporate vision (corporation) Market (business units) Financial (business units) Customer satisfaction (business operating system) Flexibility (business operating system) Productivity (business operating system) Quality (department and work centers) Delivery (department and work centers) Cycle time (department and work centers) Waste (department and work centers)

TABLE 30.1 *(Continued)*

Framework	Performance Dimensions
Results and Determinants (Fitzgerald et al., 1991)	Competitiveness (result) Financial performance (result) Quality (determinant) Flexibility (determinant) Resource utilization (determinant) Innovation (determinant)*
Advanced Manufacturing Business Implementation Tool for Europe (AMBITE) (Bradley 1996)	Time Cost Quality Flexibility Environment At least one strategic performance indicator should be identified from each performance dimensions for each of five key “macro business processes”: customer order fulfillment, vendor supply, design co-ordination, coengineering, and manufacturing
Integrated Performance Measurement System Reference Model (Bititci et al., 1998)	System 1—core production system System 2—manages core production system System 3—monitors and sets goals for system 1 and system 2 System 4—monitors external environment and sets goals for improvement System 5—highest level of organization, which sets corporate vision and strategic direction
Performance Prism (Neely et al., 2002)	Stakeholder satisfaction—what stakeholders demand (need and want) from the organization Stakeholder contribution—what the organization demands (needs and wants) from the stakeholders Capabilities—the resources required to satisfy stakeholder and the organization’s needs and wants Strategies—the policies required to satisfy stakeholder and the organization’s needs and wants Processes—the procedures required to satisfy stakeholder and the organization’s needs and wants
Dynamic Multidimensional Performance (DMP) (Maltz et al., 2003)	Financial performance—traditional financial measures Market/customer—measures focused on understanding customer needs Process—metrics related to organizational efficiency and improvement People development—metrics related to employee skills and commitment Future—metrics related to strategic alliances and partnerships, and the development of new product and service offerings
Integral Framework for Performance Measurement (IFPM) (Rouse and Putterill, 2003)	Long-term dimensions: Stakeholder expectations—long-term desires of stakeholders Contributions—resources provided by stakeholders Benefits—long-term value provided to stakeholders Vision/goals of the organization—adopted to meet stakeholder expectations Mid-term dimensions:

(continued)

TABLE 30.1 (Continued)

Framework	Performance Dimensions
European Foundation for Quality Management (EFQM) Excellence Model (EFQM, 2011)	Organizational culture/structure—driven by stakeholder expectations
	Resource capacity—of organization, based on the contributions of stakeholders
	Strategic outcomes—outcomes of mid-term activities, goal is to drive long-term benefits
	Objectives—strategic objectives of organization in pursuit of its vision and goals
	Near-term dimensions:
	Evaluation—process of tracking organizational progress toward meeting stakeholder expectations
	Resource utilization—near-term utilizations of available resources
	Outcomes—impact of near-term activities toward achieving desired near-term strategic outcomes and long-term benefits
	Plans—near-term operating plans of organization
	Short-term dimensions:
	Measurement—short-term control processes of the organization
	Inputs—resources used in production activities in short-term
	Activities—processes used to produce short-term outputs
	Outputs—products or services produced in short-term, to ultimately drive desired benefits
	Performance norms—short-term strategies driven by near-term plans
	Customer results, for example, customer satisfaction measures
	People results, for example, employee satisfaction measures
	Society results, for example, ethics and environmental responsibility
	Key results—the most important financial and nonfinancial indicators of the organization, potentially including results from the other three categories
	Leadership (enabler)
Malcom Baldrige National Quality Award (MBNQA) Results Category Dimensions (NIST, 2011)	People (enabler)
	Policy and strategies (enabler)
	Partnerships and resources (enabler)
	Processes (enabler)
	Product and process outcomes, for example, supplier and partner results, operational and process performance
	Customer-focused outcomes, for example, customer satisfaction and customer loyalty
	Workforce-focused outcomes, for example, employee well-being and satisfaction, safety, work system effectiveness
	Leadership and governance outcomes, for example, measures of ethics and stakeholder trust in organization, energy consumption
	Financial and market outcomes, for example, standard financial measures

ones (Neely, 1999); either or both of these forces can ultimately lead to dozens, hundreds, or even thousands of measures supposedly in use.

- Measures are not reliable and/or valid: many metrics suffer from reliability and/or validity problems. This is especially true for subjective measures, for example, customer satisfaction, employee satisfaction, when designed by people who do not possess expertise in the design of psychological measurements using survey questionnaires (psychometrics).
- The system lacks effective visual and portrayal tools: portrayal tools should promote statistical thinking about trends over time, and other fact-based decision-making; where possible, a standardized format should also be used across metrics (De Waal, 2002).
- The measurement system is not fully deployed throughout the organization: for instance, Brown (2007) reported that the majority of balanced scorecards exist only at senior management levels.
- Targets are set arbitrarily: that is, targets are determined based on “gut feel,” management desire or institution only without any systematic comparison to competitors, industry averages or benchmarks (Brown, 2007).
- Performance measurement software is not exploited and some software is poorly designed (too complex, etc.) (Brown, 2007).
- Measures are tracked/reviewed too infrequently: for example, annual customer satisfaction survey—and data collection is too cost intensive (Brown, 2007).
- Measures drive the wrong behaviors: this includes any behavior not aligned with organizational goals and the actual intent of the PM system, such as the functional/departmental suboptimization previously discussed. However, the PM system can also encourage wrong behaviors at the top of organizations—for instance, executive bonuses are still tied to primarily to financial metrics (Brown, 2007).
- Measures not used to support decision-making: this may be linked to the previous problems, but also to the lack of understanding of causal linkages between measures. This also occurs when an organization focuses on measures that are not within its sphere of influence or control, or when metrics to do not vary sufficiently (i.e., the organization primarily collects data on “what it already knows” (Brown, 2000)). Conversely, however, organizations should not go too far in measuring only controllable factors—important external factors that influence performance should be included but should not comprise the majority of the scorecard (Brown, 2007).

Some of these problems—particularly measures that drive the wrong behaviors—can be at least partially alleviated through the identification of causal relationships between measures, as discussed in the next section. However, others require additional steps in the process of designing and implementing/using measures, as will be discussed later in this chapter.

30.1.6 How Should Organizations Identify Causal Relationships between Measures?

As previously indicated, all of the PM frameworks above assume, either explicitly or implicitly, that balancing PM across multiple dimensions is critical not only because

organizational success has different facets, but also because the dimensions are inter-related. Thus, one key to the creation of effective performance measurement systems is developing organizational processes that document and then empirically test causal relationships between metrics.

One tool that can be used to support such processes is a *strategy map*, which is a graphical depiction of the hypothesized interrelationships between the performance measures in the measurement framework adopted by the organization. A strategy map may also include: key performance areas; strategies adopted to improve performance in each dimension/key area; specific performance measures for each dimensions/key area; shorthand versions of action plans adopted to improve performance; and targets and timelines for metrics, improvement strategies, and action plans. Other common terms for a “strategy map” include “business model,” “causal model”, or “cause-and-effect map” (Chearskul, 2010).

A simplified strategy map adapted from Kaplan and Norton (1996) is presented as Figure 30.2. This represents only a subset of a map and does not contain all necessary elements. However, it demonstrates the basic construction of a strategy map and typical hypothesized relations between the BSC dimensions—that is, that learning and growth investments influence internal business process performance, which influences customer satisfaction, which in turn influences financial outcomes. Along with logical support, there is some empirical support for this hypothesized causal chain as well. One study conducted in Sears found that a 5% improvement in employee attitude was linked to a 1.3% increase in customer satisfaction, which was linked to a 0.5% increase in store revenue (Rucci et al., 1998).

Although the simplified map in Figure 30.2 includes only the relationships between performance dimensions and specific metrics, as mentioned, a mature strategy map would typically also include specific strategic initiatives related to the metrics (e.g., “develop a fan base of repeat customers”). The map could also include specific targets for the metrics (e.g., “increase customer loyalty by 10% over last year”) and potentially also for metrics

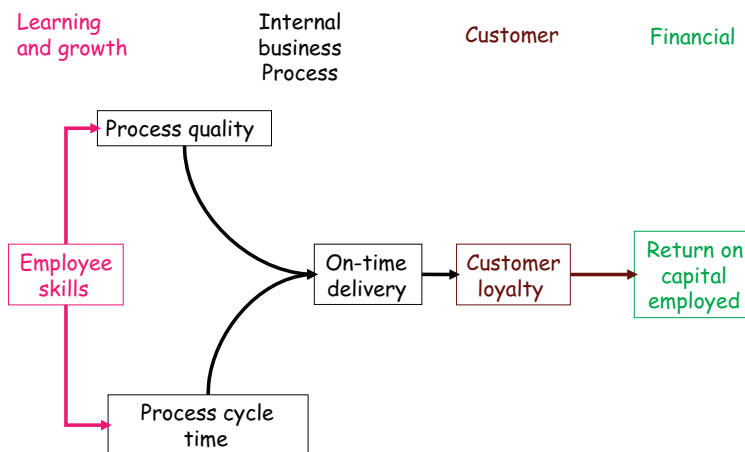


FIGURE 30.2 Simplified strategy map highlighting relationships between BSC dimensions. (Adapted from Kaplan and Norton, 1996, p. 66.)

related to strategic initiatives (e.g., “develop a fan base of at least 1000 repeat customers”), short-hand versions of action plans (e.g., “double the Facebook advertising budget and start a Facebook fan page”) and time frames (including potential time delays for seeing impacts) (e.g., 1 year).

Benchmarking, literature review, group brainstorming sessions, or historical data can be used to develop the initial strategy map for further testing. Once a strategy map has been defined, the organization must collect and analyze empirical data using its PM system to verify the extent to which the strategy map appears valid. Faulty assumptions regarding the linkages between performance dimensions can cause performance to stagnate or even worsen without the organization realizing the catalyst behind such effects. Although defining and testing a strategy map is clearly a challenging undertaking, studies have shown that this approach is necessary for achieving the most value out of the organization’s PM system (Ittner and Larcker, 2003; Chenhall, 2005). Several examples of approaches for exploring and testing causal relationships between performance measures can be seen in Santos et al. (2002) and Sousa et al. (2005).

30.1.7 What Process Should an Organization Use to Design its Performance Measurement System?

Just as many different frameworks have been proposed for defining a balanced set of measures, many different approaches to the actual PM system design process have also been identified. Some are detailed, while others are very broad. For instance, Bourne et al. (2000) identify two key parts to the PM system design process: (1) identifying the key objectives to be measured and (2) designing the measures themselves. One general process for designing PM system, the Measurement System Development Process (MSDP) (see Van Aken and Coleman, 2002), which is based on the more general TransMeth transformation methodology (Van Aken et al., 2003), is illustrated here

1. (in parallel with other tasks) create the infrastructure for the PM system design, implementation and use effort,
2. define the need for PM within the target organizational system,
3. define what the target organizational system does now and should do in the future,
4. define key performance areas in which the target organizational system must excel now and in the future,
5. define how the target organizational system will know if it is successful in its key performance areas,
6. implement the PM system,
7. utilize the PM system.

The first five steps will be discussed in this section, whereas the last two are discussed in the implementation/use section later in this chapter.

30.1.7.1 Create Infrastructure for the PM System Design, Implementation and Use Effort This step focuses on selecting the design team, communicating the results of other steps to various stakeholders within (and potentially external to) the organization,

acquiring needed resources, and providing necessary training. This is an ongoing process that occurs in parallel with the other steps.

The first thing for the organization to do will be to identify who will champion, lead and otherwise participate in the PM initiative.

The project champion, or sponsor, is usually a senior level manager within the target organization who provides guidance to the leader/project manager in terms of the vision for the project, internally links the project to the need for change within the organization, and helps the leader/project manager acquire any needed resources (personnel, budget, etc.). The champion also acts as a liaison between the leader/project manager and other members of the senior management team.

The leader or project manager is the person or persons who manage the actual PM system project, including negotiating for needed resources, scheduling activities, managing any relevant budget, managing the creation of deliverables, and reporting back to the project sponsor and other relevant stakeholders. To stimulate ownership and accountability, it is highly recommended that the same individual manages all the stages of the PM system project—from design to implementation and initial use of the PM system (“roll out”)—although this is not always possible. In addition, it is highly recommended that the leader should be internal to the organization, although in practice (particularly during the first stages of the PM initiative) it may be an external consultant working with internal liaisons, including the project champion.

In addition to the leader/project manager and the project champion/sponsor, it is essential to define a small team to assist the project manager throughout the various stages of the initiative. Members of this team can either be internal, external, or a mixture to the organization. In terms of rolling out the metric design process, one of the first decisions the organization will face is where in the organization to begin the rollout of the project management efforts. The literature suggests that PM system change can be successfully implemented using either a top-down rollout (central office first, deployed to work groups), bottom-up deployment (implemented first in one or more work groups or departments, then higher in the organization), or some combination thereof (e.g., implemented simultaneously in the central office and one or more workgroups) (Andersen and Faguerhaug, 2002; Letens et al., 2010). However, it does appear critical at this stage that the organization finds some way to control stakeholder involvement such that it is not soliciting active—or at least in depth—involvement from everyone in the organization. While open communication channels can be desirable for gathering input beyond the employees in the rollout group, attempting to involve too many stakeholders at once or to gather detailed input from everyone in the organization can lead to the failure of the initiative. Also, attempting a simultaneous rollout of the PM system to all areas of the organization appears to be destined for failure. It appears the organization should first select one or more smaller areas as test beds, implement, observe results, adjust and then continue to deploy the system across the organization. Thus, the most promising method appears to be to involve only a few key staff in the target areas, including key decision makers and a few other critical stakeholders, in the process of designing the metrics, while holding broader “all hands” meetings in the target area and other areas of the organization for information and buy-in purposes.

Once the leader, champion, and design team have been identified, a stakeholder analysis can be a useful tool to determine exactly how to communicate with the larger organization. The specific elements considered may differ based on the organization’s objectives, but the basic goals of a stakeholder analysis are to identify

- all relevant stakeholders (i.e., those who are at least affected by or interested in project outcomes);
- each stakeholder's influence on the project (e.g., has useful expertise, has decision authority);
- based on the two above relationships, how the project team should communicate with each stakeholder (from regular meetings to as needed); this analysis might also be used by the project sponsor or project manager at the beginning of the project to determine who should be targeted to participate in the project team.

Table 30.2 shows an example stakeholder analysis matrix for a hypothetical PM system project within a manufacturing department.

Needed resources initially include the PM system design/implementation team members, as well as any resources needed to support: team meetings, data collection, data analysis, actual metric/portrayal/system design, project planning, and communication to other stakeholders, among other processes. Later, necessary resources may include computing resources and personnel to support the implementation and ongoing use of the PM system.

Initial training will likely be focused on the design/implementation/use team and members of senior management. Both senior management and team members need to receive training in the basic concepts and benefits of PM systems. It will be useful for both team members and senior management to also receive additional training in group dynamics, the steps of the design/implementation/use process, and the PM systems of other companies. More detailed training can also be provided at any time during the process. As the PM system gets deployed to other organizational units, more training will be needed for the additional users of the PM system.

TABLE 30.2 Example Stakeholder Analysis for a Hypothetical PM System Project

Stakeholder	Relationship to Project						Communication/ Involvement Strategy						
	Is Affected by PM System	Can Influence PM System	Will Use PM System	Has Useful Expertise	Provides Resources	Has Decision Authority	Project Sponsor	Project Manager	Project Team member	Meet with Regularity	Invite to Team Meetings	Send Copy of Team Meeting Minutes	Speak with Informally as Needed
Department manager	X	X	X	X	X	X	X			X	X	X	X
Assistant department manager	X	X	X	X				X		X	X	X	X
Supervisors/other departmental senior staff	X	X	X	X					X	X	X	X	X
Lower level departmental personnel	X	X								X			X
Department customers	X												X
Organizational senior staff	X	X		X	X	X						X	X

30.1.7.2 Define the Need for PM in the Target Organizational System This step is focused on clarifying the objectives of the PM system, as well as who will play what role in developing, implementing and using the system. This includes

- linking the need for measurement with the general need for change in the organization;
- defining the overall purpose and impact of the PM system—that is, what questions are we trying to answer with the system and what is the expected impact of having these questions answered?
- defining the general types of information needed to answer these questions;
- identifying the users of the measurement system and other stakeholders, such as the sponsor, project manager, and design team;
- defining user expectations for the system design, for example, extent to which it is user-friendly, preferences about how information is portrayed, frequency needed to support decision-making, and how much time users expect to spend in reviewing information;
- identifying how the PM system development relates to any other ongoing improvement initiatives;
- communicating this information to relevant stakeholders.

The need for the PM system may often be identified initially by one or a few organizational personnel, who, if not members of the senior management team, may need to pass the idea up the chain of command until it gains traction.

30.1.7.3 Define What the Target Organizational System Does Now and Should Do in the Future Although the details of the various design methodologies differ, there is consistent agreement that the key objective of a PM system is to assist the target organizational system in achieving its strategic goals. In order to design a PM system that is aligned with the target system's strategic goals, this requires first identifying

- What the target system is: that is, the primary suppliers, inputs, processes, outputs, and customers of the system. This can be assessed using an input/output analysis, also called a supplier input process outcome customer (SIPOC) diagram, or similar tools. Other system models of organizations that can be instrumental in identifying organizational strengths and weaknesses include the Management Systems Model (MSM) (Kurstedt, 1985), the Malcolm Baldrige Criteria for Performance Excellence (NIST, 2011), or the European Foundation for Quality Management (EFQM) Excellence Model (EFQM, 2011).
- Why the target system exists and what value-added it provides to its customers: that is, what is the purpose of the organization, as captured in a unique mission statement. If a mission statement does not exist, the PM system development team will likely want to draft one at this time, with input from relevant stakeholders, especially the senior manager of the organizational system. However, care must be taken as creating a mission statement can easily become a controversial and lengthy task—particularly if many different organizational stakeholders are involved. Thus, some experts recommend involving only the senior manager of the organizational unit in writing

the mission statement (Brown, 2000), which can then be presented to other stakeholders for approval and fine-tuning.

- Where the organizational system wants to go in the next 3–10 years: this can be captured in narrative form through a vision statement. Here, a useful analysis tool is a strengths–weaknesses–opportunities–threats (SWOT) assessment. The SWOT attempts to identify the positive forces (internal strengths and external opportunities) and negative forces (internal weaknesses and external threats) affecting the organization. The identification of these forces can then be used as a key input to craft a vision of where the organization should go in the next 3–5 years. Other tools include environmental scanning, analysis of competitor mission and vision statements, market analysis (e.g., through Porter’s (1979) five forces model), and feedback from customers. It is recommended that members of the senior management team participate in creating ideas for the vision, but, as with the mission statement, a single individual, typically the senior manager, draft the actual vision statement and then present this draft to the rest of the management team for approval/refinement (Brown, 2000). Group creativity and brainstorming tools, such as analogies, tree diagrams, and pictorial representation of the future, can be used to help generate ideas for the vision statement. After identifying the narrative form of the core vision, the organization should then identify specific strategies to achieve the vision and qualitative or quantitative goals associated with these strategies. These in particular can be useful for identify key performance areas to be monitored, and appropriate targets for these metrics.

30.1.7.4 Define Key Performance Areas in Which the Target Organizational System Must Excel Based on the definition of the components of the target organizational system, its mission and its vision, the next step is for the organization to identify a “small number” of key performance areas (KPA) (typically three to seven), in which it must excel to achieve its mission and vision. A KPA describes a core dimension of performance that encompasses a number of potential measures, for example, customer satisfaction, and operational excellence. KPA form the core structure of organizational scorecards and strategy maps/business models. Similar terms include “critical success factor” (CSF) and “key results area” KRA. A good set of KPA will be succinct (“small number” of KPA), clearly aligned with the mission/vision, balanced, comprehensive, and unique to the target system. The set of KPA should also be reevaluated at regular intervals, typically every 3–5 years, for its continued applicability, and the KPA used should change as needed to align with changes in the mission, vision or external environment (e.g., regulatory requirements). While the organization may want to customize the number and names of dimensions to suit its unique environment, the performance measurement frameworks presented earlier—and particularly the BSC—are a starting point for identifying KPA. Identifying KPA should typically be a team process (Brown, 2000), which involves the project manager and design team members, with feedback from the sponsor and potentially other stakeholders.

30.1.7.5 Define How the Target Organizational System Will Know If It is Successful in Its Key Performance Areas Once the key performance areas have been defined, the organization should seek to identify a set of measures to evaluate each KPA. In particular, it is important to identify at least one outcome or end-result metric related to each KPA. An end-result metric is a lagging indicator of performance, measuring current—or, more

likely, recent past—success. An example of a lagging indicator may be purchases/customer. Although one is the minimum recommended number of end-result metrics, typically one to three is the desired range for each KPA. The organization should also seek to identify at least one driver metric related to each end-result metric. The driver metric is thus a leading indicator, which is believed to influence at least one end-result or lagging indicator. For instance, manufacturing batch size might be a driver metric for the end-result metric manufacturing lead-time. Again, the typical range would be one to three driver metrics per end result metric, although some driver metrics may be related to more than one end-result metric, and some end-result metrics may be drivers for others (e.g., customer satisfaction might drive purchases-per customer). In general, the PM literature seems to agree that 10–20 metrics in total appears to be an appropriate and manageable number for most organizational scorecards. However, an organization should not consider this a definite rule if more (or fewer) metrics appear appropriate for its specific case. As with the identification of KPA, identifying metrics should typically be a team process, involving the project manager and the design team. A good place to start is with an audit of existing metrics. However, the organization must be careful to avoid the temptation of “settling” for a subset of the existing metrics as its new PM system, merely because the existing metrics are already available. This is particularly true for organizations in the stage of refining their first (initial) PM system design efforts, where there is a tendency to identify too many measures. Thus, the existing number of metrics may be greater than the suggested 10–20, and the tendency is to merely cut down this set to an existing set of financial or operational control metrics without adding any new strategic improvement metrics. Organizations starting their efforts “from scratch” are likely to have few existing metrics beyond financial metrics and other required for regulatory purposes. For adding new metrics, the organization can review the metrics of competitor/comparison organizations or other benchmarks (e.g., quality award winners), review suggested sets of measures for different performance dimensions, or define entirely new measures unique to the organization. New measures for which data already exist (but simply have never been previously compiled into the metric as defined) are obviously the most tempting. However, the organization should not overlook more meaningful measures simply because the data are not available. Instead, the overall quality of each metric should be analyzed, for example, using the bullet point list below, to identify which among alternate metrics are most preferable for the system.

Finally, although not counted against the total number of metrics, it also may be useful for the organization to identify drill-down metrics for some or all end-result metrics, particularly those that are composite or index variables (Brown, 2007). For instance, an organization may have scrap rate as one of its end-result metrics within an *Internal Processes* KPA. To better understand changes in scrap rate, the organization may want to drill-down this metric into different categories of defects, to better understand the reasons for the change and specific areas (e.g., types of defects) to target for improvement. A drill-down metric therefore allows examination of the performance of different subcategories of the end-result variable and is different from a driver metric, which is a separate factor that causes changes in the end-result metric (and at least some of its subcategories). For example, employee skill level, machine maintenance audit score, and supplier quality could be driver metrics for defect rate. Meanwhile, subcategories of defect rate (drill-down metric) could include inadequate paint job, misaligned components, or failure to meet minimum functional performance standards during tests. An additional example of driver versus

end-result metrics can be found in the description of the organizational scorecard for the case study in Van Aken et al. (2003).

Similar to a lot of areas in PM system design, there are no hard and fast rules for metric design. However, from an examination of the literature, it appears that each metric should be

- *Clear*: each metric must be operationally defined (that is a formula must be developed that describes how the metric values are calculated) and the meaning of the metric should be readily understandable based on this operational definition (i.e., not obscure).
- *Aligned*: each metric must be aligned with the KPA and the vision and mission; further, the metric must be carefully selected such that it does not unintentionally drive the wrong behaviors; for example, use “first call problem resolution” instead of “talk time,” as a measure of operational excellence in a call center.
- *Reliable*: the method for collecting data and calculating the measure should be repeatable and precise, such that there is a minimum amount of measurement error in the metric. Note that measurement errors are particularly likely in the data collection stage (either the input data itself is faulty or the data collection process is unreliable), as calculations are often automated or semiautomated. However, errors can still occur at the calculation stage. Reliability is particularly a challenge for perceptual measures, such as customer or employee satisfaction.
- *Valid*: each metric must be recognized as a critical indicator of the KPA—one of the “vital few” versus the “trivial many.”
- *Variable*: measurement error aside, the underlying dimension being tracked must be likely to vary on a semiregular basis, and variation in this metric must have an impact on organizational performance. It is generally not worthwhile to track metrics that do not vary and, unless required for reporting purposes, it is hardly ever worthwhile to track a metric where variation does not noticeably impact organizational performance (either as its own performance dimension or through another metric).
- *User-focused*: metrics must be designed such that they are actionable (support decision-making) and, as much as possible, meet user preferences and expectations regarding portrayals, and so on.
- *Influenceable*: the majority of metrics must be within users’ sphere of influence or control, and it must be clear which individuals are responsible for which metrics (the limits of each users’ control should also be noted). Occasionally, it is permissible to include metrics that are not within users’ sphere of influence or control, if these are identified as key driver metrics or are needed for external reporting purposes. However, the fact that these metrics are not controllable for the organization must be noted in all communication, system documentation and metric portrayals. Individuals must never be held responsible for performance on metrics which have not been assigned to them or which are not within their sphere of influence or control, as this might lead to unacceptable or dysfunctional behavior, in particular when these performance metrics are linked to financial incentives or other rewards.
- *Cost-effective*: the cost of collecting data should not exceed the benefits of the metrics and, ideally, should be as low as possible to maintain the desired level of reliability.

TABLE 30.3 MSDM Template

Data Collection Plan						Utilization	
Metric	Tracking Tool	Data Available?	Data Collection Responsibility	Data Collection Tool(s)	Data Collection Frequency	Implementation Date	Metric Goal
Key performance area							
Key performance area							

Metric Specification				Portrayal Design		
Metric	Operational Definition and/or Formula	Purpose of Metric	Metric Owner	Portrayal Frequency	Type of Data	Portrayal Tool
Key performance area						
Key performance area						

As the metrics are developed, the required elements can be captured in a standard template, such as the metrics development matrix (MDSM) shown in Table 30.3—note that some elements here refer more to the implementation step than this design step.

Once the set of metrics has been designed, the organization should audit each metric using the criteria above (or similar criteria) to assess metric quality—and to improve the measurement system by refining, replacing, or removing individual metrics, if necessary.

The organization should then also audit the entire measurement system for balance across the different KPA to identify any potential gaps or KPA which are over-represented versus under-represented. Finally, the organization must observe the actual use of the metrics over time to assess the extent to which they are accurate, actionable, and drive the intended behaviors—as well as the extent to which the causal relationships hypothesized in the strategy map actually exist. The organization should also recognize that the set of KPA and measures which are most appropriate for its PM system will likely change over time, and must develop a system for regularly refreshing its PM system (Neely, 1999), as will be discussed in the next section.

30.1.8 How Should an Organization Implement and Use Its Performance Measurement System?

Despite significant potential benefits from successful implementation, the failure rate for PM system implementation is high—one study estimated that roughly 60% of PM system implementations fail, with exact rates depending on the context (de Waal and Counet, 2009). While deficiencies in the design steps, such as the selection of inappropriate KPA or measures, can certainly contribute to PM system failure, the literature suggests that the majority of failures are due to deficiencies in the implementation and use steps (Bourne et al., 2002; Braam and Nijssen, 2004; Franco-Santos and Bourne, 2005; Meekings, 2005; Meekings et al., 2009; Chearskul, 2010).

In addition to total failure of the system, if not carefully implemented, even a “successful” PM system can negatively impact the organization through the following effects, among others:

- system manipulation or “gaming,” for example, suboptimization (Propper and Wilson, 2003);
- information overload (Lipe and Salterio, 2000; Ittner, et al., 2003);
- high system cost (Halachmi, 2002; Ittner, et al., 2003; Marr et al., 2004; Tuomela, 2005);
- internal resistance to change (Tuomela, 2005; Martins and de Abreu, 2006).

As with PM design, there is no one accepted process for implementing a PM system, although various authors have proposed different approaches, with some common themes. The literature also typically separates “implementation” (the transition/adoption phase) from “use” (the perpetual, ongoing phase once initial adoption is achieved) in the discussion, as will be done here.

30.1.8.1 How Should an Organization Implement Its Performance Measurement System? As mentioned, the implementation step specifically refers to the transition phase where the organization is in the process of moving from current practice to the ongoing use of the new PM system. This includes three critical substeps

1. development of a change management plan;
2. development of a technology integration plan;
3. development of a plan for the regular performance review process (system use).

Developing the change management plan involves determining

- How and when to deploy (roll out) the actual changes associated with system implementation?
- What training is needed for people who use or support the system (e.g., through data collection or technical support)—and when and how this training will be provided?
- What other communication should be provided to stakeholders and how this communication will be delivered?

Deployment options will be influenced by previous decisions regarding the scope of the design/implementation efforts—that is, whether it is occurring simultaneously within many different areas of the organization or concentrated in a single area. In general, however, even if the design/implementation is occurring simultaneously in several different parts of the organization, it would be advisable to select one or two areas as pilot implementation areas, for the other areas of the organization to observe the results and the organization to fine tune the implementation process. The pilot area(s) should generally be those where the expected impact of the PM system is expected to be noticeable (to generate “short term wins” Kotter, 1996), and/or the area personnel are supportive of the PM system (or at minimum, open to change in general). If no area can be found that meets either criteria, the organization is likely not ready for implementation and the PM system design team, project manager and project sponsor should work on building support first. Similarly, it is also possible that the organization may want to deploy only part of its scorecard initially—in particular those measures that can be calculated with existing data. If possible, the organization should seek to have at least a few measures of this type on its scorecard initially—preferably balanced over the different KPA. They can always be refined or deleted with time.

Training should be provided to the individuals charged with collecting PM system data, analyzing PM system data, reporting PM system data in the performance review process, and supporting the PM system technology.

Training for individuals who collect PM system data should include training in

- the definition of the metrics within the PM system, their purpose, operational definitions, types of data required, sources of data collection, and potential errors in data sources;
- administration of any data collection tools, for example, customer satisfaction surveys, including possible errors in administration and frequency of administration;
- the recording/documentation of data, for example, in databases, spreadsheets, or directly into PM software;
- methods for retrieving PM system information;
- any other necessary aspects of interfacing with PM system software, if it exists.

Training for individuals who analyze PM system data should include the previous elements (except perhaps the details of the administration of data collection tools) plus training in any analysis tools or templates—whether built into PM system software or as stand-alone elements, for example, control charts, Pareto diagrams, regression, and simulation software.

Training for individuals who report PM system data for the performance review process should include training in

- the definition of the metrics within the PM system, their purpose, operational definitions, types of data required, sources of data collection and potential errors in data sources;
- an overview of data collection, recording, and analysis methods;
- methods for retrieving PM system information;
- preparation of reporting materials for performance review;
- any other necessary aspects of interfacing with PM system software, if it exists.

It is noted here that sometimes a single individual, often called the “metric owner,” may be charged with collecting, analyzing, and reporting PM system data for a given metric. (Note that an individual might be a metric owner for multiple metrics). However, the responsibilities might be separated into data collection/analysis versus reporting or data collection versus analysis/reporting, with the data collection-related responsibilities perhaps being conducted by a single individual within the target system (perhaps across all or at least multiple metrics) or an external department (e.g., information technology (IT) or accounting), and the reporting responsibilities being performed by a metric owner within the target system.

Finally, especially if a specialized PM software package is used, training may be needed for the individuals who manage the computers on which software used for the PM system is housed, to enable them to better manage the software and provide in house troubleshooting for other users.

Developing the technology integration plan includes determining what electronic data repository and electronic data-gathering mechanisms are required to support PM system implementation, and how these changes will interface with the organization’s existing IT system and other IT resources, including IT personnel and their capabilities (Leinonen, 2001; Franco-Santos and Bourne, 2005; Franco-Santos et al., 2007). This will likely require at least an informal audit of existing IT systems, with questions designed to gauge to what extent the current IT system and other IT resources are capable of the changes needed to support PM system implementation and what IT investments, if any, are needed to support PM system implementation. Key questions include

- What IT-related changes will be required to support PM system implementation?
- Is the current IT infrastructure capable of supporting these changes or will investment in additional IT resources be needed? (This includes personnel training as addressed above, as well as the potential hiring of additional personnel.) If so, what infrastructure resources are needed, and what is the associated cost and time frame for acquiring these resources?
- When exactly will the IT-related changes to support PM system implementation need to occur and at what cost?
- Are the above feasible given the needed implementation time frame? If not, what changes to the proposed PM system design/implementation plan will need to be made?

Finally, although the use of the PM system may to some extent evolve over time, to achieve the most benefit from its PM system and to decrease the likelihood of system failure, the organization should begin implementation with a plan for how the PM system will be used at regular, recurring intervals by a prespecified group of users to support business decisions. We refer to this as the performance review process. In general, in defining a structured review process, published guidelines suggest that it is important to determine appropriate answers to the following questions (Marr, 2006; Meekings, 1995, 2005; Neely et al., 2006; Meekings et al., 2009; Chearskul, 2010; Farris et al., 2011):

- At what organizational level(s) will each measure be reviewed and at what frequency? While a scorecard will be designed for a specific target system, different individuals within the target system may review the measures at different frequencies and some individuals may only review a subset of measures. Frequency depends on the needs of each organizational level and the role of the specific metric. However, typically it is recommended that the review occurs at least quarterly.
- Who will be the metric owner(s)? Will the metric owners collect the supporting data, or will the IT system or other individuals handle this step? What training is needed for metric owners?
- What methods will the owner(s) use to communicate performance results during performance review meetings and to other stakeholders? For example, some organizations have found it helpful to design standard graphical review templates that are used across all the different measures in the target systems scorecard (Farris et al., 2011).
- What statistical and other tools should be used to analyze current metric performance, and to what extent should analysis be done before versus during meetings? In general, a graphical representation of the metric trend over time and limits that denote the natural range of metric variation and the target for the metric are recommended for metric portrayal and analysis. Comparisons to competitors can also be added to this portrayal if it does not make the portrayal too busy, or as an additional chart.
- How should decisions be made regarding actions to improve organizational performance on metrics? What problem-solving tools or causal models, such as strategy maps, should support these decisions, who should be involved, and how should consensus be reached? It is also recommended that formal problem solving tools—for example, Pareto charts and correlation analysis—be used for drill-down analysis and other problem-solving efforts. Meanwhile, design of experiments and regression analysis can be used to verify hypothesized causal relationships. Structured group brainstorming tools, such as nominal group technique (NGT), can also be useful for the problem-solving process. Further, it is important that the organization emphasizes decision-making—a formal decision should be reached on each metric during each meeting—even if it is to do nothing at present and to continue to analyze/observe metric performance. Otherwise, it is easy for organizations to get caught up in the analysis and reporting cycle and to feel that they have achieved something through the identification of problem-solving efforts, without perhaps even realizing that they have failed to develop a clear plan for implementation (Farris et al., 2011).
- How should decisions and the effects of improvement actions be communicated to the rest of the organization? Documenting decisions and communicating them to the

rest of the organization is just as important as making them in the first place. Formal project manager tools can be used to document action plans and stakeholder responsibilities. A single individual from the performance review team, for example, a group secretary, might be charged with documenting group decisions and other highlights from the meeting, or metric owners might individually document the decisions and other findings related to their metrics, and these findings can then be combined and disseminated to the group as a whole in the full meeting minutes. Also, if the organization has developed a plan for communicating performance review information to individuals not involved in the meeting at regular intervals, for example, newsletters, visibility boards, regular updates to the PM software/organization's intranet, or briefing meetings, this will aid in communicating regular performance information to other stakeholders. As any of the stakeholders become involved in more specific initiatives to improve performance on a given metric or metrics, more detailed information (including any necessary training) should also be communicated to them under the leadership of the director of the initiative (who may or may not be the metric owner).

- How should the effect of improvement initiatives be tracked over time, and when should decisions be made to modify improvement initiatives? The organization should build the reporting on progress of improvement initiatives into their regular performance review process. However, it is likely that progress might need to be reviewed more frequently by the director of the initiative and his/her team, as well as possibly by the initiative sponsor. Although leeway is needed for individual manager judgment, the organization should also seek to identify what sort of warning limits may indicate when a modification is needed to the improvement initiatives. For example: How much time delay do we expect before we see performance improvements? What is the minimum acceptable performance improvement? Do we expect performance to get worse before it gets better? If so, how long will this worsened performance be tolerated and what is the maximum deviation allowed?

The organization should seek to develop, deploy, and continuously refine a comprehensive performance review plan addressing all the above elements as it implements its PM system. The initial performance review process may also need to be further refined in the use phase.

30.1.8.2 How Should an Organization Use Its Performance Measurement System?

Although it may be necessary to collect certain performance data purely for the purpose of reporting to external stakeholders, for example, for regulatory purposes or for the higher-level target system, for the most part, PM data that are not actively used at all to support organizational decision-making are ultimately worthless or, worse, detrimental to the organizational unit. As previously mentioned, this is a challenging prospect, with only the minority of PM system implementations achieving long-term success in ongoing use. In addition to the problems previously mentioned in the design and implementation sections, other challenges to achieving successful ongoing use include difficulty in effectively linking the PM system to: the strategic planning process, the organizational reward system, organizational learning mechanisms, and, in particular, broader organizational transformation efforts. However, if the organization is effective in achieving systematic, ongoing use of the PM system, the system can play an integrating role in overall organizational transformation (improvement) efforts. Specifically, effective PM system use can

increase employee learning and accountability, organizational alignment, and teamwork, which in turn leads to improved organizational performance.

In fact, one recent empirical study found support for this causal chain through certain PM use practices (Chearskul, 2010). Specifically, the study found that use of the PM system within organizations can be characterized as consisting of five distinct activities: monitoring, problem-finding, problem-solving, validating causal relationships, and validating improvement actions. The two most influential use variables both had positive impacts on performance: monitoring and problem-solving. Monitoring was found to directly impact nonfinancial performance, which in turn drives financial performance. Problem-solving was found to indirectly improve nonfinancial performance and financial performance by increasing organizational learning, specifically improving shared vision.

Meanwhile, the study also found that two variables (problem-finding and validating causal relationships) had smaller negative impacts on performance, while validating improvement actions had no impact on performance. Conversely to problem-solving, problem-finding was found to reduce shared vision within organizations, negatively impacting both nonfinancial and financial performance, although this effect was much smaller than the positive impact of problem-solving. Validating causal relationships had a smaller negative impact on financial performance by increasing team learning (another organizational learning variable), which, counter to the study hypothesis, was found to negatively impact financial performance. However, as the literature consistently supports the importance of validating causal relationships, it may be that there are delayed effects for this variable, which were not measured in the cross-sectional study. It also appears that the impact of team learning may be delayed or that too much emphasis on work team learning activities may reduce performance by taking too much time away from production activities.

The results from the above study, and others, suggest that organizations must carefully design and implement their PM systems to promote the right type of use activities, for example, monitoring and problem-solving, avoid focusing too much on potentially detrimental activities, for example, problem-finding, and carefully approach and study activities with effects which are not yet well understood, for example, validating causal relationships and validating improvement actions.

In order to increase the likelihood of success in long-term ongoing use, it is therefore clearly important that organizations follow the previous steps in achieving a sound design and initial implementation plan. However, the organization's efforts cannot stop there. Instead, the organization must seek to refresh and refine its PM system through regular assessment (either formal or informal) of PM system effectiveness (Bauer et al., 2004).

Although it contains elements that look beyond the actual performance measurement process, the improvement system assessment tool (ISAT) (Van Aken et al., 2005) is one tool developed to guide organizations through the process of auditing their PM systems to identify strengths and opportunities for improvement. Specifically, the ISAT uses scoring dimensions based on the MBNQA and EFQM evaluation approaches to determine the extent to which the organization demonstrates maturity in each of the three key aspects of its PM system: measurement system design (Table 30.4), measurement system development and implementation (Table 30.5), and results achieved through the PM system (Table 30.6). Measurement system design and measurement system deployment and implementation are scored using the four categories in Figure 30.3. Meanwhile, the results achieved are scored in terms of trends, goals, comparison, and causes (see Table 30.6). Each dimension is on an anchored scale of 0–100, using a format similar to the MBNQA template for scoring the results criteria.

TABLE 30.4 Assessment Template for the Measurement System Design Element within the ISAT

Scoring Dimension	Assessment Dimensions/Items	Examples of Supporting Tools
Approach	<i>Structured approach</i> for defining metrics <i>Cross-functional involvement</i> in defining metrics (e.g., leadership team, . . .) Use of group process <i>tools</i> to define metrics	Strategy mapping, BSC, Performance prism, and so on
Deployment	Metrics (scorecard design) <i>deployed</i> to lower levels if applicable Metrics clearly and consistently <i>communicated</i> Internal to the organization, promoting accessibility and real-time use External to the organization Metrics <i>cover</i> all critical Functions, processes, and work units in the organization	Visibility boards, email/intranet, newsletters, all-employee meetings, . . . Stakeholder meetings Audit of metrics deployment
Study	The scorecard has metrics that are <i>Focused</i> (the “vital few”) <i>Aligned</i> with Burning platform Vision and value-added Higher-level system Desired behaviors Reward system <i>Balanced</i> (across dimensions in multiple frameworks) . . . Proposed <i>relationships</i> across metrics have been defined. Metrics can be linked to employee work activities	Audit of metrics to vision, mission, KSFs Metrics Balance Check on System Components Metrics Balance Check on BSC/EFQM Metrics Audit Metrics Deployment Chart
Refinement	Metrics are refined based on <i>study</i> activities, if environment <i>changes</i> , and/or if they are no longer <i>needed</i>	Action reports and follow-up mechanisms

Source: From Van Aken et al. (2005), p. 406.

A case study of the use of the ISAT within the Belgian Railways demonstrated the usefulness of the framework in focusing organizational improvement efforts and providing a baseline for future assessments. Specifically, the assessment revealed the existing PM system was relatively strong in some, but not all, design elements, but weaker in implementation elements. In addition, the organization was noticeably stronger in approach than the other assessment scoring dimensions, particularly refine. Similarly, the

TABLE 30.5 Assessment Template for the Measurement System Implementation Element within the ISAT

Scoring Dimension	Assessment Dimensions/Items	Examples of Supporting Tools
Approach	<p><i>Structured approach</i> for fully implementing metrics (specify metrics, design portrayals, identify data collection process, plan for implementation and transition)</p> <p><i>Cross-functional involvement</i> in development and implementation of metrics (e.g., IT, Accounting)</p> <p>Use of group process <i>tools</i> to support development and implementation</p>	Metrics Development Matrix IT-driven metrics management
Deployment	<p>Development and implementation of metrics is <i>deployed</i> to lower levels if applicable</p> <p>Definitions and portrayals are clearly and consistently <i>communicated</i></p> <p>Internal to the organization, promoting accessibility and real-time use</p> <p>External to the organization</p> <p>Implementation activities <i>cover</i> all critical: Functions, processes, and work units in the organization</p>	<p>Visibility boards, email/intranet, newsletters, all-employee meetings, and so on</p> <p>Stakeholder meetings</p> <p>Audit of metrics deployment</p>
Study	<p><i>Metrics</i></p> <p>Have clear operational definitions with formulae if applicable</p> <p>Are consistently defined</p> <p>Have a clear purpose and need</p> <p>Enable users to make decisions</p> <p><i>Portrayal</i> mechanisms, formats, icons, etc.</p> <p>Are clearly and appropriately defined for each metric</p> <p>Are consistently used across metrics</p> <p><i>Data collection</i></p> <p>Data collection/tracking processes, tools, and roles are clearly defined and documented</p> <p>Data collection and portrayal frequencies are clearly defined for each metric</p> <p>Data collection processes are efficient (automated where possible)</p> <p>Resources are defined and available for data collection, tracking, and portrayal</p> <p><i>Target and baseline</i></p> <p>Are clearly defined for all metrics</p>	<p>Metrics deployment chart</p> <p>Metrics development matrix audit</p> <p>Portrayal design and Statistical thinking checklist</p>
Refinement	<p>Metrics are refined based on <i>study</i> activities, if environment <i>changes</i>, and/or if are no longer <i>needed</i></p>	Action reports and follow-up mechanisms

Source: From Van Aken et al. (2005), p. 407.

TABLE 30.6 Results Scoring Template for ISAT

Key Performance Area		Metric									
Rating Scale		Type of Metric: End-Result or Driver									
		0%		25%		50%		75%		100%	
Trends		No result, poor result or negative trend		Some improvement and early positive trend and/or fair performance		Positive trend and/or good performance		Sustained positive trend and/or excellent performance		Sustained excellent trend and/or sustained excellent performance	
What is our level of performance and trend?											
Goals		No result, poor result or negative trend		Some improvement and early positive trend and/or fair performance		Positive trend and/or good performance		Sustained positive trend and/or excellent performance		Sustained excellent trend and/or sustained excellent performance	
How are we performing against goals?											
Comparison		No comparison and/or unfavorable comparison		Somewhat favorable comparison, mostly internal sources		Somewhat to moderately favorable comparison with external sources		Favorable comparison, mostly external sources		Best in class and/or industry leader	
How are we performing against comparisons?											
Causes		No investigation of causes		Causes proposed and assumed		Causes monitored and some evidence for relationships		Causes monitored and strong evidence for relationships		Causes monitored and controlled	
How are we managing performance proactively?											
Overall Rating		No result, poor result or negative trend		Some improvement and early positive trend and/or fair performance		Positive trend and/or good performance		Sustained positive trend and/or excellent performance		Sustained excellent trend and/or sustained excellent performance	
Strengths		No result, poor result or negative trend		Some improvement and early positive trend and/or fair performance		Positive trend and/or good performance		Sustained positive trend and/or excellent performance		Sustained excellent trend and/or sustained excellent performance	
- T:		No result, poor result or negative trend		Some improvement and early positive trend and/or fair performance		Positive trend and/or good performance		Sustained positive trend and/or excellent performance		Sustained excellent trend and/or sustained excellent performance	
- G:		No result, poor result or negative trend		Some improvement and early positive trend and/or fair performance		Positive trend and/or good performance		Sustained positive trend and/or excellent performance		Sustained excellent trend and/or sustained excellent performance	
- C:		No result, poor result or negative trend		Some improvement and early positive trend and/or fair performance		Positive trend and/or good performance		Sustained positive trend and/or excellent performance		Sustained excellent trend and/or sustained excellent performance	
- C:		No result, poor result or negative trend		Some improvement and early positive trend and/or fair performance		Positive trend and/or good performance		Sustained positive trend and/or excellent performance		Sustained excellent trend and/or sustained excellent performance	
To be verified		No result, poor result or negative trend		Some improvement and early positive trend and/or fair performance		Positive trend and/or good performance		Sustained positive trend and/or excellent performance		Sustained excellent trend and/or sustained excellent performance	

Source: From Van Aken et al. (2005), p.405.

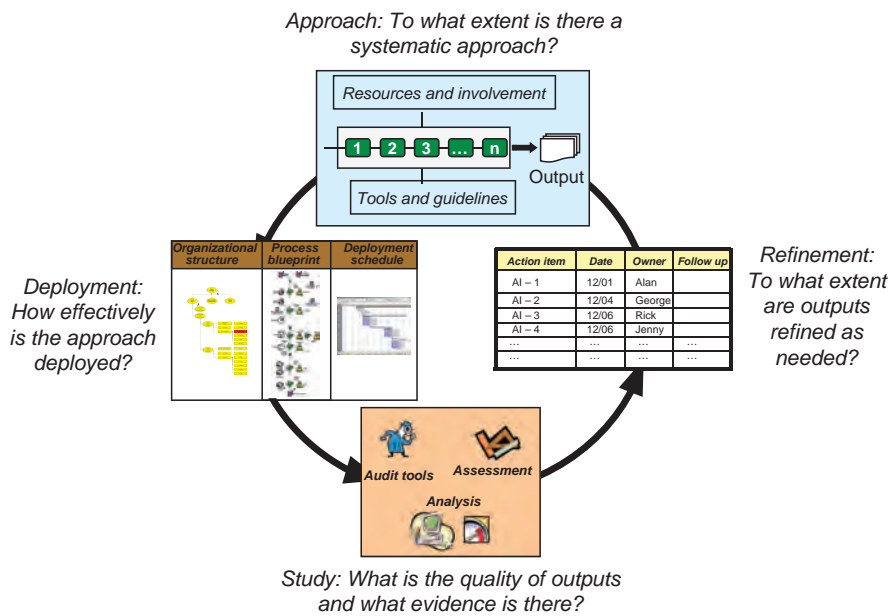


FIGURE 30.3 ADSR scoring dimensions used within the ISAT (Van Aken et al., 2005).

results assessment revealed the need to significantly improve performance on five of the six key measures assessed, and also to identify better benchmarks for comparisons and improve the monitoring of presumed driver.

30.2 SUMMARY

Developing an effective PM system is clearly a difficult prospect, with many different factors related to the design, implementation, and ongoing use of the system to which the organization must attend. The goals of this chapter were to provide an overview of the purpose and content of PM systems, as well as to provide high-level guidelines for their design, implementation, and use. The reader is encouraged to consult one of the full-length books on the subject—for example, Andersen and Faguerhaug (2002), Brown (2000), Brown (2007), de Waal (2002), and Neely et al. (2002)—for more detail and additional viewpoints on the design, implementation, and use processes.

REFERENCES

Andersen B, Faguerhaug T. *Performance Measurement Explained: Designing and Implementing Your State-of-the-Art System*. Milwaukee (WI): ASQ Quality Press; 2002.

Andersen H, Cobbold I, Lawrie G. Balanced scorecard implementation in SMEs—Reflection in literature and practice. SMESME Conference, Copenhagen, Denmark; 2001.

Azzone G, Masella C, Bertele U. Design of performance measures for time based companies. *International Journal of Operations & Production Management* 1991;11(3):77–85.

- Bauer J, Tanner SJ, Neely A. Developing a performance measurement audit template: a benchmarking study. *Measuring Business Excellence* 2004;8(4):17–25.
- Beischel ME, Smith KR. Linking the shop floor to the top floor: Here's a framework for measuring manufacturing performance. *Management Accounting* 1991;73:9–25.
- Bititci U, Carrie A, McDevitt L, Turner T. Integrated performance measurement systems: a reference model. In: *Organising the Extended Enterprise*, Schonsleben P, Buchel A. editors. London: Chapman & Hall; 1998. p. 191–203.
- Bourne M, Neely A, Platts K, Mills J. The success and failure of performance measurement initiatives: Perceptions of participating managers. *International Journal of Operations & Production Management* 2002;22(11):1288–1310.
- Bradley P. *A performance measurement approach to the reengineering of manufacturing enterprises [doctoral dissertation]*. CIMRU, NUI Galway, Ireland; 1996.
- Braam GJM, Nijssen EJ. Performance effects of using the Balanced Scorecard: A note on the Dutch experience. *Long Range Planning* 2004;37(4):335–349.
- Brown MG. *Winning Score*. New York: Productivity Press; 2000.
- Brown MG. *Beyond the Balanced Scorecard*. New York: Productivity Press; 2007.
- Chearskul P. *An empirical investigation of performance measurement system use and organizational performance [doctoral dissertation]*. Grado Department of Industrial and Systems Engineering, Virginia Tech, Blacksburg, VA; 2010.
- Chenhall RH. Integrative strategic performance measurement systems, strategic alignment of manufacturing, learning and strategic outcomes: An exploratory study. *Accounting, Organizations and Society* 2005;30(5):395–422.
- Crabtree AD, DeBusk GK. The effects of adopting the Balanced Scorecard on shareholder returns. *Advances in Accounting* 2008;24(1):8–15.
- Davis S, Albright T. An investigation of the effect of Balanced Scorecard implementation on financial performance. *Management Accounting Research* 2004;15(2):135–153.
- De Waal A. *Quest for Balance*. New York: John Wiley and Sons, Inc.; 2002.
- De Waal A, Counet H. Lessons learned from performance management systems implementations. *International Journal of Productivity and Performance Management* 2009;58(4):367–390.
- European Foundation for Quality Management (EFQM). *The EFQM excellence model*. 2011. <http://www.efqm.org/en/tabid/132/default.aspx>.
- Farris J, Van Aken EM, Letens G, Chearskul P, Coleman G. Improving the performance review process: A structured approach and case application. *International Journal of Production and Operations Management* 2011;31(4):376–404.
- Fitzgerald L, Johnston R, Brignall TJ, Silvestro R, Voss C. *Performance Measurement in Service Businesses*, London: CIMA; 1991.
- Franco-Santos M, Bourne M. An examination of the literature relating to issues affecting how companies manage through measures. *Production Planning & Control* 2005;16(2):114–124.
- Franco-Santos M, Kennerley M, Micheli P, Martinez V, Mason S, Marr B, Gray D, Neely A. Towards a definition of a business performance measurement system. *International Journal of Operations & Production Management* 2007;27(8):784–801.
- Gimbert X, Bisbe J, Mendoza X. The role of performance management systems in strategy formulation processes. *Long Range Planning* 2010;43(4):477–497.
- Halachmi A. Performance management: A look at some possible dysfunctions. *Work Study* 2002;51(4/5):230–239.
- Hoque Z, James W. Linking balanced scorecard measures to size and market factors: Impact on organizational performance. *Journal of Management Accounting Research* 2000;12:1–17.

- Iselin ER, Mia L, Sands J. The effects of the balanced scorecard on performance: The impact of the alignment of the strategic goals and performance reporting. *Journal of General Management* 2008;33(4):71–85.
- Ittner CD, Larcker DF. Coming up short on nonfinancial performance measurement. *Harvard Business Review* 2003;81(11):88–95.
- Ittner CD, Larcker DF, Randall T. Performance implications of strategic performance management in financial services firms. *Accounting, Organizations and Society* 2003;28(7–8):715–741.
- Kaplan RS, Norton DP. The balanced scorecard—Measures that drive performance. *Harvard Business Review* 1992;70(1):71–79.
- Kaplan RS, Norton DP. Linking the balanced scorecard to strategy. *California Management Review* 1996;39(1):53–79.
- Keegan D, Eiler R, Jones C. Are your performance measures obsolete? *Management Accounting* 1989;70(12):45–50.
- Kennerley M, Neely A. Performance measurement frameworks: a review. In: *Business Performance Measurement: Theory and Practice*. Neely A, editor. Cambridge, UK: Cambridge University Press; 2002. pp 145–155.
- Kotter John P. *Leading Change*. Cambridge (MA): Harvard Business School Press; 1996.
- Kurstedt HA. *Management system model helps your tool work for you*. White paper, Management Systems Laboratory, Department of Industrial and Systems Engineering, Virginia Tech, Blacksburg, VA; 1985.
- Lebas M. Managerial accounting in France: Overview of past tradition and current practice. *European Accounting Review* 1994;3(3):471–487.
- Leinonen M. A survey on performance measurement system design and implementation. Proceedings of the International Business and Economic Research Conference Reno, NV, Oct 8–12; 2001.
- Letens G, Verweire K, Slagmulder R, Van Aken EM, Farris J, Chearskul P. 2010. Implementing customer intimacy through integrated performance management and organizational learning. Proceedings of the 2010 American Society for Engineering Management Conference. Rogers, AR, Oct; 13–16; 2010, CD-ROM.
- Lingle JH, Schiemann WA. From balanced scorecard to strategic gauges: Is measurement worth it? *Management Review* 1996;85(3):56–61.
- Lipe MG, Salterio SE. The balanced scorecard: Judgmental effects of common and unique performance measures. *The Accounting Review* 2000;75(3):283–298.
- Lynch RL, Cross KF. *Measure Up-The Essential Guide to Measuring Business Performance*. Mandarin, London; 1991.
- Malina MA, Selto FH. Communicating and controlling strategy: an empirical study of the effectiveness of the balanced scorecard. *Journal of Management Accounting Research* 2001;13:47–90.
- Maltz AC, Shenhar AJ, Reilly RR. Beyond BSC: Refining the search for organizational success measures. *Long Range Planning* 2003;36(2):187–204.
- Marr B. *Strategic Performance Management: Leveraging and Measuring Your Intangible Value Driver*. Oxford, UK: Butterworth-Heinemann; 2006.
- Marr B, Bourne M, Kennerley M, Franco M, Wilcox M, Adams C, Mason S. *Business Performance Management: Current State of the Art*. Cranfield, UK: Cranfield University Press; 2004.
- Martinez V, Kennerley M. Performance management systems: Mixed effects. EURAM Conference. Munich, Germany; 2005.
- Martins RA, de Abreu ALT. Enhancement of understanding of relationship between performance measures at the operations level. Performance Management and Management: Public and Private Conference. London, UK; 2006; p. 473–480.

- Meekings A. Unlocking the potential of performance measurement: a practical implementation guide. *Public Money & Management* 1995;15(4):5–12.
- Meekings A. Effective review meetings: the counter-intuitive key to successful performance management. *International Journal of Productivity and Performance Management* 2005;54(3): 212–220.
- Meekings A, Povey S, Neely A. Performance plumbing: installing performance management systems to deliver lasting value. *Measuring Business Excellence* 2009;13(3):13–19.
- National Institute of Standards and Technology (NIST). 2011. *2011–2012 Criteria for Performance Excellence*. http://www.nist.gov/baldrige/publications/upload/2011_2012_Business_Nonprofit_Criteria.pdf.
- Neely A. The performance management revolution: why now and what next? *International Journal of Operations & Production Management* 1999;19(2):205–228.
- Neely A, Gregory M, Platts K. Performance measurement system design: a literature review and research agenda. *International Journal of Operations & Production Management* 1995;15(4): 80–116.
- Neely A, Adams C, Kennerley M. *The Performance Prism: The Scorecard for Measuring and Managing Business Success*. London: Financial Times/Prentice Hall; 2002.
- Neely A, Bourne M, Mills J, Platts K, Richards H. *Getting the Measure of Your Business*. Cambridge, UK: Cambridge University Press; 2002.
- Neely A, Micheli P, Martinez V. Action on information: performance management for the public sector. *Executive Briefing Series*. London, UK: The Advanced Institute of Management Research; 2006.
- Porter M. How Compétitive Forces Shape Strategy. *Harvard Business Review*: 57(2):137–145.
- Propper C, Wilson D. The use and usefulness of performance measures in the public sector. *Oxford Review of Economic Policy* 2003;19(2):250–267.
- Rigby RS, Bilodeau DP. Bain's global 2007 management tools and trends survey. *Strategy & Leadership* 2007;35(5):9–16.
- Rouse P, Putterill M. An integral framework for performance measurement. *Management Decision* 2003;41(8):791–805.
- Rucci AJ, Kirn SP, Quinn RT. The employee-customer-profit chain at Sears. *Harvard Business Review* 1998;76(1):82–97.
- Santos SP, Belton V, Howick S. Adding value to performance measurement by using system dynamics and multicriteria analysis. *International Journal of Operations and Production Management* 2002;22(11):1246–1272.
- Slizyte A, Bakanauskiene I. Designing performance measurement system in organization. *Management of Organizations: Systematic Research* 2007;43:135–148.
- Simons R. *Performance Measurement & Control Systems for Implementing Strategy*, Upper Saddle River (NJ): Prentice-Hall, Inc.; 2000.
- Sink DS, Tuttle TC. *Planning and Measurement in Your Organization of the Future*. Norcross (GA): Industrial Engineering and Management Press Norcross; 1989.
- Scott TW, Tiessen P. Performance management and managerial teams. *Accounting, Organizations and Society* 1999;24(3):263–285.
- Sousa GWL, Carpinetti LCR, Groesbeck RL, Van Aken EM. Conceptual design of performance measurement and management systems using a structured engineering design approach. *International Journal of Productivity and Performance Management* 2005;54 (5/6):385–399.
- Tuomela TS. The interplay of different levers of control: A case study of introducing a new performance measurement system. *Management Accounting Research*: 16(3):293–320.

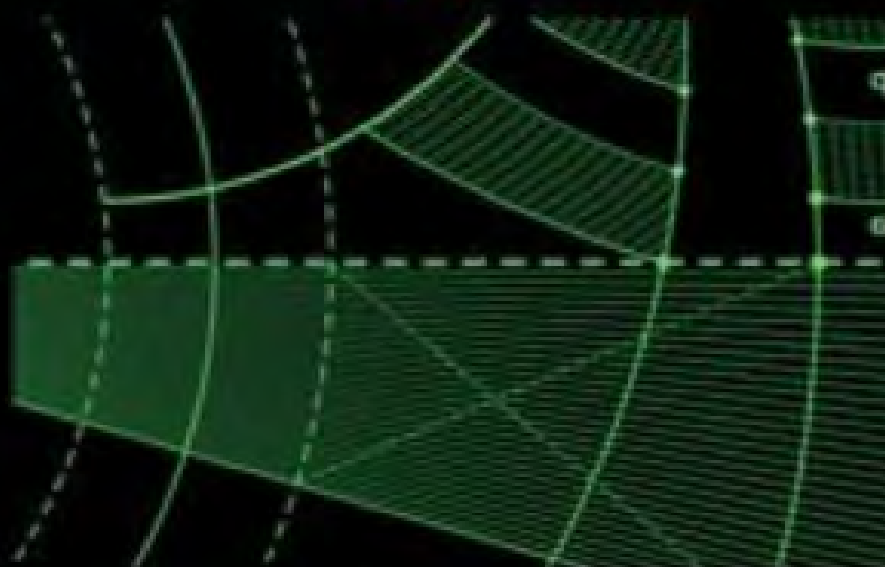
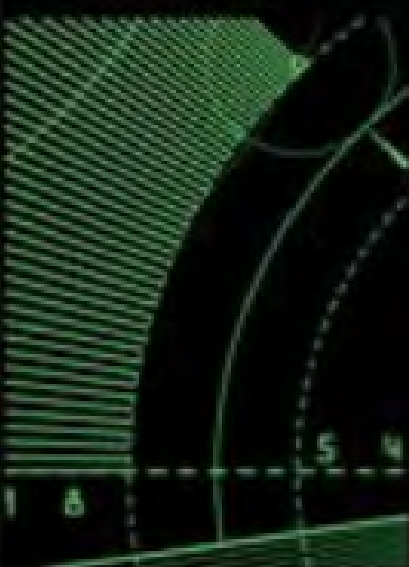
- Van Aken EM, Coleman GD. A measurement process for driving value-added. *Industrial Management* 2002; July/August, 28–33.
- Van Aken EM, Van Goubergen D, Letens G. Integrated enterprise transformation: case application in engineering project work in the Belgian armed forces. *Engineering Management Journal* 2003;15(2):3–16.
- Van Aken EM, Letens G, Coleman GD, Farris J, Van Goubergen D. Assessing maturity and effectiveness of enterprise performance measurement systems. *International Journal of Productivity and Performance Management* 2005;54(5/6):400–418.



HANDBOOK OF MEASUREMENT

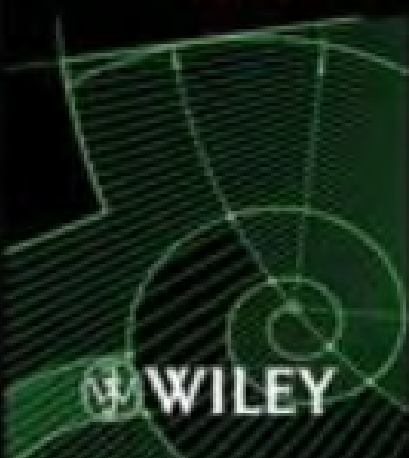
IN SCIENCE AND ENGINEERING

Volume 2



EDITED BY

MYER KUTZ



 WILEY

PART IV

MATERIALS PROPERTIES AND TESTING

PART VI

MEASUREMENT STANDARDS

31

VISCOSITY MEASUREMENT

ANN M. ANDERSON, BRADFORD A. BRUNO, AND LILLA SAFFORD SMITH

- 31.1 Viscosity background
 - 31.2 Common units of viscosity
 - 31.2.1 Absolute viscosity, μ
 - 31.2.2 Kinematic viscosity, ν
 - 31.2.3 Nonstandard units
 - 31.2.4 Distinction between rheology and viscometry
 - 31.2.5 Mathematical formalism
 - 31.2.6 Relation of viscosity to molecular theory
 - 31.2.7 Effect of pressure and temperature on viscosity
 - 31.2.8 Correlations of viscosity with temperature for gases
 - 31.2.9 Correlations of viscosity with temperature for liquids
 - 31.2.10 Effect of pressure on viscosity
 - 31.3 Major viscosity measurement methods
 - 31.3.1 Drag-type viscometers
 - 31.3.2 Bubble (tube) viscometers
 - 31.3.3 Rotational viscometers
 - 31.3.4 Flow-type viscometers
 - 31.3.5 Orifice-type (cup) viscometers
 - 31.3.6 Vibrational (resonant) viscometers
 - 31.4 ASTM standards for measuring viscosity
 - 31.5 Questions to ask when selecting a viscosity measurement technique
- References

31.1 VISCOSITY BACKGROUND

The most significant mechanical difference between materials classified as “fluids” and those classified as solids is in their reaction to shear stresses. (Recall that a shear stress is a distributed force, or force per unit area, whose direction of action is within the

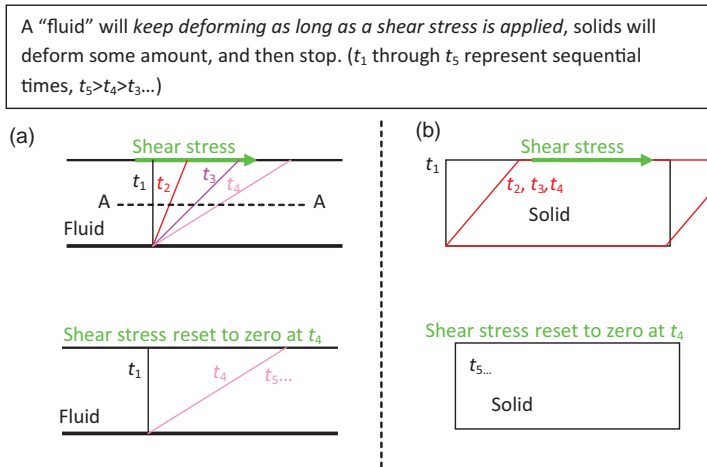


FIGURE 31.1 Fluid (viscous) versus Solid (elastic) behavior. (a) *Fluid Behavior*: A thin layer of fluid held between two parallel plates, the top plate caused to move relative to the bottom plate by application of a shear stress, will continue to deform (experience shear strain) as long as the shear stress is applied. The *rate* of shear strain is related to the magnitude of the shear stress. When the shear stress is removed the fluid will remain in its deformed state. (b) *Solid (Elastic) Behavior*: A solid will deform (experience shear strain) through some fixed angle when a shear stress is applied. The *amount* of shear strain is related to the magnitude of the shear stress. When the shear stress is released the solid will return to its original form.

plane of application. If you slide your open hand over a desk top the friction between your hand and the desk will create a shear stress on your hand.) As illustrated in Figure 31.1, a solid (within its elastic limit) will deform through some limited angle while a shear stress is applied, and will then return to its original configuration when the stress is removed. A fluid on the other hand will deform continuously as long as a shear stress is applied, and will not return to its original shape when the stress is removed. For a fluid the *rate* of shear deformation is related to the magnitude of the shear stress applied, while for a solid the *amount* of shear deformation is related to the magnitude of the shear stress applied. This seemingly innocuous mechanical difference in reaction to shear stresses gives rise to the large difference in character between solids and fluids. In fact it gives rise to all of the behaviors that one thinks of as inherently "fluid": the ability to flow, the ability to fill volumes of arbitrary shape, the ability to spread out and "wet" certain surfaces, etc. It also gives rise to the complex nature of fluid mechanics because it allows for the very large material deformations that in turn give rise to phenomena like turbulence.

Viscosity (or more precisely the *shear viscosity*, defined below) is the material property that defines the quantitative relation between the applied shear stress and the shear deformation rate in a fluid. Qualitatively the viscosity indicates the "thickness" or resistance to flow of a fluid. Since viscosity is the property that controls and quantifies the shear stress/shear rate behavior that is definitional to fluids, it is in many regards the most important physical property of a fluid.

Unfortunately, as alluded to above, the term “viscosity” is actually used to denote several related, *but different*, physical properties. It is important to understand these distinctions in terms from the outset. First, the term “viscosity” is most commonly used in conjunction with effects arising from *shear* forces and *shear* deformations in fluids. When used in this context, the most common one, the property is more precisely called the “shear viscosity” or the “first coefficient of viscosity.” However, when used in this sense, it is almost always simply referred to as “viscosity.” This is contrasted with the “bulk viscosity,” associated with volume dilatation. Bulk viscosity is rarely an important parameter and hence is not as well known or understood as the more common shear viscosity. Bulk viscosity is discussed briefly in Section 31.1.3. Second, it should be noted that even the “shear viscosity” described above is often stated in two different forms, the *absolute* or *dynamic* viscosity, μ , and the *kinematic* viscosity or *momentum diffusivity*, ν , where $\nu = \mu/\rho$ and ρ is the fluid’s density. Although the *dynamic* and *kinematic* viscosities are clearly related properties, they are dimensionally dissimilar and it is critically important to always distinguish between them. More is said on the distinction between dynamic and kinematic viscosity in the following section on common viscosity units.

The remainder of this chapter begins by discussing the units in which viscosity is measured. Then the distinction between the larger field of rheology and its subfield viscometry is made in the context of differentiating between the so-called Newtonian and non-Newtonian fluids. After that the chapter provides a brief theoretical and mathematical overview of viscosity. Finally, the majority of the chapter provides detailed and practical information on methods for measuring viscosity.

31.2 COMMON UNITS OF VISCOSITY

There are several systems of units used with viscosity; many of them are archaic and/or closely tied to one specific viscosity measuring technique (e.g., the Saybolt cup and the “Saybolt Universal Second,” and the Krebs unit) or one particular industry (e.g., SAE oil grade and the automotive industry). It is impossible to capture all of these systems in one document, but an attempt is made below to define and relate the most common and standard units associated with viscosity measurement.

31.2.1 Absolute Viscosity, μ

In terms of the SI (Le Systeme Internationale d’Unites) system of fundamental units the derived units for absolute viscosity, μ , are $\text{kg}/\text{m} \times \text{s}$ which is equivalent to $\text{Pa} \cdot \text{s}$ (Pascal-seconds). This grouping of units has not received a name of its own. In the closely related cgs (centimeter, gram, and second) system of units, the derived unit of $\text{g}/\text{cm} \times \text{s}$ or $\text{dyne} \times \text{s}/\text{cm}^2$ is called a “Poise” (after Poiseuille). More commonly a centipoise, $\text{cP} = 1/100\text{th}$ of a Poise is used. In the FPS (foot, pound, and second) system of units, the units of absolute viscosity are $\text{lb}_F \cdot \text{s}/\text{in}^2$, which is called the Reyn (after Osborne Reynolds). Refer to Table 31.1 for a collection of units of absolute viscosity.

TABLE 31.1 Units for Absolute/Dynamic Viscosity, μ

Units System	Derived Viscosity Unit		Unit Name	Equivalence
SI	$\frac{\text{kg}}{\text{m} \times \text{s}}$ $\text{Pa} \times \text{s}$	or	none	$1 \text{ Pa} \cdot \text{s} = 10 \text{ Poise} = 1000 \text{ Centipoise}$
cgs	$\frac{\text{g}}{\text{cm} \times \text{s}}$ $\frac{\text{dyne} \times \text{s}}{\text{cm}^2}$	or	Poise	$100 \text{ Centipoise} = 1 \text{ Poise}$
English (FPS)	$\frac{\text{lb}_F \times \text{s}}{\text{in}^2}$		Reyn	$1 \text{ Reyn} = 68,948 \text{ Poise}$

31.2.2 Kinematic Viscosity, ν

Recall that the dynamic (kinematic) viscosity, ν , is defined as the absolute viscosity divided by the fluid density, ρ . In the SI system of fundamental units the units for kinematic viscosity are meter square per second, which is not a named grouping. It should be noted that the units of kinematic viscosity (m^2/s) are identical to the units of thermal diffusivity used in heat transfer, and mass (species) diffusivity used in diffusion. This leads to the kinematic viscosity being referred to as the coefficient of momentum diffusivity by analogy. In the cgs system the unit of kinematic viscosity is the centimeter square per second called the “Stokes” (after G.G. Stokes). More commonly the kinematic viscosity is given in centistokes (cSt) where $100 \text{ cSt} = 1 \text{ Stokes}$. In the FPS system kinematic viscosity would be foot square per second or inch square per second, neither of which is a named unit.

31.2.3 Nonstandard Units

Kinematic viscosity is also often given in “Saybolt Universal Seconds” or SUS (also sometimes SSU “Saybolt Seconds Universal” or SUV “Saybolt Universal Viscosity”), which is directly related to the Saybolt viscosity cup measuring system (see Section 31.2.2). Of course, the unit of “seconds” is not a dimensionally correct unit for the physical quantity of kinematic viscosity, so this system is problematic. The Saybolt measurement system is based on ASTM method D88 and measurements in SUS can be converted into more standard (dimensionally correct) viscosity units using procedures provided in ASTM 2161. There are countless other such “legacy” scales of viscosity associated with different industries, and unfortunately there is often no standard method for converting these legacy measures into dimensionally correct viscosity units. A number of online viscosity converters exist (see www.coleparmer.com, www.gardco.com, or www.cannon.com, for example) (Table 31.2).

31.2.4 Distinction Between Rheology and Viscometry

A simple linear relationship between shear stress and shear strain rate is observed in a wide variety of fluids (Figure 31.2a). The constant slope of the line labeled Newtonian is

TABLE 31.2 Units for Kinematic Viscosity, ν

Units System	Derived Viscosity Unit	Unit Name	Equivalence
SI	m^2/s	None	$1 \text{ m}^2/\text{s} = 10,000 \text{ Stokes}$
cgs	cm^2/s	Stokes	$100 \text{ Centistokes} = 1 \text{ Stokes}$
English (FPS)	in^2/s , ft^2/s	None	$1 \text{ in}^2/\text{s} = 645.16 \text{ Centistokes} = 0.00694 \text{ ft}^2/\text{s}$

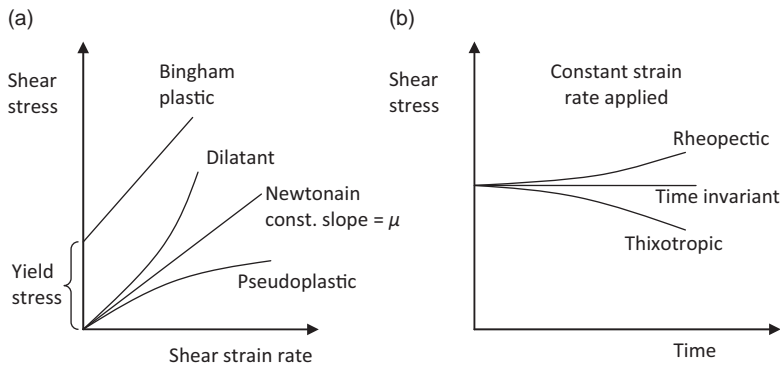


FIGURE 31.2 Newtonian and non-Newtonian fluid behavior (a) shear stress vs. strain rate (b) shear stress vs. duration of applied strain rate.

the (shear) viscosity of the fluid. Fluids demonstrating such a relationship are known as *Newtonian fluids*. Many common fluids like air, gases in general, water, or simple oils demonstrate Newtonian behavior meaning constant viscosity with respect to strain rate over a very wide range (many orders of magnitude) of strain rates. The measurement of the shear viscosity of Newtonian fluids is referred to as *viscometry* and is the focus of this chapter.

Fluids with more complicated molecular structures (e.g., polymers) or fluids with other phases suspended in them (e.g., mixtures, slurries, and colloids) often demonstrate more complicated shear stress to strain rate behaviors (see Figure 31.2a). Fluids exhibiting such behaviors are broadly characterized as “non-Newtonian” fluids. Non-Newtonian fluids can be further classified according to how they react to changes in shear deformation rates, to the duration of application of the applied loading, and to whether or not they exhibit a threshold elastic (solid-like) shear resistance prior to deforming like a fluid.

Fluids that show increasing *apparent viscosity* (the *apparent viscosity* is the local slope of the stress versus strain rate curve, see Figure 31.2a) as the applied strain rate increases are called “shear thickening” or *dilatant* fluids. The classic example of a shear thickening fluid is a mixture of cornstarch in water. If one attempts to shear this fluid quickly (e.g., hit it with a hammer) the viscosity will rise to such a level that the fluid seems almost solid, the hammer blow will bounce off the surface. Yet at lower shear rates the mixture will act like a “normal” fluid (e.g., a hammer set on its surface would sink right into the fluid). Fluids which show the opposite behavior (decrease in apparent shear viscosity with increasing strain rate) are called “shear thinning” or *pseudo-plastic* fluids. A common example of a shear thinning fluid would be “no drip” paint, which behaves as a fairly thick (viscous) fluid while adhering to a paintbrush (a low shear rate circumstance), but which spreads easily (i.e., exhibits lower viscosity) when the paintbrush is dragged along a surface thereby increasing the shear strain rate applied to the fluid (Cengel and Cimbala, 2010).

Some fluids will “thin” (produce a lower shear stress resisting the motion) or some will “thicken” (produce a higher shear stress resisting the motion) as the *duration* for which a constant strain rate is applied increases, (see Figure 31.2b). Fluids exhibiting the former behavior are referred to as “thixotropic” the latter as “rheopectic.” Such fluids are also sometimes referred to as “time-thinning” or “time-thickening” fluids. Examples of thixotropic fluids include yogurt and some classes of paint. Rheopectic behavior is much rarer.

Examples include gypsum paste and printers ink. Newtonian fluids exhibit constant strain rate with regard to loading duration for a constant applied shear stress.

Newtonian fluids will exhibit constant strain rate to shear stress behavior down to very low (theoretically zero) applied shear stresses. However some fluids, called “Bingham plastic” fluids will initially show “solid-like” behavior until a threshold shear stress (called the “yield stress”) is applied; after which they will show “fluid-like” behavior (continuously deforming while the shear stress is applied) (see Figure 31.2a). A common example of this type of fluid is toothpaste, which will not flow at all until a threshold shear value is exceeded. Broadly this kind of behavior is described as viscoelasticity. Bingham plastic materials can show dilatant, Newtonian, or pseudo-plastic behavior after their yield point. It is also worth noting that these terms are often not consistently applied.

The study and measurement of these more complicated, non-Newtonian, shear stress/-shear strain rate behaviors is a subset of the larger science referred to as rheology and is largely beyond the scope of this chapter.

31.2.5 Mathematical Formalism

In this section we develop the mathematical formulations governing viscosity, and explain the roles and relations between “shear” viscosity and “bulk” viscosity. The discussion begins with the consideration of a very simple one-dimensional (1D) flow situation, and then introduces the more general 3D form of the equations.

Shear viscosity is defined mathematically by Newton’s Law of Viscosity. Newton’s law defines viscosity as the physical property that relates the shear stress produced as a reaction to an applied strain deformation rate. If one considers the simple flow shown in Figure 31.1a, a thin layer of fluid confined between two infinite parallel flat plates the upper moving within its own plane relative to the lower, the total *shear strain rate* on any layer in the fluid is given by $\partial u/\partial y$ which is the rate of change of x direction velocity in the perpendicular (y) direction. Newton’s law states that the shear stress experienced on the lower face of such a layer is given by:

$$\tau = -\mu \frac{du}{dy} \quad (31.1)$$

where μ is the (shear) viscosity of the fluid at the applied strain rate. If the shear viscosity is constant with regard to strain rate then the fluid is said to be “Newtonian.” If the fluid exhibits a more complex relationship between shear stress and strain rate the fluid is defined as “non-Newtonian.” Distinctions between Newtonian and non-Newtonian behavior is discussed in greater detail in Section 31.1.2.

The viscosity picture becomes more complicated if we allow for more complex (3D) motions of the fluid. Any motion that a “particle” of fluid can undertake can be constructed from a superposition of the following four simpler types of motion: pure translation (movement without rotation or deformation), pure rotation (rotation without movement or deformation), pure linear strain (deformation without motion that does not disturb angles within the fluid particle), and pure shear strain (deformation without motion which does change angles within the fluid particle). Figure 31.3 illustrates these types of motions in two dimensions. In a fully general flow any combination of these motions can occur in any or all of the three coordinate directions. In such cases independent shear deformations (Figure 31.3d) can occur on any or all of three orthogonal planes.

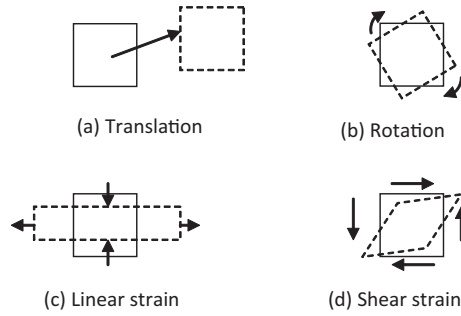


FIGURE 31.3 Types of fluid motion and deformation illustrated in 2-D.

The fluid can also undergo purely extensional deformations (i.e., elongations without shear, Figure 31.3c) in any or all of the three dimensions. These extensional deformations will in some circumstances also contribute to the stress response of the fluid, for example, if they combine in such a way that the volume of the fluid element changes then the “bulk viscosity” described below will also be important. Thus for general 3D deformations it is necessary to use tensors (a branch of mathematics that describes vectors pointing in several directions) to describe the full relation between the stress at a point in the fluid and the resulting strain.

The stress response of a Newtonian fluid element in response to a fully general deformation is given in White (2006) as

$$\tau_{ij} = -P\delta_{ij} + \mu \left(\frac{\partial u_i}{\partial x_j} + \frac{\partial u_j}{\partial x_i} \right) + \delta_{ij}\gamma(\nabla \cdot \vec{V}) \quad (31.2)$$

Here μ is the shear viscosity, γ is the “Bulk Viscosity” coefficient and δ_{ij} is the Kronecker delta function (i.e., $\delta = 1$ when $i = j$, and $\delta = 0$ when $i \neq j$), and i and j are indices used to refer to the three orthogonal planes. Those interested in the derivation of this relation are directed to White or any other graduate level fluids text.

One important point to note arising from Equation (2) is that the stress response to purely extensional strains is described by the bulk viscosity, γ , or through the closely related “second coefficient of viscosity” μ' (the bulk viscosity, γ is equal to the second coefficient of viscosity, μ' , plus $2/3 \mu$, see Owczarek (1964), for example, for a more thorough discussion). The topic of bulk viscosity is largely beyond the scope of this chapter, but these few comments are made to inform the reader of when it may be safely ignored, and when it may become important. First it is important to note that the bulk viscosity is not as well understood or characterized as the more common shear viscosity. Fortunately, for Newtonian fluids, the bulk viscosity coefficient occurs only in combination with the divergence of the velocity field ($\nabla \cdot \vec{V}$); and for incompressible fluids conservation of mass (i.e., continuity) requires that $\nabla \cdot \vec{V} = 0$. Hence the bulk viscosity will play no role in a truly incompressible fluid.

Of course no substance is truly 100% incompressible, so if one is concerned with acoustic issues (which are inherently a compressibility phenomenon) or flows at significant Mach number the bulk viscosity may play a role. One way to gain insight on when the bulk viscosity may play a significant role is to examine its role in viscous dissipation.

Ultimately the role of viscosity (shear or bulk) is to irreversibly convert the mechanical energy in a flow into thermal energy (heat). This effect is known as “viscous dissipation.” Hence the magnitude of the dissipation caused by the viscosity can give one indication of the qualitative importance of viscosity in the flow. The dissipation (Φ) caused by the bulk viscosity is given by Shaughnessy et al. (2005) as:

$$\Phi_{bulk} = \frac{\gamma}{\rho^3} \left(\frac{d\rho}{dt} \right)^2 \quad (31.3)$$

where γ is the “bulk viscosity” and ρ is the density. In order to be significant, either large changes in density are required, or the changes in density must occur over very short time scales. Thus, due to the relative magnitudes of the extensional strains typically involved, the dissipation due to bulk viscosity (and indeed the bulk viscosity itself) may safely be ignored for almost all practical applications except shock waves (which involve large changes in density) and attenuation of high frequency (small dt) sound or “ultrasound”. Henceforth in this chapter the term “viscosity” shall refer to “shear viscosity” only.

31.2.6 Relation of Viscosity to Molecular Theory

As discussed above, ultimately the role of viscosity is to dissipate the ordered kinetic energy associated with the macroscopic motions of a flow to disordered, randomly distributed, microscopic molecular energy (i.e., thermal energy). As such it becomes clear that the quantity that we refer to as viscosity is the macroscopic manifestation of molecular level effects; much in the same way that the macroscopic quantity of “pressure” represents the average net force per unit area caused by countless individual molecular collisions on a surface. That is to say that viscosity, like pressure, temperature, and density, is a “continuum property” of a substance. It makes no more sense to talk about the “viscosity” of a single molecule than it does to talk about the “density” of a single atom of a gas. The fact that the viscosity is an emergent property arising only for large collections of molecules places an important limitation on the concept of fluid viscosity, namely the continuum limit. Essentially the (continuum) concept of viscosity breaks down when applied on length scales that are comparable to the mean free path of the molecules in the fluid, or when applied on time scales that are comparable to the mean time period between molecular collisions in the fluid. For ordinary macroscopically sized flows at ordinary pressures the continuum limit is not a concern. However in applications such as nanotechnology (where the length scales of concern become very small) or rarefied gas dynamics (where the mean free path of molecules in very low density gases becomes very large) this limit should be kept in mind.

Examining how the macroscopically observed property of viscosity arises from molecular effects can provide insight and physical intuition about viscosity. If we examine how momentum is transported by the thermal (random) motions of molecules within a flow we can come to understand the molecular basis for viscosity.

First, let us consider the physics qualitatively. To do so consider again the simple 1D shear flow shown in Figure 31.1. Specifically consider the molecules near a plane parallel to the top plate and half way between the top plate and the bottom plate. All of these molecules will have velocities that are composed of their individual, random (thermal)

velocities (which average out to zero bulk velocity), plus a small extra velocity component which depends on the local value of the bulk velocity. Those molecules slightly above the plane will, on average, have slightly larger values of velocity in the x direction and therefore also have slightly higher values of x direction momentum. Similarly, on average, those below the plane will have slightly lower values of x velocity and momentum. All of the molecules (above and below the plane) will have random thermal velocities with components in all three directions. The y component of these random velocities will occasionally carry molecules across the plane in both directions. But because of the asymmetry in velocity above and below the plane (i.e., because of the gradient of velocity perpendicular to the plane) the net effect of these random cross plane exchanges will be to transport x momentum from above the plane to below the plane. That is, there will be a net flux of x momentum in the negative y direction.

In fact the quantity we call “viscosity” is precisely defined by this “diffusive” transport (i.e., transport caused by random molecular motions rather than bulk macroscopic fluid motions) of momentum in the direction opposite to a velocity gradient. A simple dimensional analysis will convince the reader that a stress, such as shear stress (measured in $\text{Pa} = \text{kg/m s}^2$), is dimensionally identical to a flux of momentum (transport of momentum per second per unit area $= \text{kg/m s}^2$). Thus, referring back to Equation (1), we can consider viscosity to be the fluid property that relates the diffusive momentum flux (τ) to the velocity gradient driving it.

31.2.7 Effect of Pressure and Temperature on Viscosity

31.2.7.1 Low Density Gases We can make this relationship more quantitative by examining the actual molecular interactions occurring in the fluid. The following development parallels that given by Bird et al. (1960), but similar developments can be found in any text covering the kinetic theory of gases. The simplest model of viscosity arises from a consideration of simple kinetic theory of a low density gas where we assume that the gas molecules are rigid spheres with diameter “ d ” that only interact through collisions (i.e., there are no forces causing “action at a distance” between molecules). Thus molecules will exchange momentum and come into equilibrium with their surroundings only through collisions. Kinetic theory for rigid sphere molecules also provides these other important results that will be used in this development:

- The average random (thermal) velocity of the molecules in the gas will be

$$\bar{V} = \sqrt{\frac{8kT}{\pi m}} \quad (31.4)$$

where k is Boltzman’s Constant, m is the mass of the individual molecules and T is the absolute temperature.

- The mean free path of a molecule between collisions, λ , will be

$$\lambda = \frac{1}{\sqrt{2}\pi d^2 n} \quad (31.5)$$

where n is the number density (number per unit volume) of the gas molecules.

- The average frequency of collision per unit area (from one side) on any plane in the flow

$$Z = \frac{1}{4} n \bar{V} \quad (31.6)$$

If one again considers the situation of a plane between the two plates in Figure 31.1 it can be shown that a molecule will travel an average vertical distance of $2/3 \lambda$ between collisions. The flux of x momentum transferred across the plane from below is then $Zm v_{x,y-2/3\lambda}$ where $v_{x,y-2/3\lambda}$ is the average excess (nonthermal) x velocity component at a y location “one collision distance” below the plane (i.e., at $y-2/3\lambda$). Similarly the flux of x momentum transferred across the plane from above is, $Zm v_{x,y+2/3\lambda}$. Thus the net x momentum flux is:

$$\tau = Zm v_{x,y-2/3\lambda} - Zm v_{x,y+2/3\lambda} \quad (31.7)$$

And, if the x velocity gradient in the vicinity of y is linear we can replace the v_x at locations above and below the plane with a first order Taylor series in terms of the gradient at the plane, yielding:

$$\tau = Zm \left\{ \left(v_{x,y} - \frac{2}{3} \lambda \frac{dv_x}{dy} \right) - \left(v_{x,y} + \frac{2}{3} \lambda \frac{dv_x}{dy} \right) \right\} \quad (31.8)$$

Simplifying and substituting in the definition of Z gives:

$$\tau = \frac{-1}{3} nm \bar{V} \lambda \frac{dv_x}{dy} \quad (31.9)$$

Comparing back to Equation (31.1) we can see that:

$$\mu = \frac{1}{3} nm \bar{V} \lambda \quad (31.10)$$

Further substitution for \bar{V} and λ yields:

$$\mu = \frac{2}{3} \frac{1}{\pi^{3/2}} \frac{\sqrt{mkT}}{d^2} \quad (31.11)$$

Thus if the molecules are considered as perfectly rigid spheres (the simplest model) then $\mu \sim \sqrt{T}$. Note also that the viscosity of such a gas is not expected to be a function of pressure based on this very simple molecular model. This is an important result which is largely borne out by experimental observations of real low-density gases; their dynamic viscosity is observed to depend only very weakly on pressure.

Of course molecules are not perfectly rigid spheres for that would imply no force whatsoever between molecules until they come into contact (when their center to center distance equals d) and then an infinite repulsion force. Clearly the idea of an infinite repulsion force is unphysical. In fact all molecules will show some “action at a distance” and, more importantly, act as if they have some “give” or flexibility when they “collide”

which will eliminate the unphysical infinite forces inherent in the simple rigid sphere model. Typically these interactions are modeled as an intermolecular “potential” function from which the magnitudes of attractive and/or repulsive forces as a function of center to center distance can be calculated; as can an “effective” molecular diameter. The exact nature of these intermolecular potential functions is complicated, and has received extensive study, but is largely beyond the scope of this chapter. The interested reader is directed to Kogan (1969) or Vicenti and Kruger (1965), for example. Here it is sufficient to make a few points. First, as more complex and precise relations for the potential theory are used, the predictions of macroscopic properties such as viscosity improve markedly. Second, the “action at a distance” effects associated with these potentials allow for pressure to have an effect on the viscosity of gases. But, as stated above, this is typically found to be a weak effect for common gases. And finally, the functional relation between the viscosity and the temperature in low-density gases depends critically on the details of this potential function.

The next simplest model of intermolecular relations (after the rigid sphere model) is one proposed by Maxwell. In such a model the potential function drops off as $1/s^4$ where s is the center to center distance between the molecules. With such a model it can be shown that the “effective molecular diameter” will vary as: $d^2 \sim 1/\sqrt{T}$ and so, referring back to Equation 11, $\mu \sim T$. That is, a gas composed of “Maxwellian” molecules will have a viscosity that will increase linearly with temperature, as opposed to with the square root of temperature predicted from the rigid sphere model.

31.2.8 Correlations of Viscosity with Temperature for Gases

In reality, gases typically exhibit behaviors between the two extremes discussed in the section above (rigid and “Maxwellian”) and so their viscosity’s dependence on temperature is commonly correlated with a power-type law, as given by White (2006):

$$\frac{\mu}{\mu_o} = \left(\frac{T}{T_o} \right)^n \quad (31.12)$$

Where the parameters n , T_o , and μ_o are specific to the particular gas. Values for “ n ” for most simple gas molecules fall between 0.5 and 1, as predicted by the above discussion. Some values for n , T_o , and μ_o can be found in White.

Another common correlation technique for gas viscosities is based on the work of Sutherland (1893) (as covered in Vicenti and Kruger, 1965). Here viscosity is correlated as

$$\frac{\mu}{\mu_o} = \left(\frac{T}{T_o} \right)^{\frac{3}{2}} \left(\frac{T_o + S}{T + S} \right) \quad (31.13)$$

where S is the so-called Sutherland Parameter and T_o , and μ_o are reference values. A completely equivalent form of this equation:

$$\mu = \left(\frac{C_1 T^{\frac{3}{2}}}{T + S} \right) \quad (31.14)$$

where

$$C_1 = \left(\frac{\mu_o}{T_o^{\frac{3}{2}}} \right) (T_o + S) \quad (31.15)$$

is often used as well. Poling et al. (2004) provided a detailed discussions of several much more detailed methods for estimating the variation of viscosity of both pure gases and mixtures of gases at various temperatures.

31.2.9 Correlations of Viscosity with Temperature for Liquids

In real (higher density, more complex molecular structure) gases and especially in liquids intermolecular forces (beyond the “collisional” forces discussed previously) play a critically important role. Molecules in such substances can exert significant force (and hence transfer significant momentum) “at a distance” without colliding. Since viscosity as a property arises from transfers of momentum, these “actions at a distance” must be accounted for in any physical model that hopes to adequately predict a material’s viscosity. The nature and magnitude of these non-collisional interactions are so complex and so large in liquids that currently no one general model exists that will adequately predict the viscosity of all liquids. Instead many specialized empirical and semi-empirical relations are available. Some general trends however are observed for liquids, the most important being that *the viscosity of liquids falls off strongly with increasing temperatures*. One type of curve fit that is recommended for liquid viscosity, recommended by White (2006) is

$$\ln \frac{\mu}{\mu_o} \cong a + b \left(\frac{T_o}{T} \right) + c \left(\frac{T_o}{T} \right)^2 \quad (31.16)$$

where a , b , and c are curve fit parameters and T_o and μ_o are reference values. Another, slightly simpler, empirical correlation often used is “Andrade’s equation” (Munson et al., 2009)

$$\mu = D e^{\frac{B}{T}} \quad (31.17)$$

which is often presented in the alternate form

$$\ln \mu = A + \frac{B}{T} \quad (31.18)$$

Viswanath et al. (2007) provide a lengthy discussion of such correlation methods and coefficients A and B for a wide variety of liquids.

31.2.10 Effect of Pressure on Viscosity

The effects of pressure on viscosity are not nearly as significant as the effects of temperature. In many practical circumstances it is entirely sufficient to simply neglect the effect

of pressure on viscosity. This has lead to pressure effects being much less studied, and to data on viscosity at different pressures being much more sparse.

For low density gases, the molecular dynamics models discussed above (refer to Equation 31.10) indicate that the absolute viscosity, μ , should not depend on pressure at all; due to the competing effects of increasing number density (n) and decreasing mean free path (λ) as pressure is increased. Of course the value of the kinematic viscosity, $\nu = \mu/\rho$, of gases will decrease with increasing pressure due to the increase in density, ρ , of the gas as the pressure increases. These predictions are largely borne out by experimental data for common low-density (ideal) gases. Poling et al. (2004) provide a detailed discussion of several methods for estimating the viscosities of both pure gases and mixtures of gases at higher pressures.

Viscosity data for liquids at high pressure is sparse when compared to the quantity of data available at or near atmospheric pressures. In general, as pressure increases, so does the viscosity of most liquids. Both Viswanath et al. (2007) and Poling et al. (2004) recommend a method for estimating the effect of pressure on liquid viscosity attributed to Lucas (1981):

$$\frac{\mu}{\mu_{SL}} = \frac{1 + D \left(\frac{\Delta P_r}{2.118} \right)^A}{1 + C \omega \Delta P_r} \quad (31.19)$$

where μ is the viscosity of the liquid at pressure P ; μ_{SL} is the viscosity of the liquid at vapor pressure P_{vp} ; P_{vp} is the the vapor pressure; $\Delta P_r = (P - P_{vp})/P_c$; $T_r = T/T_c$; P_c, T_c are the critical pressure and temperature; ω is the acentric factor

$$A = 0.9991 - \left[\frac{4.674 \cdot 10^{-4}}{1.0523 T_r^{-0.03877} - 1.0513} \right] \quad D = \left[\frac{0.3257}{(1.0039 - T_r^{2.573})^{0.2906}} \right] - 0.2086$$

$$C = -0.07921 + 2.1616 T_r - 13.4040 T_r^2 + 44.1706 T_r^3 - 84.8291 T_r^4 + 96.1209 T_r^5 - 59.8127 T_r^6 + 15.6719 T_r^7$$

In spite of the theory and correlation techniques described above, in many practical situations one must resort to simply measuring viscosity. The following section describes the different measurement techniques available.

31.3 MAJOR VISCOSITY MEASUREMENT METHODS

Viscometers are designed to make use of the theoretical relationship between shear stress and strain rate to measure viscosity. They do this using simple flows (1D, steady, fully developed) in which both the shear stress and strain rate can be measured. There are three primary types of viscometers: flow, drag, and resonant. The flow-type viscometers measure the rate of flow of the fluid in a tube or through an orifice. The shear stress can be calculated from theory (e.g., capillary tube viscometer) or estimated based on theory (e.g., orifice cup viscometers). Use of these types of viscometers yields values for kinematic viscosity. Design parameters for flow-type viscometers include minimizing entrance and exit effects, maintaining a constant pressure head (which drives the flow), minimizing surface tension effects and mitigating effects of temperature variation.

Drag-type viscometers measure either the force on an object as it moves at a specified rate in the fluid (rotational viscometers) or measure the time it takes for an object to move a specified distance through the fluid (falling object and bubble tube viscometers). Use of these types of viscometers yields values for absolute viscosity (except for the bubble tube which measures kinematic viscosity). Design parameters for drag-type viscometers include minimizing the effects of turbulence and flow separation through the specification of a flow condition (generally a low relevant Reynolds number), controlling for transients, minimizing surface tension effects and mitigating effects of temperature variation. The third type of viscometer is the resonant or vibrational viscometer which is most commonly used in in-line process applications. These are designed so that changes in the viscous damping bring about significant changes in the resonance behavior of the instrument. Use of these viscometers yields values for kinematic viscosity.

This section presents information on each of the three major viscometer types. It begins with the drag-type viscometers (falling object, bubble tube, and rotational) followed by the flow-type viscometers (capillary and orifice) and concludes with a discussion of vibrational viscometers. Each section includes information on the theory of operation, a description of the types of viscometers available, a list of available manufacturers, and the capabilities and advantages/limitations. This chapter focuses on the use of laboratory-type viscometers; however some information is included on the use of process viscometers. This list of viscometers is not intended to be exhaustive but includes many of those that are most readily available commercially. There are other specialized methods for measuring viscosity and the reader is referred to Viswanath et al. (2007) for more information.

31.3.1 Drag-Type Viscometers

31.3.1.1 Falling Object Viscometers

Theory of Operation Falling object viscometers determine viscosity (μ) by measuring the drag force acting on a falling object under specific flow conditions. Use of the falling object viscometer requires a separate measurement of density to calculate kinematic viscosity. Figure 31.4 illustrates the forces acting on a falling object. This case shows a spherical object (a ball), however there are a variety of falling objects that can be used such as needles, and cylinders. There are three forces acting on the object: F_B the buoyancy force and F_D the drag force act upwards while F_G the gravitation force (weight) acts down. The buoyancy force is calculated using Archimedes principle and is equal to the weight of the fluid displaced by the object:

$$F_B = \nabla_B^* \rho_f^* g \quad (31.20)$$

where ∇_B is the volume of the object, ρ_f is the density of the fluid, and g is the gravitational constant. The weight of the object is simply:

$$F_G = \nabla_B^* \rho_B^* g \quad (31.21)$$

where ρ_B is the density of the object. If the object is travelling at terminal speed, the acceleration will be zero and application of Newton's second law yields an equation relating the three forces:

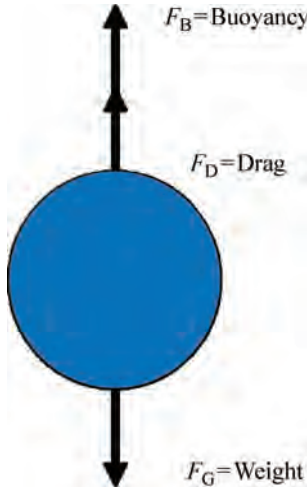


FIGURE 31.4 Schematic showing the forces acting on a falling object.

$$F_D = F_G - F_B = V_B^*(\rho_B - \rho_f)^*g \quad (31.22)$$

The drag force is composed of a shearing force (due to the fluid) and a pressure force (due to flow separation). Falling object viscometers are generally designed to operate in the Stokes (creeping) flow regime which is characterized by a lack of flow separation and occurs for very low Reynolds numbers ($Re < 0.1$). In this case the drag force is due only to the shearing force. For example, if the object is a sphere then the Reynolds number is calculated as

$$Re = \frac{\rho_f VD}{\mu} \quad (31.23)$$

Where V is the terminal speed of the object, D is the diameter of the sphere, and μ is the viscosity. For the low Reynolds number situation, the drag force is related to Reynolds number by Stokes Law:

$$F_D = \frac{3\pi\mu VD}{Re} \quad (31.24)$$

Combination of these equations yields an equation for viscosity in terms of the speed, diameter, and density difference:

$$\mu = \frac{gD^2}{18V}(\rho_{obj} - \rho_{fluid}) \quad (31.25)$$

This theory applies to balls moving at low Reynolds number in an infinite media (see Brizard et al. 2005 for development of the theory accounting for more realistic conditions). In many commercial applications, the falling object is placed in a tube of specified diameter (see Figure 31.5). The object, typically a ball, will fall or slide down the tube and

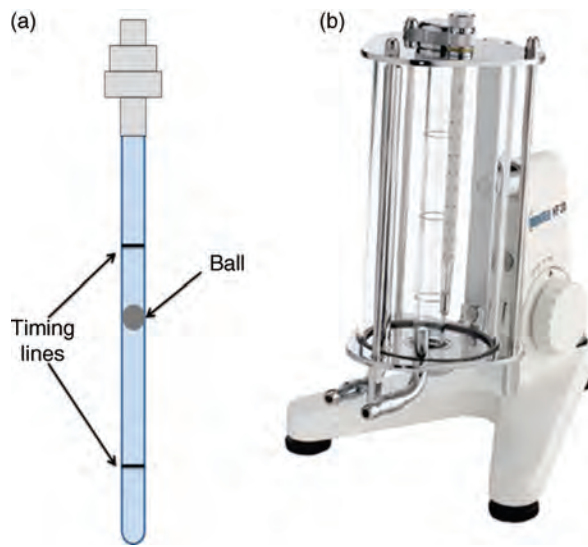


FIGURE 31.5 Falling ball viscometers (a) Gilmont and (b) Hakke type (Courtesy of Brookfield Engineering Labs).

the user measures the time it takes for the ball to travel between two timing lines. The first timing line is placed sufficiently far from the top of the viscometer to allow the ball to reach terminal velocity. The manufacturer supplies a calibration equation of the sort:

$$\mu = K^* t^* (\rho_{\text{obj}} - \rho_{\text{fluid}}) \quad (31.26)$$

where K is a calibration constant and t is the measured time to fall the specified distance. To obtain the calibration constant the manufacturer measures the fall time of the ball in a series of liquids of known viscosity.

Types of Viscometers/Options Falling object viscometers use a variety of objects including spheres, needles, and cylinders. Falling object viscometers often come with a set of objects, each with different mass/density which allows one to measure viscosity over a range of values. The most readily available commercial falling object viscometers are the relatively inexpensive Gilmont-type falling ball viscometers and the more expensive, more accurate Haake-type falling ball viscometers as shown in Figure 31.5a and b. The Haake viscometers include a mounting mechanism and an outer chamber that can be used for temperature control of the sample during testing. Falling needle viscometers use thin needle like objects which are designed to minimize wall effects and are more stable as they fall (see Davis and Brenner, 2001). They can be used to measure viscosity of non-Newtonian fluids. Falling cylinder viscometers involve a more complex flow field subject to significant end effects; however, they are useful for measuring viscosity at high pressure (Cristescu et al., 2002). Table 31.3 provides further information about suppliers of falling object viscometers.

Summary Although falling object viscometers are relatively inexpensive, the use of one requires some skill and is labor intensive. The tubes must be carefully cleaned before use and when filling the tube with the fluid of interest, care needs to be taken to avoid air

TABLE 31.3 Sampling of Common Falling Ball Viscometer Types

Type	Name	Manufacturer/Vendor	Range (cP)	Price
Falling ball	Gilmont Falling-ball Viscometers	Gilmont; Cole Parmer/ Thermo Scientific; Gardco	0.2–200	\$ (\$200)
Falling ball	HAAKE Falling Ball Viscometer	HAAKE/Thermo Scientific	0.5–10 (up to 7500)	\$\$\$ (\$3500)
Falling ball	Brookfield Falling Ball Viscometer	Brookfield/Gardco	0.5–70,000	\$\$\$ (\$3000)
Falling needle	PDV-100 Portable Field Viscometer	Stony Brook Scientific/Gardco; Cole Parmer	5–106	\$\$ (\$700)

bubbles. They cannot be used with opaque liquids. After setup, each individual measurement can take 1–2 min to complete. Gilmont-type viscometers must be mounted or held vertically and care needs to be taken when handling them to avoid heating up the fluid in the viscometer. Haake-type viscometers are pre-mounted at a specified angle and the falling ball tube is located inside an outer glass tube that can be easily connected to a circulating water bath for temperature control. They are not automated and require manual timing. However, it is relatively easy to compare the viscosity of different fluids by using multiple viscometers. The primary sources of error that arise in the use of a falling ball viscometer are related to temperature effects, handling and contamination.

31.3.2 Bubble (Tube) Viscometers

Theory of Operation Although most drag-type viscometers measure absolute viscosity and require a separate measurement of density to calculate kinematic viscosity, bubble viscometers measure kinematic viscosity (ν) by measuring the drag on a rising bubble of air which has low density compared to the fluid. The bubble tube consists of a glass tube (see Figure 31.6a) which is filled with the liquid of interest (leaving space for a bubble to

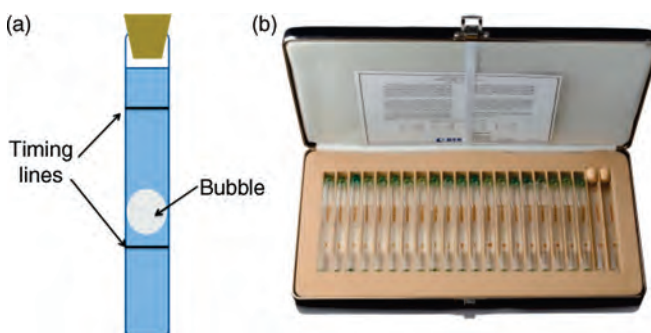


FIGURE 31.6 (a) Schematic of a bubble tube viscometer and (b) BYK-Gardner Bubble Viscometer Set (Courtesy of BYK Gardner USA).

form). The tube is marked with timing lines and the time for the bubble to rise is measured and compared to times for liquids with known viscosities.

The theory in this case is similar to that for a falling object, except that here one measures the time for the bubble to rise a specific distance (see Goldsmith et al., 1962, for more information on the theory). If we ignore the effect of bubble viscosity the same equation for viscosity applies:

$$\mu = \frac{gD^2}{18V}(\rho_{\text{obj}} - \rho_{\text{fluid}}) \quad (31.27)$$

However in the case of a rising gas bubble one can assume that the object (air) density is much less than that of the fluid and the calibration equation becomes:

$$\mu = \frac{gD^2}{18V}(\rho_{\text{fluid}}) \quad (31.28)$$

Or in terms of kinematic viscosity

$$\nu = \frac{\mu}{\rho_{\text{fluid}}} = \frac{gD^2}{18V} \quad (31.29)$$

In practical applications, bubble viscometers use carefully manufactured precision tubes which are calibrated using known viscosity standards. The kinematic viscosity is then calculated as:

$$\nu = k^* t \quad (31.30)$$

where k is the calibration constant for the tube and t is the time for the bubble to rise a specific distance. The faster the bubble travels, the lower the viscosity.

Types of Viscometers/Options Bubble tube viscometers are used to measure viscosity using either a direct measurement method or a comparison method. In the direct method, a single tube (like that shown in Figure 31.6a) is used to measure the bubble rise time and that time is converted to viscosity using information from the manufacturer. In the comparison method, the user selects reference tubes of known viscosity (see Figure 31.6b) and directly compares the bubble rise time in the fluid of interest to the rise time in the reference tubes. The fluid of interest is then assumed to have the same viscosity as that of the reference tube with the closest rise time. Table 31.4 includes a list of suppliers for bubble tube viscometers.

Summary Bubble tube viscometers are relatively inexpensive. Their use requires some skill and measuring viscosity is labor intensive. The tubes must be carefully cleaned before use. After setup, each individual measurement typically takes 1–2 min to complete.

TABLE 31.4 Sampling of Common Bubble Tube Viscometer Types

Type	Name	Manufacturer/Vendor	Range (St)	Price
Bubble viscometers	Gardner Standard Bubble Viscometers	Gardner/Gardco	0.5–5.5	\$\$
Bubble viscometers	Bubble Viscometer Kits	Cole Parmer	0.005–0.320	\$–\$\$

The viscometer needs to be held in a vertical position during measurement (some manufacturers supply a stand to hold and flip the tubes) and care must be taken not to heat the liquid in the tube when handling. They are not suitable for use with opaque liquids. They are not automated and require manual timing; however one can easily use multiple tubes to compare the viscosity of different fluids. The primary sources of error that arise in the use of a falling ball viscometer are related to temperature effects (it is not easy to supply external temperature control), handling, and contamination.

31.3.3 Rotational Viscometers

Theory of Operation Rotational viscometers determine viscosity by measuring the resistance on a shaft rotating in the liquid of interest. They are designed to make a direct measurement of the absolute viscosity μ . A schematic of a rotational viscometer is shown in Figures 31.7a and 31.8a. Figure 31.7b shows a commercially available Brookfield rotational viscometer. The viscometer includes a spindle (cylinder), attached to a rotating shaft. The spindle is placed in the liquid of interest and rotated at constant speed. The torque required to rotate the spindle is measured and then related to the fluid viscosity. Although their setup is somewhat more elaborate than that of the falling ball or capillary viscometers, they can be used to study the behavior of non-Newtonian fluids.

The theory of operation of a rotational viscometer is based on the Couette flow model for fully developed, steady, incompressible laminar flow between two surfaces, one of which is moving. If we model a viscometer as a rotating cylinder inside a stationary cylinder as shown in Figure 31.7a, the Couette flow solution for the wall shear τ , assuming the gap between the two cylinders is small and the velocity varies linearly, is given as

$$\tau = \mu \frac{du}{dy} = \mu \frac{\omega R_i}{R_o - R_i} \quad (31.27)$$

where ω is the rotational speed of the cylinder, R_i is the radius of the rotating cylinder, and R_o is the radius of the outer, stationary cylinder. The torque due to this wall shear is then equal to the shear force times the area over which the shear acts times the moment arm

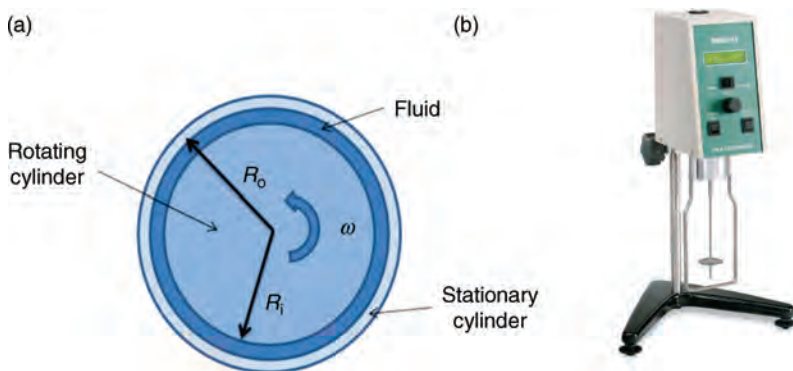


FIGURE 31.7 (a) Schematic of a cylinder in cylinder rotational viscometer and (b) spindle-type viscometer (Courtesy of Brookfield Engineering Labs).

(in this case the inner radius):

$$T = \tau A R_i = \mu \frac{\omega R_i}{R_o - R_i} (2\pi R_i L) R_i \quad (31.28)$$

where T is the torque and A is the area of the rotating cylinder ($2\pi R_i L$). Under steady state conditions the torque due to the viscous forces equals that which is applied to keep the cylinder rotating. Therefore, the viscosity can be measured using the measured torque value and the geometry of the system:

$$\mu = T \frac{R_o - R_i}{(2\pi \omega R_i^3 L)} \quad (31.29)$$

In practice, rotational viscometers come with a variety of rotating spindles and calibration information that allows one to relate the torque measurement to viscosity. A typical calibration equation is in a form such as:

$$\mu = \frac{T}{c\omega} \quad (31.30)$$

where T is the measured torque at a given rotation rate, w and c is the calibration constant used to account for geometry and end effects. Some viscometers are equipped with a digital readout.

Types of Viscometers/Options There are three main types of rotational viscometers: Concentric cylinders, cone and plate, and parallel plate. The concentric cylinder theory is described above for a rotating inner cylinder which is more common. Systems are also designed to have a rotating outer cylinder to minimize the centrifugal forces which lead to the formation of Taylor vortices. A variation on the concentric cylinder-type viscometer uses a rotating spindle in an infinite medium.

Cone and Plate viscometers (see Figure 31.8b) are designed to provide a uniform shear rate across the rotating plate. They consist of a conical surface and a flat plate, separated by a small gap that is filled with the liquid of interest. One of the plates is rotated and the torque required to hold the other in place is measured. The theory of operation is similar to that described above for the concentric cylinder viscometer, although the geometry is different. The cone angle in Figure 31.8b is designed to provide the uniform shear rate.

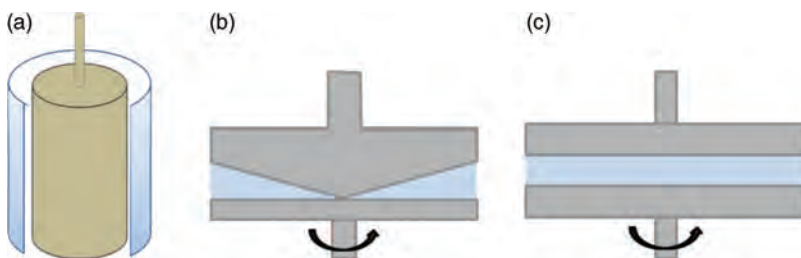


FIGURE 31.8 Geometries typically used in rotational viscometer (a) cylindrical, (b) cone and plate, and (c) parallel plate.

Parallel plate viscometers provide a non-uniform shear rate. They consist of two parallel plates (see Figure 31.8c) separated by a gap filled with the liquid of interest. The shear rate acting on the fluid depends on the radial location. One plate typically rotates and the torque required to hold the other is measured. Again, the theory is similar to that described above; however in this case the shear rate varies with radial distance. Table 31.5 presents a listing of types and manufacturers of rotational viscometers.

Summary Rotational viscometers are more expensive than the other drag-type viscometers but they are somewhat easier to use. Measurements can be made rapidly. Rotational speed can be altered to vary shear rates and test non-Newtonian fluids. The spindles, cones and plates must be carefully cleaned before use so the testing of multiple fluids is more complex. The primary sources of error that arise in the use of a rotational viscometer are related to temperature effects, set up errors (the system must be level and care must be taken to avoid end effects and eccentricity) and calibration of the torque measurement system. Some rotational viscometers have built in temperature control. Process versions exist that can be used in-line (see Table 31.5).

31.3.4 Flow-Type Viscometers

31.3.4.1 Capillary Viscometers

Theory of Operation Capillary viscometers determine viscosity through measurement of the flow rate of the liquid traveling through a capillary tube. A capillary tube is one with a large length to diameter ratio (i.e., long and skinny). A schematic of a capillary tube

TABLE 31.5 Sampling of Common Rotational Viscometer Types

Type	Name	Manufacturer/Vendor	Range (cP)	Price
Cylindrical	Dial Reading	Brookfield Eng./Gardco	1–64 M	
Cylindrical	DV-E; DV-I Prime; DV-II+ Pro; DV-II+ Pro EXTRA	Brookfield Eng./Gardco	1–320 M	\$\$–\$\$\$\$
Cylindrical	HAAKE Rotational Plus Viscometer	Thermo Scientific; Cole Parmer	0.3–4000	\$\$\$\$\$
Cylindrical	Thomas Stormer Viscometers	Thomas Stormer/Gardco	332–2000	\$
Cylindrical	KU-2	Brookfield Eng./Gardco	27–5,274	
Cylindrical in-line	VTA Pneumatic Viscosel; VTE Electric Viscosel	Brookfield Eng.	0–10,000	
Cylindrical in-line	TT Series In-Line PV Process	Brookfield Eng.	2–10,000,000	
Cone and plate	LV, RV, HA and HB Series Viscometers	Wells-Brookfield/ Brookfield Eng.; Gardco	0.3–7,864,000	
Cone and plate	High Shear CAP-1000+, CAP-2000+	Wells-Brookfield/ Brookfield Eng.; Gardco	20–1,500,000	\$\$\$\$\$

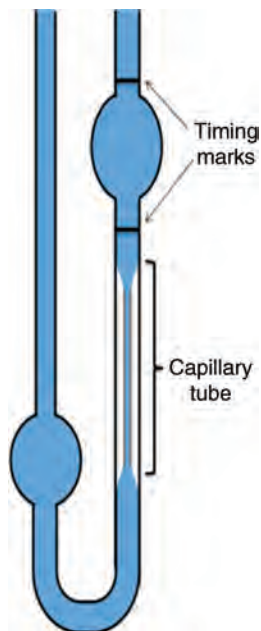


FIGURE 31.9 Capillary tube viscometer (Ostwald type).

viscometer is shown in Figure 31.9. Capillary viscometers are typically made of glass and consist of a bulb reservoir connected to the capillary tube. The theory of operation for a capillary tube viscometer is based on the Poiseuille model of laminar flow which describes flow through a round pipe. The volume flow rate, Q in a pipe can be derived from the Navier–Stokes equations for steady, laminar, and fully developed, incompressible flow as

$$Q = -\frac{\pi R^4}{8\mu} \frac{dP}{dy} \quad (31.31)$$

where R is the pipe radius, μ is the viscosity and dP/dy is the pressure gradient (i.e., the change in pressure over the length of the pipe) which is the driving head for the flow. In the case of a vertical tube with both ends open to the ambient, the pressure gradient is caused by the hydrostatic pressure gradient:

$$\frac{dP}{dy} = -\rho g \quad (31.32)$$

where ρ is the density and g is the gravitational constant. Combining equations and rearranging

$$\frac{\mu}{\rho} = \nu = \frac{\pi R^4 g}{8Q} \quad (31.33)$$

Which allows for calculation of the kinematic viscosity, ν , from the measured flow rate and the geometry of the capillary tube. In practical applications, capillary viscometers are carefully calibrated and the manufacturer supplies a calibration constant. The user is

instructed to measure the time for the fluid to travel a specified distance and then the kinematic viscosity is calculated as:

$$\nu = Kt \quad (31.34)$$

where K is the manufacturer supplied calibration constant and can be a function of temperature depending on the viscometer type.

Types of Viscometers/Options There are four primary types of capillary tube viscometers. They are the original Ostwald, the Modified Ostwald, the Suspended level (Ubbelohde) and the reverse flow capillary viscometers. Each is described below.

The Ostwald Viscometer is one of the simplest capillary tube viscometers. As shown in the schematic (Figure 31.9), the viscometer consists of a bulb connected to a long capillary tube. To use the viscometer one partially fills it and then draws the fluid to the upper mark above the right side bulb (typically using a syringe system). The fluid is released to flow through the capillary tube and the time for the upper bulb to empty (fluid level at upper marks to lower marks) is measured. Some of the problems associated with the use of the Ostwald viscometer include the need to keep the viscometer vertical, the requirement for a specific volume of fluid and the effect of temperature on the viscosity measurement.

A number of Modified Ostwald-type viscometers exist. These include the Cannon–Fenske routine viscometer, Pinkevitch viscometer, Zeitfuchs, (see Viswanath et al. 2007). Each is designed to address some of the sources of error found in the Ostwald type. For example, the Cannon–Fenske Routine viscometer is designed to minimize the effect of tilt angle by placing the upper and lower bulbs along the same vertical axis.

Suspended Level Viscometers are designed to address loading issues by using a constant pressure gradient (driving head) during measurement of viscosity. They do this by suspending the test liquid above the capillary tube and using a pressure equalization tube. They include the Ubbelohde and Cannon–Ubbelohde-type viscometers. To determine viscosity, the test liquid is loaded into the upper bulb and then released. The liquid flowing through the capillary is separated from the reservoir bulb at the bottom. The third tube which connects the bottom of the capillary tube to the ambient ensures that the only pressure difference between the top of the bulb and the bottom of the capillary is that due to the hydrostatic pressure that is, the weight of the liquid.

Reverse Flow Viscometers are used to measure the viscosity of opaque fluids (although they can also be used to measure that of transparent liquids). They measure the flow rate through a “dry” capillary tube so that the leading edge of the opaque fluid can be easily identified. Reverse Flow viscometers must be cleaned between each measurement.

There are a number of variations on the standard capillary tube viscometer, including small volume (micro or semi-micro) viscometers requiring 1 mL or less of fluid (useful in the measurement of the viscosity of blood and plasma), dilution viscometers with extra large reservoirs for dilution of the sample and vacuum viscometers for fluids with high viscosities such as asphalt. There also exist more rugged capillary tube viscometers that are used under continuous flow conditions in industrial applications, and disposable capillary models to avoid the labor (and error potential) associated with cleaning. Table 31.6 provides a sampling of the various capillary tube viscometers.

Summary Capillary tubes are used to measure viscosity for a wide range of fluids from oils to blood/plasma and even asphalt. They are relatively inexpensive (although generally

TABLE 31.6 Sampling of Common Capillary Viscometer Types

Type	Name	Manufacturer/Vendor	Range (St)	Price
Cannon-Fenske Routine	Cannon-Fenske Routine	Cannon	0.5–100,000	\$
Ubbelohde	Cannon-Ubbelohde; Ubbelohde	Cannon	0.5–100,000	\$
Reverse flow	BS/U-Tube; Zeitfuchs Cross-Arm; Cannon-Fenske Opaque	Cannon	0.5–100,000	\$
Small volume	Cannon-Manning; Cannon-Ubbelohde Semi-Micro	Cannon	0.5–20,000	\$
Dilution	Cannon-Ubbelohde Four-Bulb Shear; Cannon-Ubbelohde Dilution	Cannon	0.5–20,000	\$
Vacuum	Asphalt Institute; Cannon-Manning; Modified Koppers Vacuum; and Zeitfuchs Cross-Arm	Cannon	0.036–5,800,000 P	\$
In-line viscometer	KV100 Capillary Viscometer	Brookfield Eng.	0–500 cP	

more costly than the falling ball and bubble tube viscometers). They require skill to use and can be labor intensive. Measurements take 1–5 min. Use requires manual timing and a method for filling the viscometer (a syringe system is typically used). The glass tubes are fragile. The tubes must be carefully cleaned before use but the viscosity of different fluids is easily measured using different tubes. The primary sources of error that arise in the use of a capillary viscometer are related to temperature effects, set-up errors (the tubes must be mounted vertically) and contamination. The tubes can be mounted in a temperature-controlled bath to minimize temperature effects and care must be taken not to affect the fluid temperature when handling. The calibration constant can be sensitive to temperature. Process versions exist that can be using in-line.

31.3.5 Orifice-Type (Cup) Viscometers

Theory of Operation Orifice-type viscometers measure kinematic viscosity by comparing the time it takes the fluid to pass through an orifice (the efflux time) to the time that it takes a fluid of known viscosity to pass through the same orifice. They generally consist of a fluid reservoir from which the fluid flows, an orifice and a capturing reservoir (see Figure 31.10). Although they were originally designed using the Poiseuille flow model, it was determined that entrance and exit effects in this type of device are significant. Therefore, cup methods only provide a relative measurement. Absolute values of viscosity cannot be measured using this type of viscometer.

In practice, a “cup” or reservoir is filled with a specified quantity of the fluid of interest and allowed to equilibrate thermally. Then a valve is opened at the bottom of the

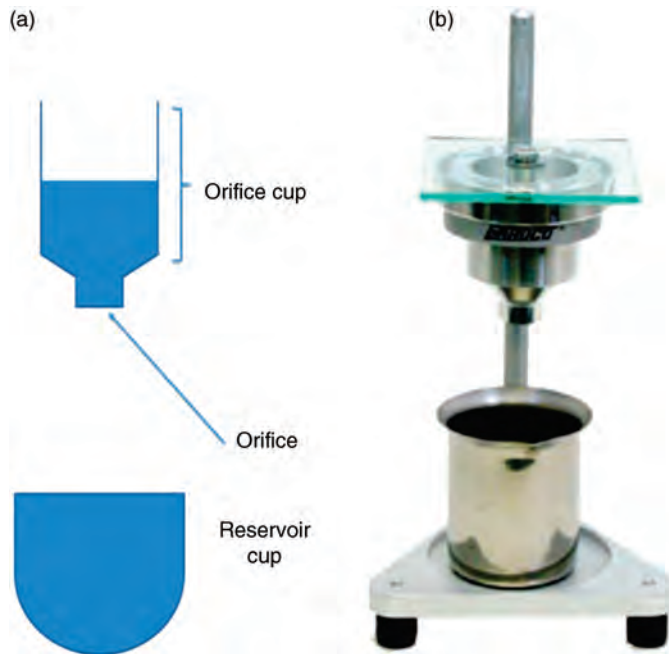


FIGURE 31.10 (a) Schematic of cup viscometer and (b) image of a Ford cup viscometer (Courtesy of Paul N. Gardner Company, Inc.).

cup and the time for the cup to empty is measured. Viscosity is then calculated using an *empirical* equation:

$$\nu = k * t \quad (31.35)$$

where k is supplied by the manufacturer. These types of viscometers are commonly used to measure the viscosity of oils, varnishes, and paints. Specifications in certain industries or for certain industrial processes are often closely tied to one particular type of cup measurement, and thus it may be difficult to use or generalize viscosity information found with other viscosity measuring methods to these industrial applications.

Types of Viscometers/Options Cup viscometers are commonly used for in field measurements. There are a variety of cup types including the early models used in the petroleum industry such as the Saybolt, Redwood, and Engle Cups. The Ford, Zahn, and Shell viscosity cups (to name just a few) are more commonly used for measuring the viscosity of paints, varnishes, and lacquers. A sampling of cup viscometer types is listed in Table 31.7.

Summary Cup viscometers are relatively inexpensive but generally not very accurate. In fact cup “viscometers” do not produce true viscosity measurements, only relative ones. Their primary drawback is that information gained using them is not directly translatable or comparable to information obtained by other methods (although empirical relations making comparisons to other methods abound). In practice this leads to significant “legacy” effects, necessitating the continued use of the same exact cup method if comparisons to historical data are desired. Thus a particular type of cup’s use is typically closely

TABLE 31.7 Sampling of Common Cup Viscometer Types

Type	Name	Manufacturer/Vendor	Range (cSt)	Price
Dip viscosity cups	EZ (Equivalent Zahn) Viscosity Cups	Zahn/Gardco	10–1401	\$
Dip viscosity cups	S90 Zahn Signature Cups	Zahn/Gardco; Spec- trum Chemical	15–1627	\$
Dip viscosity cups	Gardco/DIN 4mm Dip Viscosity Cups	DIN/Gardco	38–545	\$
Dip viscosity cups	Ford Dip Viscosity Cups	Ford/Gardco	2–1413	\$
Dip viscosity cups	Gardco/Fisher Dip Viscosity Cups	Fisher/Gardco	11–1125	\$
Dip viscosity cups	Gardco/ISO Mini Dip and Orifices	ISO/Gardco	4.6–823	\$
Dip viscosity cups	Norcross [®] Shell Viscosity Cup	Gardco	3.3–400	\$
Laboratory or “Ring Stand” viscosity cups	Ford Standard Viscosity Cups	Gardco	29–1413	\$
Laboratory or “Ring Stand” viscosity cups	Gardco/ISO Viscosity Cups	Gardco	4.6–2611	\$
Standard cups	Gardco/DIN Standard Viscosity Cups	DIN/Gardco	38–545	\$
Standard cups	The Parlin Cups	Gardco	7–15000	\$
Standard cups	Gardco/Fisher Standard Viscosity Cups	Fisher/Gardco	11–1125	\$
Standard cups	Gardco/ISO Standard Viscosity Cups and Orifices	ISO/Gardco	4.6–2611	\$
Standard cups	Stein Hall Viscosity Cup	Gardco		\$

tied to a certain industry, industrial process, or product and the primary benefit of using such an “industry standard cup” is that their measurements can be directly compared to previous industrial knowledge or standards developed using the same method. Cup viscometers also require some skill to use (the cup requires steady holding during measurement) and can be labor intensive. Some models include a ring stand, others have handles, to hold the cup still and improve accuracy. Typical measurements are made in the field and each can take 1–2 min. The primary sources of error that arise in the use of a cup viscometer are related to temperature effects, setup errors, unsteadiness, and contamination. Most cups do not come with temperature control and this can affect the measurement although Saybolt-type cup viscometers do include temperature control. Cups must be properly cleaned because even slight clogging of the orifice can cause significant error.

31.3.6 Vibrational (Resonant) Viscometers

Theory of Operation Vibrational viscometers determine viscosity by forcing a resonator to vibrate in the fluid of interest and then measuring the damping that is associated with fluid viscosity. This dampening effect is measured by one of three methods: (a) through

TABLE 31.8 Sampling of Common Vibrational Viscometer Types

Type	Name	Manufacturer/Vendor	Range (cP)	Price
Vibrational	Process viscometers	Vindum Eng.		
Vibrational	Portable viscometers	Vindum Eng.		
Vibrational	Special viscometers	Vindum Eng.		
Vibrational	Reactor viscometer	Vindum Eng.		
Vibrational	SV-A100	Gardco; Spectrum Chemical	1,000–100,000	\$\$\$\$\$
Vibrational	SV-10	Gardco	0.3–10,000	\$\$\$\$
Vibrational	SV-100	Gardco	1,000–100,000	\$\$\$\$
Vibrational	In-line viscometer	Nametre/Galvanic Applied Sciences		
Tuning fork vibrational	SV-A1 tuning fork vibration viscometer	Gardco; Spectrum Chemical; Cole Parmer	0.3–1,000	\$\$\$\$–\$\$\$\$\$
Tuning fork vibrational	SV-A10 tuning fork vibration viscometer	Gardco; Spectrum Chemical; Cole Parmer	0.3–10,000	\$\$\$\$\$–\$\$\$\$\$

measurement of the power required to maintain a constant vibration (the higher the viscosity of the fluid, the higher the power need); (b) measurement of the signal decay time upon halting the vibration (higher viscosity fluids will have a shorter decay time); or (c) looking at the frequency of the resonator as a function of the phase angle between the excitation and response signals.

Types of Viscometers/Options Vibrational viscometers are commonly used in in-line process measurement applications, although lab versions exist. Table 31.8 presents information on a sampling of commercially available vibrational viscometers. The most common vibrational viscometers are configured as a tuning fork, an oscillating sphere, or a vibrating rod. Tuning fork vibrational viscometers have two sensor plates which are immersed in the liquid of interest. The plates (which have the same natural frequency) are made to vibrate at the same amplitude using an electromagnetic force. The viscosity of the liquid is determined by measuring the current required to maintain the constant amplitude motion. Figure 31.11 shows a schematic of one type of tuning fork viscometer.

The oscillating sphere viscometers use the dampening associated with the torsional motion of a sphere immersed in the fluid of interest to measure viscosity. Vibrating rod viscometers do the same with a rod shaped sensor. Tuning fork vibrational viscometers are capable of simultaneously measuring density and viscosity; however the oscillating spheres and rod types require separate measurement of density (see Retsina et al., 1987, for more details).

Summary Vibrational viscometers are more expensive than the other types of viscometers but they are much easier to use. Once they are installed inline they require minimal skill and measurements can be made quickly. Their fundamental lack of moving parts makes them more robust and resilient, and thus better suited for use in harsh environments. The primary sources of error that arise in the use of a vibrational viscometer are related to calibration and drift errors.

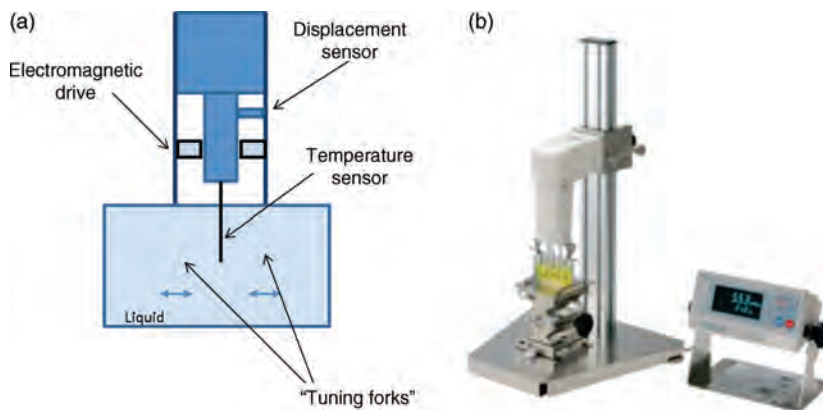


FIGURE 31.11 (a) Schematic of a vibrational viscometer and (b) Image of an SV-A Series Sine Wave Vibro Viscometer (Courtesy of A&D Weighing).

31.3.6.1 Viscosity Calibration Fluids Viscosity standards are standard fluids of known viscosity that are used to calibrate a viscometer. Most manufacturers recommend three or more fluids specified by NIST that can be used to calibrate a viscometer. Viscosity standard oils can be found through Sigma–Aldrich, Gardco, Cole–Parmer or other similar companies.

31.4 ASTM STANDARDS FOR MEASURING VISCOSITY

There are a variety of published standards available. In this section we focus on the standards put out by the American Society for Testing and Materials (ASTM). Many industries adhere to specific guidelines provided by these or similar standards.

The ASTM viscosity standards come in two types: (a) standards focused on methods (method standards) and (b) standards based on materials (material standards). A method standard describes a viscosity measurement method that can be applied in a variety of circumstances. An example of this would be the Standard Test Method for High-Shear Viscosity Using a Cone/Plate Viscometer, D-4287, which specifies a method for using a cone and plate viscometer and in particular applies this to paint. Because such standards specify the method, they can be applied beyond the particular material discussed in the standard. A material standard outlines multiple techniques that can be used to measure the viscosity of one specific material. For example, the Standard Test Methods for Viscosity of Adhesives, D1084, provides four different methods to test free-flowing adhesives with a viscosity range from 50 to 20 k cP. One limitation of the material standards is that they have a narrow focus.

Table 31.9 provides a starting point for finding both types of standards. In this table the different methods (viscometer type) are listed by column and the material types are listed by row. For example, if one wishes to use a cup-type viscometer to measure the viscosity of paints ASTM D1200-10, D4212-10 and D5125-10 apply. One can also utilize the ASTM website to search for standards. The bolded standards in Table 31.9 are the recommended starting standards as they are the most general and provide a good list of referenced documents.

TABLE 31.9 ASTM Standards

Material/Method	Rotational Viscometers	Capillary Viscometers	Falling Object Viscometers	Cup-Type Viscometers	Bubble Viscometer	Tapered Plug/Bearing Viscometer
Iron and steel products	D562; D2196					
Construction	D4016-08; D7226-06e1; D4402-06; D7496-09; D562; D2196	D2171M-10; D2170M-10; D4957-08; D4603-03	D1823			
Petroleum Products, lubricants and fossil fuels	D2983-09; D7042-04; D6896-03; D2161-05e1; D7279-08	D5481-10; D446-07; D445-09	D7483-08		D1545-07	D6616-07; D4683-09; D4741-06
Paints, related coatings and aromatics	D2196-10; D7394-08; D562-10; D4287; D6606; D7395-07		D5478-09; D6606-00; D4040-10; D1343-95	D1200-10; D4212-10; D5125-10;	D1725-04	
Plastics	D1824-95	D5225-09; D4603-03; D446	D1823			
Rubber	D1992-91; D2196	D1646-07				
Medical and surgical materials	D2196	D4603				
Adhesives	D2556-93a; D562; D1084-08					
High temperature	D2669-06; D6267-08					

31.5 QUESTIONS TO ASK WHEN SELECTING A VISCOSITY MEASUREMENT TECHNIQUE

This section is intended as a guide to the practicing engineer on how to choose the appropriate viscometer. It is structured as a series of nonhierarchical questions. Table 31.10 also provides a summary comparison of the different viscometer types.

1. *Do you need an online/process measurement, or an offline measurement?*

If you need to continuously monitor the viscosity of an industrial process stream online or *in situ* your choices of available technique will be limited, and the cost of your system will be higher. On the positive side, very little operator labor will be required once the system is installed and there will be no need to remove and handle samples from the process stream. Measurements can be taken at much shorter time intervals, and the viscosity will be known at exactly the conditions (temperature, pressure) associated with the process. The main viscometry techniques available for online measurement are resonant, capillary, and rotational. More information on the measurement range and suppliers of these viscometers can be found in Tables 31.5, 31.6 and 31.8.

2. *What is the viscosity range that you will be measuring?*

Matching the viscometer to the viscosity range to be measured is one of the most important tasks associated with selecting a viscometer. Since many techniques measure the time for a certain amount of fluid to flow (e.g., cup or capillary types), or for an object to fall through the fluid (e.g., falling ball or rising bubble types), if the fluid viscosity is higher than optimal for a given viscometer the measurements may take an exceedingly long time to conduct. On the other hand, if the viscosity is too low key assumptions related to the operation of the viscometer (e.g., the low Reynolds number assumption) may be violated and accuracy will be lost. Rotational-type viscometers simply will not give a reading, and may in fact be damaged, if the fluid being measured is too viscous for the instrument and settings selected. On the other hand if the viscosity is such that the reading is less than 10% of the instrument's full range (for a given setting) the uncertainty of the measurement will become unacceptably high. Thus good results necessitate careful matching of the viscosity range and the instrument. Operating ranges for many common types of viscometers are listed in Tables 31.3–31.8.

3. *What other characteristics of the fluid must be considered?*

Other physical properties of the fluid beyond viscosity are also important to consider. For example, certain techniques or instruments lend themselves to use with opaque fluids (e.g., the reverse flow viscometer) while others cannot be used because the technique depends on being able to see through the fluid (e.g., falling ball). Likewise, chemical compatibility of the fluid with the exposed viscometer surfaces is important to consider, as are any issues associated with safe handling of the fluid. Also various properties such as how difficult the fluid will be to clean off of the viscometer surfaces, and how likely the fluid is to solidify and clog are critical. Certain specialized fluids, like blood, have specific viscosity instruments designed specifically for use with them.

4. *Do you wish to measure kinematic or dynamic viscosity?*

Each type of viscometer fundamentally measures either the absolute viscosity of a fluid (μ) or the kinematic viscosity (ν). In order to convert between the two a separate

TABLE 31.10 Comparison of Viscometer Types

Type	Falling Object	Bubble Tube	Rotational	Capillary	Cup	Vibrational
Viscosity range	0.2–70 k cP	0.005–5.5 cSt	0.3–320 M cP	0.5–100 k cSt (general), 0.5–20 k cSt (small vol- ume) up to 5.8 MP (vacuum)	4–2600 cSt	4–2600 cSt
Labor/skill level	High	High	Med	High	Med	Low
Automated?	No	No	Yes	Some	No	Yes
In-line?	No	No	Some	Some	No	Yes
Expense	\$100–\$200 Gilmont \$1000 Haake	(\$50–100 each or \$400–1000 for a kit)	Expensive	\$200–\$300		Expensive

measurement of the fluid's density is required. Thus care must be exercised in selecting a viscometer if one wishes to avoid the need for this additional measurement.

5. *How much and what type of labor is required?*

The quantity of labor (time) and skill level required to make accurate viscosity measurements varies significantly among different viscometer types. Broadly speaking the lower cost viscosity measurement instruments (falling ball, rising bubble, and capillary tube) are more labor intensive to use and require significantly more skill (operator training) than the more expensive viscometer types (rotating disk, vibrating rod, etc.). The frequency of measurements required, as well as the labor associated with maintenance, calibration, and cleaning of the viscometer should be carefully considered in this context.

6. *What is the required accuracy?*

Another major factor influencing the type of viscometer to choose is the required accuracy. Here some of the simplest models (e.g., capillary and falling ball) compare quite well on the basis of accuracy with more expensive techniques, assuming a sufficiently skilled and careful operator. However the old truism that increased accuracy leads to increased expense does generally hold within a given class of viscometer. As discussed in point 5 above, labor costs may be significant in regards to the cost/accuracy decision.

7. *Are there standard methods associated with your industry or fluid?*

Viscosity measurements for many substances, for many industries, and for the use of many types of viscometers are the subject of government or industrial standards. For example, the ASTM standards governing viscosity measurement are discussed at length in Section 31.3. Likewise many industries or industrial processes are closely tied to particular viscosity measurement techniques, even if these are out-moded. In such cases there may be a considerable body of industrial or process control knowledge that is not readily translatable to more standard viscosity measurements (e.g., the use of the "Stein–Hall Cup" measurement with starch based adhesives used in corrugated box manufacturing). In such cases sticking with the "standard method" may be the most desirable approach.

8. *Is temperature control needed? Is viscosity as a function of temperature needed?*

The viscosities of liquids and gases are a strong function of temperature, and this should always be kept in mind when collecting and reporting viscosity data. Some viscosity measurement techniques lend themselves more readily to maintaining temperature control of the fluid during the measurement (e.g., the Haake falling ball viscometer has a built-in temperature bath capability, for other types that capability must be added *ad hoc*). Likewise some viscometers (e.g., Brookfield rotational viscometer with small sample adapter) lend themselves to measuring the viscosity of a fluid over a wide range of temperatures.

9. *Quantity of fluid needed for the measurement.*

This is an important factor to consider when dealing with scarce or expensive fluids (e.g., blood). Viscosity measurement techniques can vary by more than an order of magnitude in the amount of liquid required to make a measurement.

10. *Specialized needs.*

If your process involves specialized needs or requirements, that is, measurements at high pressure or measurements at temperature extremes there are viscometers available specifically for those purposes.

REFERENCES

- Bird RB, Stewart WE, Lightfoot EN. *Transport Phenomena*. New York: John Wiley and Sons; 1960.
- Brizard M, et al. Design of a high precision falling-ball viscometer. *Review of Scientific Instruments* 2005;76:025109.
- Cengel YA, Cimbala JM. *Fluid Mechanics, Fundamentals and Applications*. 2nd ed. New York: McGraw Hill; 2010.
- Cristescu ND, Conrad BP, Tran-Son-Tay R. A closed form solution for falling cylinder viscometers. *International Journal of Engineering Science* 2002;40.6:605–20.
- Davis AMJ, Brenner H. The falling-needle viscometer. *Physics of Fluids* 2001;13:3086.
- Goldsmith HL, Mason SG. The movement of single large bubbles in closed vertical tubes. *Journal of Fluid Mechanics* 1962;14.01:42–58.
- Kogan MN. *Rarefied Gas Dynamics*. New York: Plenum Press; 1969.
- Lucas K. Die Druckabhängigkeit der Viskosität von Flüssigkeiten eine einfache Abschätzung. *Chemie Ingenieur Technik* 1981;53:959.
- Munson BR, Young DF, Okiishi TH, Huebsch WW. *Fundamentals of Fluid Mechanics*. 6th ed. New York: John Wiley and Sons; 2009.
- Owczarek JA. *Fundamentals of Gas Dynamics*. Scranton, PA: International Textbook Co.; 1964.
- Poling BE, Prausnitz JM, O'Connell, JP. *The Properties of Gases and Liquids*. 5th ed. New York: McGraw Hill; 2004.
- Retsina T, Richardson SM, Wakeham WA. The theory of a vibrating-rod viscometer. *Applied Scientific Research* 1987;43:325–346.
- Shaughnessy JJr., EJ, Katz IM, Schaffer JP. *Introduction to Fluid Mechanics*. New York: Oxford University Press; 2005.
- Viswanath DS, Ghosh TK, Prasad DHL, Dutt NVK, Rani K. *Viscosity of Liquids. Theory, Estimation, and Data*. New York: Springer; 2007.
- White FM. *Viscous Fluid Flow*. 3rd ed. New York: McGraw Hill Inc.; 2006.
- Vicenti WG, Kruger CH. *Introduction to physical gas dynamics*. New York: Wiley; 1965.

TRIBOLOGY MEASUREMENTS

PRASANTA SAHOO

- 32.1 Introduction
- 32.2 Measurement of surface roughness
 - 32.2.1 Surface profilometer
 - 32.2.2 Optical microscopy
 - 32.2.3 Advanced techniques for surface topography evaluation
- 32.3 Measurement of friction
 - 32.3.1 Inclined plane rig
 - 32.3.2 Pin-on-disc rig
 - 32.3.3 Conformal and non-conformal geometry rig
 - 32.3.4 Environment control
 - 32.3.5 Techniques for friction force measurement
- 32.4 Measurement of wear
- 32.5 Measurement of test environment
 - 32.5.1 Temperature measurement
 - 32.5.2 Thermocouples
 - 32.5.3 Thin film sensors
 - 32.5.4 Radiation detectors
 - 32.5.5 Metallographic observation
 - 32.5.6 Liquid crystals
 - 32.5.7 Humidity measurement
 - 32.5.8 Measurement of oxygen and other gases
- 32.6 Measurement of material characteristics
 - 32.6.1 Hardness
 - 32.6.2 Young's modulus and the elasticity limit
 - 32.6.3 Fracture toughness
 - 32.6.4 Residual stresses
 - 32.6.5 Chemical composition of a surface
- 32.7 Measurement of lubricant characteristics
 - 32.7.1 Analysis of chemical changes
 - 32.7.2 Viscosity measurement
 - 32.7.3 Lubricant oxidation tests

- 32.8 Wear particle analysis
 - 32.8.1 Chemical analysis of particles in lubricant
 - 32.8.2 Analysis based on separation of wear particles
- 32.9 Industrial measurements
- 32.10 Summary

32.1 INTRODUCTION

Tribology as a subject is relatively new, but the practice of tribological principles is older than recorded history. Tribology is defined as the science and technology of interacting surfaces in relative motion and of related subjects and practices. It deals with the technology of lubrication, control of friction, and prevention of wear. Successful design of machine elements depends essentially on the understanding of tribological principles. During contact of two nominally flat surfaces, contact occurs at discrete spots due to surface roughness and adhesion occurs due to intimate contact. When one solid body moves over another it experiences the resistance to motion called friction. The surface damage or material removal that takes place in a moving contact is termed as wear. Surface coatings and treatments are provided to monitor friction and to control wear. The most effective way of friction and wear control is by using proper lubricants either liquid, solid or gas. The lubricant properties and the mechanisms of lubrication constitute the essence of optimum performance and reliability of bearings. The recent emergence of proximal probes and high capability computational techniques has stimulated systematic investigations of interfacial problems with high resolution in micro and nano components leading to the development of the new field of micro/nano tribology.

Tribology is different from other science branches, for example, fundamental physics where theoretical predictions are made long before the experimental validation of the same. Tribology uses a more empirical methodology based on experimental observation and theoretical concepts follow the findings later. A classical example of experimental observation leading to the development of basic tribological concept of a hydrodynamic pressure field operating in a lubricated bearing is Tower's friction experiments. The railway axle bearings were fitted with numerous oil holes in order to supply oil to the bearing. But during operations, oil leaked through the holes and even wooden plugs were unable to prevent the leakage. Tower fitted pressure gauge at the oil holes to find that the oil pressure was capable to support the bearing load. Figure 32.1 is a schematic representation of the partial bearing used by Tower. Tower's results led to the development of hydrodynamic theory of lubrication by Osborne Reynolds.

Most tribological phenomena, for example, friction, wear, and frictional heating, are not intrinsic material properties. These depend on a number of competing factors. For example, to evaluate load capacity of a hydrodynamic bearing one needs to consider viscous heating of the lubricant, cavitation and turbulent lubricant flow, elastic deformation of bearing structure and so on. Friction and wear being chaotic processes, these are described in terms of specific experimental findings that are systematically analyzed to get an engineering model of friction, wear, and lubrication. Tribological investigation

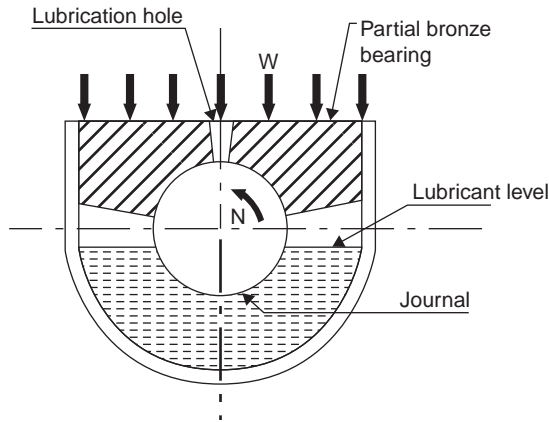


FIGURE 32.1 Schematic representation of the partial bearing used by Tower.

may be categorized into two groups: fundamental research for understanding of basic mechanisms of friction and wear; and applied research for resolving specific industrial friction and wear issues. On both counts, tribological measurements mainly include surface roughness, friction, wear, test environment, material characteristics, and lubricant characteristics. Wear particle analysis and industrial tribology form an important part of tribological measurement.

32.2 MEASUREMENT OF SURFACE ROUGHNESS

The surfaces of any engineering component contain a vast number of peaks and valleys and it is not possible to measure the height and location of each of the peaks. So what is done is to take measurements from a small and representative sample of the surface so chosen that there is a high probability for the surface lying outside the sample to be statistically similar to that lying within the sample. Over the years different methods have been devised to study the topography of surfaces. A brief outline of some of the methods is presented here.

32.2.1 Surface Profilometer

The most common method of studying surface texture features is the stylus profilometer, the essential features of which are illustrated in Figure 32.2. A fine, very lightly loaded stylus is dragged smoothly at a constant speed across the surface under examination. The transducer produces an electrical signal, proportional to displacement of the stylus, which is amplified and fed to a chart recorder that provides a magnified view of the original profile. But this graphical representation differs from the actual surface profile because of difference in magnifications employed in vertical and horizontal directions. Surface slopes appear very steep on profilometric record though they are rarely steeper than 10° in actual cases. The shape of the stylus also plays a vital role in incorporating error in measurement. The finite tip radius (typically $1\text{--}2.5\text{ }\mu\text{m}$ for a diamond stylus) and the

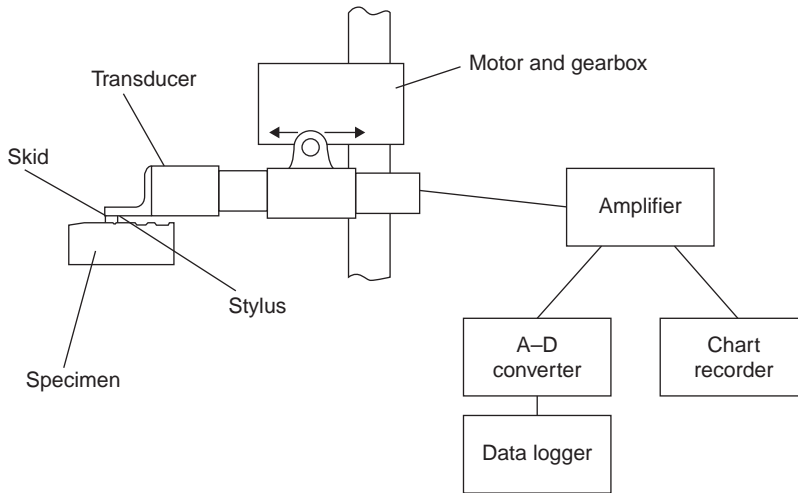


FIGURE 32.2 Component parts of a typical stylus surface-measuring instrument.

included angle (of about 60° for pyramidal or conical shape) results in preventing the stylus from penetrating fully into deep and narrow valleys of the surface and thus some smoothing of the profile are done. Some error is also introduced by the stylus in terms of distortion or damage of a very delicate surface because of the load applied on it. In such cases noncontacting optical profilometer having optical heads replacing stylus may be used. Reflection of infrared radiation from the surface is recorded by arrays of photodiodes and analysis of the same in a microprocessor result in the determination of the surface topography. Vertical resolution of the order of 0.1 nm is achievable though maximum height of measurement is limited to a few microns. This method is clearly advantageous in case of very fine surface features.

32.2.2 Optical Microscopy

In this method, the surface of interest is held to reflect a beam of visible light and then these are collected by the objective of the optical microscope. An image of the surface is produced and is analyzed at very high rates of resolution (up to $0.01\ \mu\text{m}$) by optical interferometers. Depth of field achievable is up to $5\ \mu\text{m}$. But success of the method depends on the reflective property of the material, which limits the use of the same.

Optical methods may be divided into two groups: geometrical methods and physical methods. Geometrical methods include light-sectioning and taper-sectioning methods. Physical methods include specular reflection; diffuse reflection, speckle pattern, and optical interference.

In *light-sectioning method*, the image of a slit is thrown onto the surface at an incident angle of 45° . The reflected image appears as a straight line if surface is smooth and as an undulating line if the surface is rough. In *taper-sectioning method*, a section is cut through the surface to be examined at an angle of θ , thus effectively magnifying the height variation by a factor of $\cot\theta$ and is subsequently examined by an optical microscope. The surface is supported with an adherent coating that prevents smearing of the contour during

the sectioning process. The taper section is lapped, polished, and lightly heat tinted to provide good contrast for optical examination. The process suffers from the disadvantages like destruction of test surface and tedious specimen preparation.

In *specular reflection method*, gloss or specular reflectance that is a surface property of the material and function of reflective index and surface roughness, is measured by gloss meter. Surface roughness scatters the reflected light and affects the specular reflectance. Thus a change in specular reflectance provides a measure for surface roughness.

Diffuse reflection method is particularly suitable for on-line roughness measurement during manufacture since it is continuous, fast, non-contacting and non-destructive. This method employs three varieties of approaches. In total integrated scatter (TIS) approach, one measures the total intensity of the diffusely scattered light and the same is used to generate the maps of asperities, defects and particles rather than micro-roughness distribution. The diffuseness of scattered light (DSL) approach measures a parameter that characterizes the diffuseness of the scattered radiation pattern and relates the same to surface roughness. In angular distribution (AD) approach, the scattered light provides roughness height, average wavelength or average slope. With rougher surfaces, this may be useful as a comparator for monitoring both amplitude and wavelength surface properties.

In *speckle pattern method*, surface roughness is related to speckle, which is basically the local intensity variation between neighbouring points in the reflected beam when a surface is illuminated with partially coherent light. The principle of laser speckle roughness measuring system is schematically shown in Figure 32.3.

The *optical interference technique* involves looking at the interference fringes and characterizing the surface with suitable computer analysis. Common interferometers include the Nomarski polarization interferometer and Tolanski multiple beam interferometer.

32.2.3 Advanced Techniques for Surface Topography Evaluation

A further improvement in the resolution of surface topographic examination is possible by the use of electron microscopes. Two basic types of electron microscopes are available: scanning electron microscopes and transmission electron microscopes. In scanning

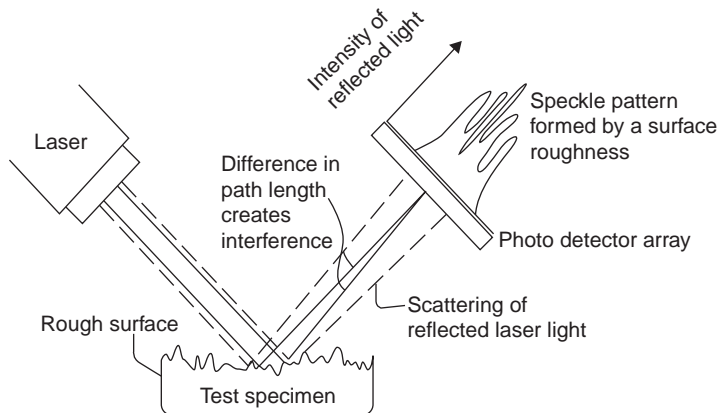


FIGURE 32.3 Schematic of the principle of laser speckle roughness measuring system.

electron microscopy (SEM) a focused beam of high-energy electrons is incident on the surface at a point resulting in the emission of secondary electrons. These are then collected and fed to an amplifier to send an electric signal to a cathode ray tube (CRT). The electron beam is scanned over the surface to have a complete picture. The CRT screen gives a topographical image of the entire area of interest. Depth of field is up to $1000\text{ }\mu\text{m}$, which acts as a primary advantage of this method over optical method. The requirement on size of the specimen to be placed within the vacuum chamber of the instrument raises a drawback of the method. This can be overcome by preparing a replica of the surface.

In transmission electron microscopy (TEM), the focused beam of high-energy electrons is made to transmit through a very thin specimen. Then the deflection and scattering of the electrons is recorded to analyze the surface topography. Preparation of a specimen thin enough to transmit electrons plays a vital role. Sometimes a replica of the surface retaining all the texture features but of a material having greater electron transparency is produced for the same purpose.

Recently, a different type of electron microscopy called scanning tunneling electron microscopy (STM) is in use. It incorporates the electron-tunneling phenomenon through an insulating layer separating two conductors. The sharp pointed tip of a probe forms one electrode and the surface of the specimen the other. The probe is moved by a highly precise positional controller to keep the tunneling current at a steady value. The probe provides an image of the surface under examination. The method is superior to the earlier ones in a sense it does not require any vacuum. Only disadvantage is the proper design of the controller mechanism. The principle of the STM is very simple. Just like in a record player, the instrument uses a sharp needle, referred to as the tip, to investigate the shape of the surface. But the STM tip does not touch the surface. The schematic of the method is shown in Figure 32.4. A voltage is applied between the metallic tip and the specimen, typically between a few millivolts and volts. The tip touching the surface of the specimen results in a current and when the tip is far away from the surface, the current is zero. The STM operates in the regime of extremely small distances between the tip and the surface of only $0.5\text{--}1.0\text{ nm}$, which are typically $2\text{--}4$ atomic diameters. At these distances, the

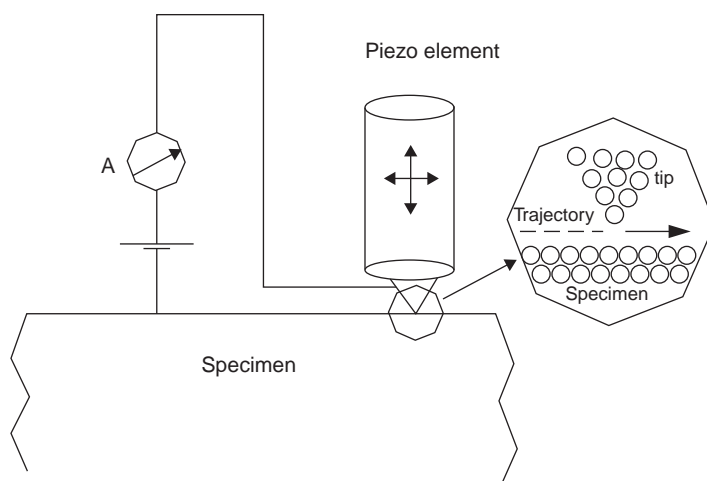


FIGURE 32.4 Schematic of STM.

electrons can jump from the tip to the surface or vice versa. This jumping is necessarily a quantum mechanical process known as “tunneling” and hence the name scanning tunneling microscope. The STMs usually operate at tunneling currents between a few pico-Amperes (pA) and a few nano-Amperes (nA). The tunneling current depends critically on the precise distance between the last atom of the tip and the nearest atom or atoms of the underlying specimen. When this distance is increased a little bit, the tunneling current decreases heavily. As a thumb rule, for each extra atom diameter that is added to the distance, the current becomes a factor of 1000 lower. Thus the tunneling current provides a highly sensitive measure of the distance between the tip and the surface. The STM tip is attached to a piezo-electric element, which changes its length a little bit, when it is put under an electrical voltage. The distance between the tip and the surface can be regulated by adjusting the voltage on the piezo element. In most STMs, the voltage on the piezo elements is adjusted in a manner that the tunneling current always has the same value, say 1 nA. Thus the distance between the last atom on the tip and the nearest atoms on the surface is kept constant. Using the so-called electronics; the distance regulation is done automatically. The feedback electronics continually measures the deviation of the tunneling current from the desired value and accordingly adjust the position of the tip. While this feedback system is active, two other parts of the piezo elements are used to move the tip in a plane parallel to the surface to scan over the surface. In the scanning process, every time that the last atom of the tip is precisely over a surface atom, the tip needs to be retracted a little bit, while it has to be brought slightly closer when the tip atom is between the surface atoms. This automatically leads the tip to follow a bumpy trajectory, which replicates the atoms of the surface. Then this information about the trajectory is available in the form of the voltages that have been applied by the feedback electronics are then finally visualized in the form of a collection of individual height lines or in the form of grey scale/color scale representation or in the form of some three dimensional perspective views.

More recently AFM (atomic force microscope) is developed to investigate surfaces of both conductors and insulators on an atomic scale. Like the STM, AFM relies on a scanning technique to produce very high-resolution, three-dimensional images of sample surfaces. In AFM, the ultra-small forces (less than 1 nN) present between the AFM tip and sample surface are measured by measuring the motion of a very flexible cantilever beam having an ultra small mass. The AFM combines the principles of the STM and the stylus profiler. The important difference between the AFM and the STM is that in the AFM, the tip gently touches the surfaces. The AFM does not record the tunneling current but the small force between the tip and the surface. The AFM tip is attached to a tiny leaf spring, known as the cantilever, which has a low spring constant. The bending of the cantilever is detected with the use of a laser beam, which is reflected from the cantilever. The AFM thus measures contours of constant attractive or repulsive force. The detection is made very sensitive such that the forces as small as a few pico Newton can be detected. Forces below 1 nN are usually sufficiently low to avoid damage to either the tip or the surface. Since AFM does not rely on the presence of a tunneling current, it can also be used on non-conductive materials. Soon after the introduction of the AFM, it was realized that the same instrument could be used to also measure forces in the direction parallel to the surface, that is the friction forces. When modifications are incorporated for atomic scale and micro scale studies of friction, it is termed as the friction force microscope (FFM) or the lateral force microscope (LFM). The FFM usually detects not only the deflection of the cantilever perpendicular to the surface, but also the torsion of the cantilever, resulting

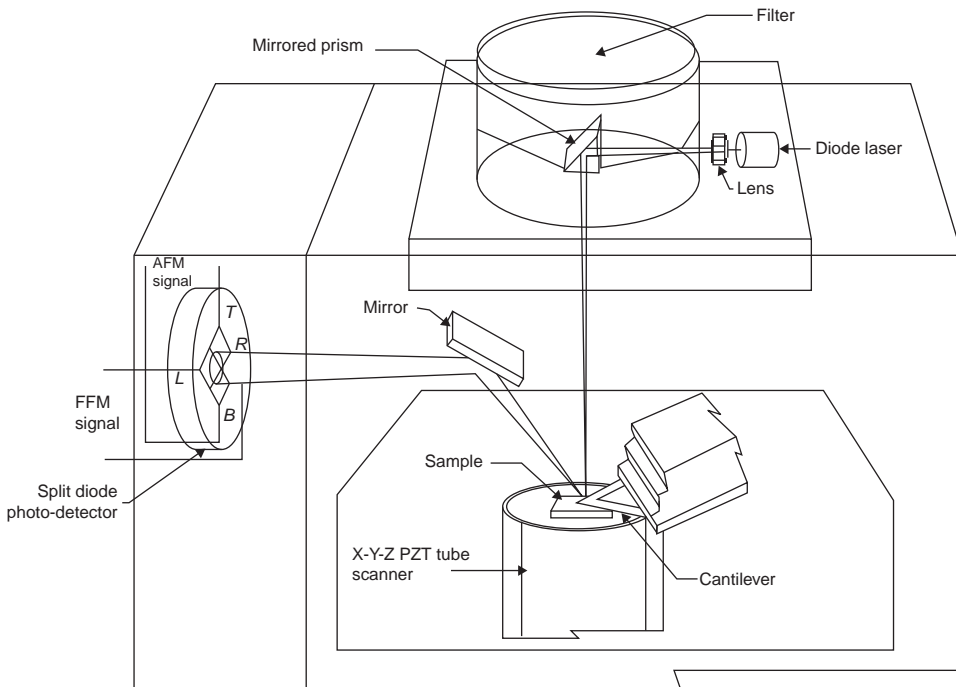


FIGURE 32.5 Schematic operation of AFM/FFM.

from one lateral force. Schematic of the AFM/FFM commonly used for measurements of surface roughness, friction, adhesion, wear, scratching, indentation and boundary lubrication from micro to atomic scales is shown in Figure 32.5.

In all surface profilometric methods, roughness (small-scale irregularities) and form error (deviation from its intended shape) remain coupled in the recorded data. Form error may be subtracted from the recorded data to provide only the roughness features by different means. The two most common methods used in stylus profilometer are (a) use of datum-generating attachments and (b) use of large radius skids or flat shoes. With these the average local level is used as datum and form error or waviness is not recorded. Other methods include the use of filtering the displacement signal corresponding to waviness. Table 32.1 summarizes the comparison of the different roughness measuring methods.

32.3 MEASUREMENT OF FRICTION

Friction is defined as the force of resistance to motion that occurs when a solid body moves tangentially with respect to the surface of another body that it touch. The friction force acts in a direction opposite to that of motion. Even when an attempt is made to initiate the motion, the friction force exists. The friction force required to initiate the sliding is called the static friction force and that required to maintain sliding is called the kinetic friction force the value of which is usually lower than the former for the same combination of material and other parameters. The basic

TABLE 32.1 Summary and Comparison of Roughness Measurement Methods

Method	Quantitative Data	3-D Data	Resolution at Maximum Magnification, nm		On-Line Measurement Capability	Limitations/Comments
			Horizontal	Vertical		
Stylus method	Yes	Yes	15–100	0.1–1	No	Operates along linear track, contact type can damage sample, slow speed in 3-D mapping
Optical methods						
Light sectioning	Limited	Yes	500	0.1–1	No	Qualitative
Taper sectioning	Yes	No	500	25	No	Destructive, tedious specimen preparation
Specular reflection	No	No	10^5 – 10^6	0.1–1	Yes	Semiquantitative
Diffuse reflection	Limited	Yes	10^5 – 10^6	0.1–1	Yes	Smooth surfaces (<100 nm)
Speckle pattern	Limited	Yes			Yes	Smooth surfaces (<100 nm)
Optical interference	Yes	Yes	500–1000	0.1–1	No	Operates in vacuo, limits on specimen size
SEM	Limited	Yes	10	0.1	No	Operates in vacuo, requires replication of surface
TEM	Limited	Yes	0.5	0.02	No	Operates in vacuo, requires replication of surface
STM	Yes	Yes	0.2	0.02	No	Requires a conducting surface, scans small areas
AFM	Yes	Yes	0.2–1	0.02	No	Scans small areas

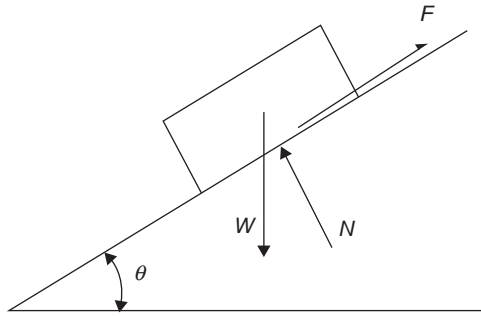


FIGURE 32.6 Measuring friction by an inclined plane test rig.

principle of any friction-measuring instrument is to place two specimens together under a specified normal load and in relative motion while the tangential force resisting motion is measured. Many methods of varying specimen geometry, loading condition and resisting force measurement are available. Different researchers use many ingenious set-ups to investigate different specific cases.

32.3.1 Inclined Plane Rig

The simplest arrangement is the inclined plane test shown in Figure 32.6. A specimen is placed on a flat plane whose inclination with the horizontal is gradually increased until the specimen on it starts sliding. If the inclination at this moment be θ , then $\mu_s = \tan\theta$. Obviously, this method is not capable of evaluating friction in continuous sliding.

32.3.2 Pin-on-Disc Rig

In continuous sliding cases, the rig based on pin-on-disc configuration (Figure 32.7) is used. The pin is held stationary under a normal load while the disc is made to rotate. The

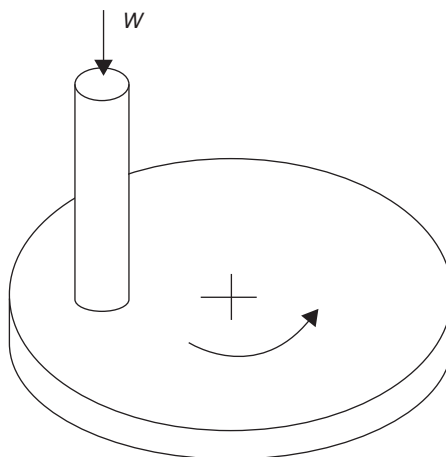


FIGURE 32.7 Pin-on-disc friction measuring device.

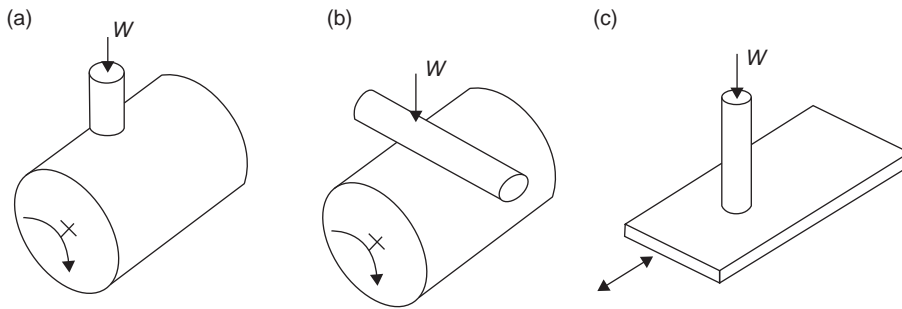


FIGURE 32.8 Friction-measuring devices: (a) pin-on-cylinder, (b) crossed cylinders, and (c) reciprocating rig.

loading can be provided by simple dead weight or by spring loading or hydraulic or pneumatic pressure. The friction force is measured with the help of the calibrated tangential movement of a capacitive or inductive transducer mounted on the stationary specimen. For a multiple-pass arrangement the pin is held at a constant radial distance from the centre of the disc but in single-pass arrangement it is moved radially during the experiment. Other standard arrangements such as pin-on-cylinder, crossed cylinders, and reciprocating arrangement has been shown in Figure 32.8.

32.3.3 Conformal and Non-Conformal Geometry Rig

The test rigs can be classified into two groups depending on the test geometry: conformal and non-conformal geometry test. In conformal geometry test, the profiles of the two contacting surfaces are matched carefully before the experiment is started. In this case the contact pressure is moderate and normally held constant throughout the experiment. The test may then be used to simulate the situations such as brakes, thrust bearings, plane bearings, face seals, and clutches. On the other hand, in non-conformal geometry test (with spherically profiled pin) contact pressure is initially high because on first loading contact is made at a single point and with time pressure reduces due to development of small wear scars. This can be used to simulate the heavily loaded contacts such as gear teeth or to provide accelerated tests of friction and wear of a number of candidate material pairs for specified applications.

32.3.4 Environment Control

For accurate investigation the friction test must be carried out in an enclosed environment having simulated environmental conditions. Different friction pairs are susceptible to the presence of liquid lubricants, water vapour, gases and so on. If pin-on-disc tests are carried out in presence of liquid lubricant, the results vary due to poorly controlled hydrodynamic conditions at the interface. Sometimes the small deformations of the rig caused by the thermal loading or pressure loading may give rise to experimental scatter. So many other forms of test rigs have been developed for specific applications involving different environmental conditions. For example, space environments are simulated by the use of high-vacuum condition where the entire test rig is installed inside the vacuum chamber.

32.3.5 Techniques for Friction Force Measurement

Two basic types of device are commonly used for the measurement of friction force, viz., piezoelectric force gauge and strain gauge transducer. Piezoelectric force gauges give a direct measurement of friction force as an electrical impulse that is recorded electronically. Piezoelectric gauges operate by elastic deflection of a piezoelectric crystal and are sensitive to temperature, vibration and corrosive agents. Piezoelectric gauges are also relatively expensive. Strain gauged beams are comparatively cheaper. In this case, friction force is usually measured from the bending of a beam arranged perpendicularly to the direction of the friction force and strain gauges are mounted on the beam to record the deflection of the beam. Strain gauged beams are effective in recording steady friction forces where piezoelectric force gauges are unsuitable. Only difficulty with strain gauged beam is its failure to record rapid change in friction force.

32.4 MEASUREMENT OF WEAR

Wear is the removal of material from one or both of two solid surfaces in relative motion (sliding, rolling, or impact). Wear occurs as a natural consequence and mostly through surface interactions at asperities. Though wear generally refers to material removal, surface damage due to material displacement with no net change in volume or weight is also termed as wear. Wear is a system response and it is not a material property. Interface wear is strongly dominated by operating conditions. Thus the measurement of wear is mainly influenced by the definition of wear that is considered. For the definition of loss of mass, it is possible to measure wear by simply weighing the worn object before and after a wear test. But this does not allow to measure wear if material displaced by wear gets attached to the worn object. Thus there are three basic methods of measuring wear, viz., detection of change in mass, measurement of reduction in dimension of a worn specimen and profilometry of the worn object. Measurement of mass change is usually done using a sensitive analytical balance since mass changes in wear are usually small. Reduction in dimensions of the worn object is usually measured by connecting a displacement transducer to the surface of the worn object that is directly above the wear spot. Linear variable differential transformer (LVDT) transducers or non-contact inductive proximity probes along with an electronic amplifier are commonly used. The other commonly used technique for evaluation of the worn volume from the wear scar is profilometry where either an optical profile projector, or stylus profilometer or a laser scanning profilometer is used. The optical profile projector is based on projecting an image of the object on a screen and measuring the change in dimensions of the worn specimen. In stylus profilometry technique, a picture of the wear scar is taken and compiled by using several evenly spaced traverses of the stylus over the scar. Then wear is determined from the deepest wear scar profile. In scanning profilometry technique, light from a laser is focused on the surface to measure the dimensions of the wear scar. Other specialized techniques for wear measurement are thin layer activation by radioactivity and ultrasonic interference measurements of dimensional changes. Noise emission and thermal emission from a dynamic contact can also be used for recoding changes in wear condition.

Many different experimental arrangements are used to study sliding wear. These are usually carried out either to examine the process by which wear takes place or to simulate practical situations to generate design data on wear rates and coefficients of friction.

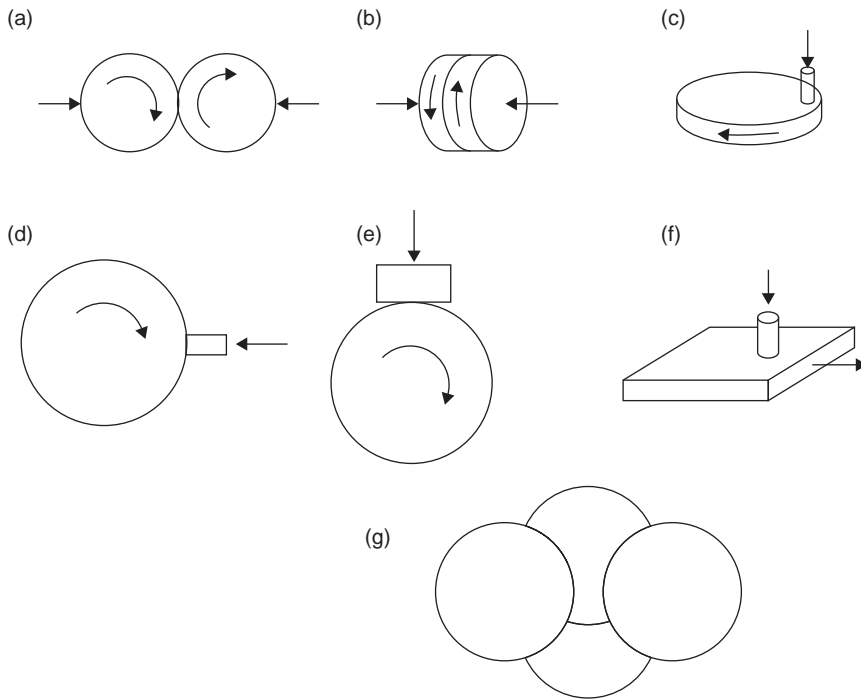


FIGURE 32.9 Sliding wear test arrangements (a) roller/roller, (b) disk/disk, (c) pin on disk, (d) pin on roller, (e) disk on roller, (f) pin on flat, and (g) four ball.

Close control and monitoring of all the variables which may influence wear are essential if the results of a test are to be useful for wider scientific purposes. Figure 32.9 shows the geometrical arrangements in several common types of wear testing apparatus. For adhesive wear between identical materials, the two surfaces are made of same material. For abrasive wear testing, one of the surfaces, generally the larger one, is made of abrasive material. Changes in geometries and arrangements are done for testing of different mechanisms of wear. For two-body abrasive wear, commercial bonded-abrasive paper or cloth is usually used for the counter-face, carrying evenly distributed grit particles of narrow size distribution, bonded to the substrate by a strong resin. In simulating three body abrasion, silica (quartz) particles of a narrow size distribution and from a specified source are fed at a constant rate into the contact region. Figure 32.10 shows schematic diagrams of four types of testing methods for erosive wear. In jet impingement method (a), particles are accelerated in a fluid stream along a nozzle to strike the target material, which is held some way from the end of the nozzle at a fixed angle. In re-circulating loop test (b), a two-phase flow of particles and fluid is driven around a loop of pipe-work where the specimen is kept completely immersed in the flow. In centrifugal accelerator (c), a continuous stream of particles, generated by circular motion of a rotor, strikes the stationary specimens arranged around the rim after rotor. In whirling arm rig (d), two specimens at the ends of a balanced rotor move at high speed through a slowly falling stream of particles, striking them at the peripheral speed of the rotor.

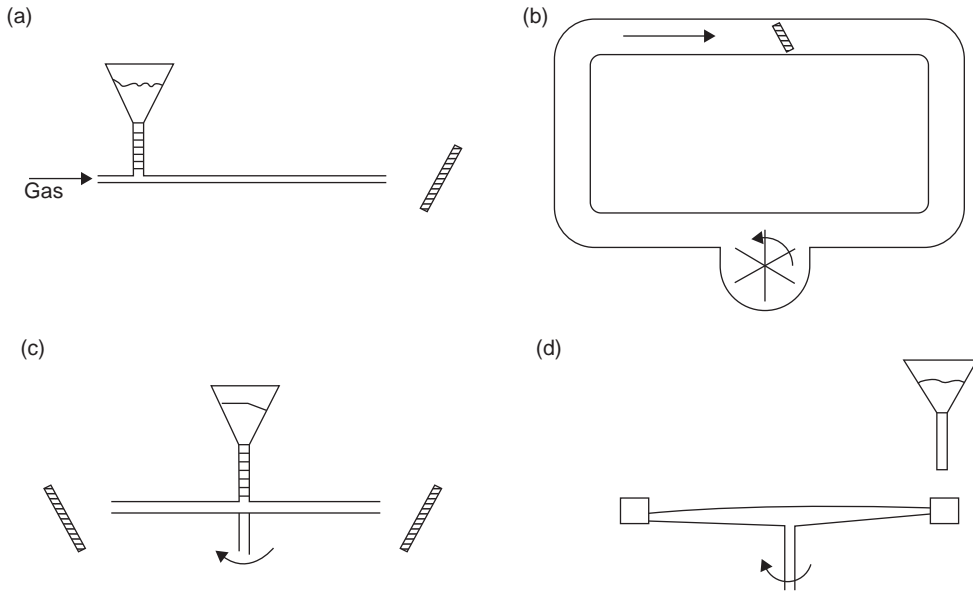


FIGURE 32.10 Schematic of erosive wear test arrangements: (a) jet impingement method, (b) recirculating loop, (c) centrifugal accelerator, and (d) whirling arm rig.

32.5 MEASUREMENT OF TEST ENVIRONMENT

Many experiments often omit the measurement of environmental factors such as temperature, humidity, and oxygen and lead to apparent conflict in experimental data. Careful measurement and control of these environmental factors is necessary for tribology experiments.

32.5.1 Temperature Measurement

In a dynamic contact situation, most of the frictional energy input is generally used up in plastic deformation which is directly converted to heat in the material close to interface with a consequent rise in temperature. Since the heat is produced as a continuous process, temperature gradients will develop in the contacting bodies with the highest temperature occurring at the point of heat release (heat source), that is, the contact surface. In the absence of lubricants, this heat is conducted into the two sliding bodies through contact spots. Contact between two bodies may be approximated as a single contact or as multiple contacts depending on the stress situation. For high contact stress situation, the real area of contact approaches the apparent area of contact and a single contact may be assumed to occur during sliding. Contact of two very smooth surfaces, even at low load, may be assumed as a single contact. For a low contact stress-sliding situation, most engineering contacts of interest are of this type, asperity interaction results in numerous high, transient temperature flashes of as high as several hundred degrees Celsius over a very small area within a few nanoseconds to a few microseconds. As the sliding continues, these temperature flashes shift from one place to another. Measurement of temperature rise over isolated

micro contacts is very difficult. A number of techniques have been used to measure the transient temperature rise in a sliding contact, but with limited success.

Techniques for interface temperature measurement include thermocouples, thin film sensors, radiation detector, metallographic observation, and liquid crystals. Thermocouples are most commonly used but do not measure true flash temperature. Radiation detectors come close to true measurement while liquid crystals and metallographic observations provide a rough estimate of flash temperature.

32.5.2 Thermocouples

A thermocouple uses wires of two dissimilar metals connected together at the two ends and gives rise to a thermal electromotive force (EMF) which depends on the difference in the temperature between two junctions, while one junction is kept at a known reference temperature (cold), the temperature of the other measuring junction (hot) can be obtained from measured EMF using a calibration chart. Two types of thermocouples are in use; embedded thermocouples and dynamic thermocouples. In embedded type, a small hole is drilled through the stationary component of a sliding pair and a small thermocouple is held within the hole such that its measuring junction rests just beneath the sliding surface. In dynamic variety, a thermocouple junction is created at the sliding interface of the contacting bodies. Schematics of the both the types are shown in Figure 32.11. While embedded type cannot measure true flash temperature because of its position and finite mass of measuring junction, the dynamic variety gives a good measure of an average surface temperature rise. Dynamic thermocouples are found to give higher values of measured temperatures and faster transient response than embedded thermocouples. The major disadvantage with dynamic type is that it can only be used for metallic pairs of dissimilar metals and requires electrical contact with a moving body.

32.5.3 Thin Film Sensors

Thin film temperature sensors are used on surfaces to measure temperature rise on a small region and on a very small mass. One of the first of its kind was the thermistor used to measure surface temperature on gear teeth. The resistance of titanium being sensitive to temperature and pressure, measurement of any change in resistance indicates the change in temperature and pressure. Magneto resistive (MR) sensors are used to measure interface flash temperature at magnetic head–medium interfaces. Thin film thermocouples produced by vapour deposition techniques are also used for measurement of sliding interface temperatures.

32.5.4 Radiation Detectors

All materials emit thermal radiation that depends on the surface temperature and structure of the material. It is concentrated in the infrared region if the temperature of the material lies between 10 and 5000 K. Knowing the radiation characteristics of the material and radiant heat transfer properties of particular geometry configuration, temperature of the material may be determined. This method has been successfully used to measure the transient flash temperatures. However, it requires one of the bodies to be transparent to the radiation to be detected and sapphire is a useful material

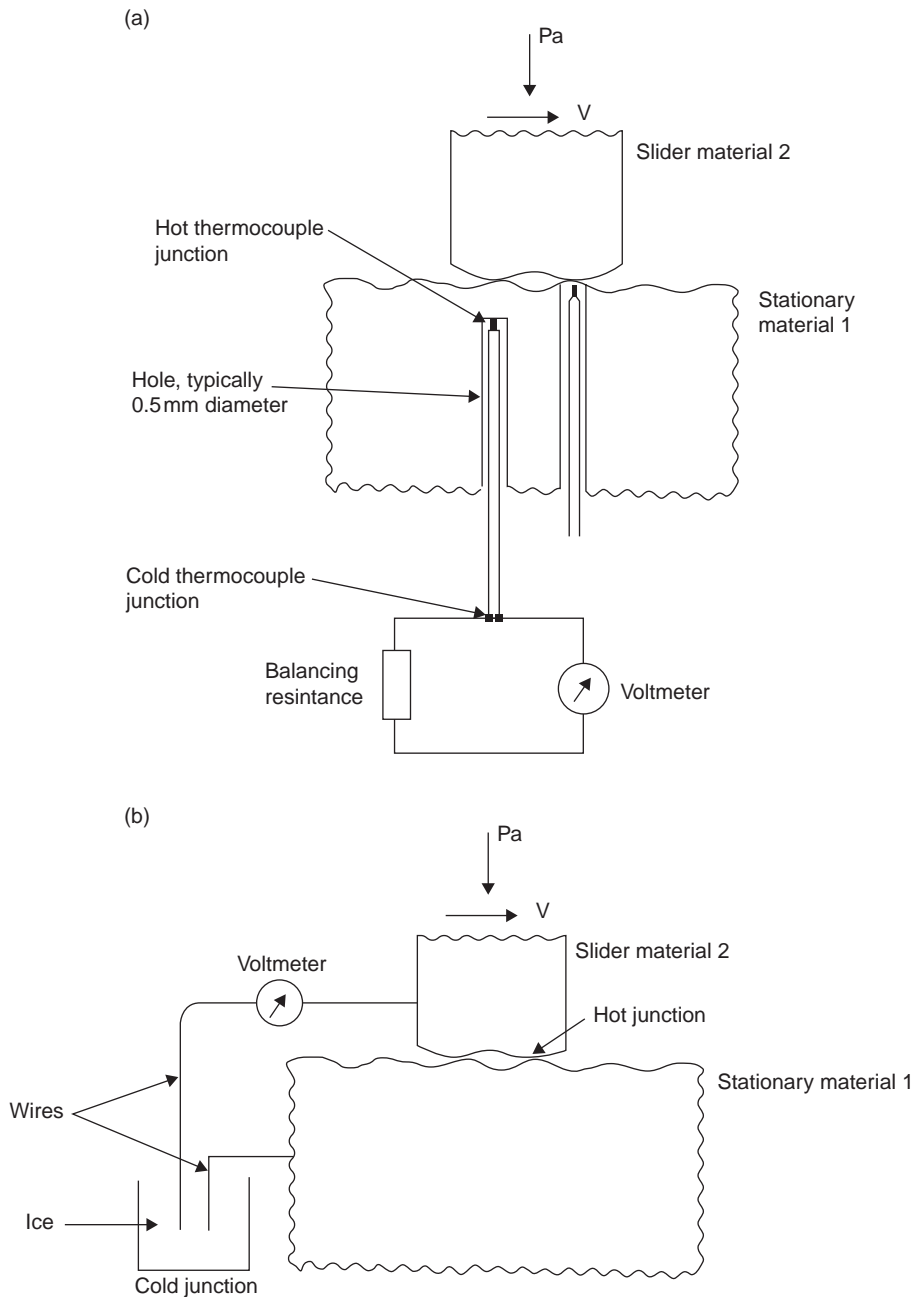


FIGURE 32.11 Schematics of thermocouples: (a) embedded type and (b) dynamic type.

in this respect. Different radiation measurement techniques have been successfully used to measure surface temperatures. These methods include photography, pyrometry, photon detection, and thermal imaging. In infrared detection technique, an infrared (IR) radiometric microscope is used for local surface temperature measurements

where the detector is equipped with optics to limit the field of view to a small spot size in order to permit a small spatial resolution. In photon detection technique, a photo multiplier is used to collect photons emitted by a hot contact spot. The response time of the photo multiplier being very small (less than 30 ns), the technique can be used for detecting flash temperatures of very short duration (2 μ s or less). In thermal imaging technique, a charged couple device (CCD) camera with an IR pass filter is used. The infrared radiation emitted at the contact surface is detected through the sapphire material by the CCD camera, which is linked to a computer-based image processing system. The IR-pass filter permits only IR radiation to reach the CCD camera. The video signal is then processed for thermal imaging and the contact temperature rise may be obtained by suitable transformation through calibration.

32.5.5 Metallographic Observation

Microstructure of many materials undergoes changes due to surface and near surface frictional heating. Using scanning electron microscopic examination of the cross-section of the sliding body in a perpendicular plane to sliding direction, these changes can be detected and may give a rough estimate of the temperature rise. For some material, this can be done by micro hardness measurements. In metal cutting, the flash temperatures during contact can be determined from the micro structural changes in the metal chips, but with limited accuracy. The major limitation of the method is that the micro structural changes may occur due to plastic deformation in surface or near surface regions, in addition to being due to temperature rise.

32.5.6 Liquid Crystals

Cholesteric liquid crystals are also used in surface temperature measurements because these crystals undergo changes in colour with very small change in temperature. When a surface of a body is coated with a specific liquid crystal and subjected to body temperature changes, the colour change occurring in the liquid crystal material estimates the temperature change. Liquid crystals can be sprayed or applied by brush. These are used to measure bulk temperature of bodies rather than flash temperatures.

32.5.7 Humidity Measurement

The routine measurement of humidity and its control at a constant level is necessary as it has a strong influence on friction and wear. Conventionally wet and dry bulb thermometers are used for humidity measurement during tribological studies.

32.5.8 Measurement of Oxygen and Other Gases

The variation in concentration of oxygen and other gases in the atmosphere affect the friction and wear processes to a large extent. Thus there is a need to observe their concentration during tribological tests for effective conclusion of the tests. For this purpose, gas chromatography is commonly used. Gas chromatography is a technique for separating different components of a mixture for analysis by diffusing a gaseous or vaporized sample through a column of liquid. This is based on the principle that different substances are transported at different speeds through a medium. A carrier gas, usually

nitrogen, hydrogen or helium transports the sample through the chromatograph. The mixture of sample and carrier gas is transferred to a long capillary tube filled with liquid, usually a hydrocarbon fluid like silicone oil or squalene, which acts as the separating medium. The components of the sample pass through the capillary tubes at different speeds and are then detected on leaving the capillary to obtain the composition of the sample. Oxygen analysis electrodes are readily available for direct measurement of oxygen concentration in water.

32.6 MEASUREMENT OF MATERIAL CHARACTERISTICS

Tribological behaviour of any material depends on its material characteristics. Each basic type of material, metal, polymer or ceramic, requires different specification of material properties. For metals, the chemical composition and metallurgical phases are significant. For polymers, the organic components, additives and degree of crystallinity are essential. For ceramics, apart from composition, grain size and material toughness are usually important parameters. Accurate measurements of hardness and microhardness are essential for analysis of surface contact and deformation in friction and wear. The materials are in general characterized using four quantities: (a) hardness, that describes the resistance to plastic deformation, (b) Young's modulus and the elasticity limit, that describe the materials' elastic properties, (c) toughness, that accounts for the relative brittleness of the material, and (d) residual stresses, that play an important role in the resistance to wear and cracking.

32.6.1 Hardness

Hardness tests use indenters to make an impression on a sample of a material under a normal load (F) and measure the surface area (S) of the residual impression left after removal of the load. Then hardness is expressed as F/S . Indenters can be of various shapes. For the Vickers test, a square-based pyramid is used while for the Brinell test, spherical indenters are used. For the Rockwell test, either conical or spherical indenters are used while Shore hardness tests use two types of cones.

In Vickers test, the indenter used is a diamond square-based pyramid with an angle of 136° between opposite faces as shown in Figure 32.12.

Vickers hardness is given by the formula:

$$HV = 1.854 \frac{F}{D^2}$$

where F is the applied load (kg) and D is the diagonal of the residual indentation (mm).

Brinell hardness is determined using tungsten carbide spherical indenters of 10, 5, 2.5 or 1 mm in diameter (shown in Figure 32.13). Brinell hardness can be expressed as:

$$HB = \frac{0.204F}{\pi D \left(D - \sqrt{D^2 - d^2} \right)}$$

where F is in N and both the diameter of the spherical indenter (D) and the diameter of the residual impression (d) are in millimetres.

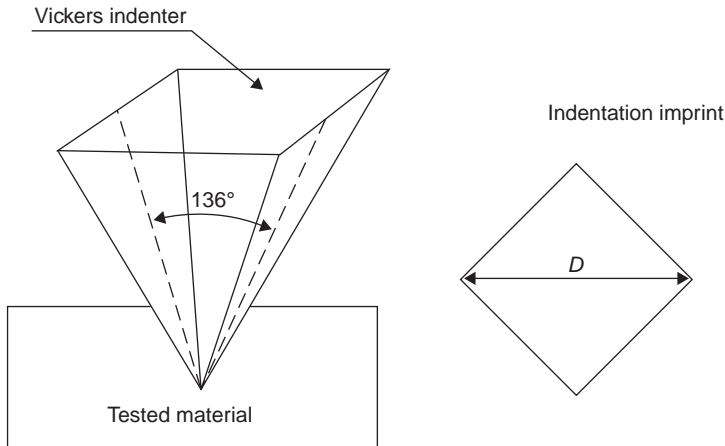


FIGURE 32.12 Vickers hardness test: (a) Vickers indenter and (b) residual square indentation of diagonal D .

In Rockwell hardness test, the depth of the residual impression, h , is measured instead of its diameter. First a preliminary (minor) load of 98 N is applied and it forces the indenter into the surface to a depth of l (say). This depth of penetration is not taken into account in hardness measurement but used as the zero penetration reference point. Then a major normalized load is applied to the surface for a few seconds leading to a total depth of penetration t (say). The major load is then reduced to the minor load and the depth of the residual impression h is measured in millimeters. In Rockwell B test, a steel spherical indenter of diameter 1.58 mm is used with a normal load of 980 N. In Rockwell C test, the load is 1470 N and the indenter is a conical diamond indenter with radius of curvature of 0.2 mm and apex angle of 120° . Rockwell B and Rockwell C hardness can be expressed as:

$$\text{HRB} = 130 - \frac{h}{0.002}$$

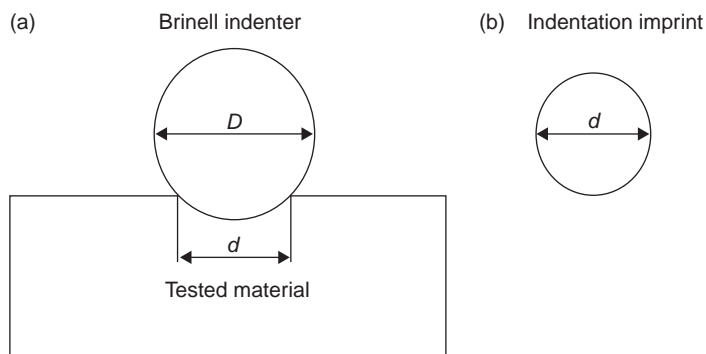


FIGURE 32.13 Brinell hardness test: (a) indenter, and (b) residual impression.

and

$$\text{HRC} = 100 - \frac{h}{0.002}$$

The Shore hardness test is suitable specifically for polymers and is conducted with an apparatus called a durometer and a calibrated spring that generates a known force on the indenter. The penetration depth is recorded and the corresponding value of Shore hardness is obtained in a scale of 0–100 with these values corresponding to maximum penetration and zero penetration respectively. For Shore A test, a truncated cone with an apex angle of 35° is used while for Shore D test, a sharp cone with an apex angle of 30° is used.

32.6.2 Young's Modulus and the Elasticity Limit

For an isotropic material, Young's modulus is the constant of proportionality between the stress applied to a specimen rod and its elastic deformation. Generally uniaxial tensile tests are carried out to determine Young's modulus as well as the elasticity limit. However, the same can be determined from indentation tests also.

32.6.3 Fracture Toughness

Different modes of fracture can occur within a brittle material. Brittleness is characterized using a parameter known as fracture toughness (denoted K_{IC}), that describes the resistance of the material to fracture propagation. K_{IC} is generally determined from impact tests on a single-edge notch bend specimen. It can also be obtained using Vickers indentation tests by measuring the length of the radial crack during indentation.

32.6.4 Residual Stresses

Machining methods and surface production processes incorporate strains on the surface layers of materials that in turn induce internal residual stresses. These may be either tensile or compressive. Tensile stresses are dangerous as they help fracture propagation while compressive stresses are beneficial and often added deliberately to harden the surface of a material in order to increase its resistance to wear. Measurement of residual stresses is done either by destructive techniques or non-destructive techniques. The destructive method relies on the cutting-away of some stressed material and measuring of the resulting deformation in the adjacent material that has occurred due to the relaxation of the residual stresses. Non-destructive method relies on the analysis of the changes of the physical and crystallographic properties of the material due to residual stresses. These include the ultrasound method that uses the effect of residual stresses on the variation of speed of ultrasound waves, the Barkhausen technique that uses the interaction between elastic strain and magnetization in ferromagnetic materials, method of X-ray diffraction and so on.

32.6.5 Chemical Composition of a Surface

Chemical composition of a surface can be determined using a number of methods. Generally, a primary particle beam (electronic, optical or ionic) is directed at the surface under investigation and it produces the emission of secondary particles (photons, electrons or

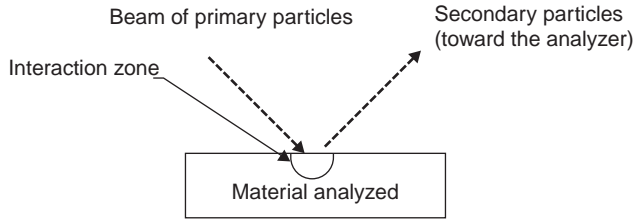


FIGURE 32.14 Principle of spectroscopic surface analysis.

ions). The characterization of these secondary particles allows the identification of the composition of the surface under study (Figure 32.14). The most widely used techniques include Energy dispersive X-ray analysis (EDX), X-ray photoelectron spectroscopy (XPS), Auger electron spectroscopy (AES), Glow discharge optical emission spectroscopy (GDOS), Rutherford backscattering spectroscopy (RBS), Secondary ion mass spectroscopy (SIMS), Infrared (IR) spectroscopy. In EDX, primary particle is electron and secondary particle is photon. EDX helps in elemental analysis of materials to a depth of a few microns but this is not suitable for light element analysis. EDX is generally used in analysis of precipitation and segregation in metallurgical materials, characterization of thermal oxide films, hard coatings. In XPS, primary particle is photon and secondary particle is electron. It helps in elemental analysis of materials along with the information of the environment of the atoms and nature of chemical bonds. In AES, both primary and secondary particles are electrons. It helps in elemental analysis of the top few-atom layers of a solid. It is generally used for analysis of adsorbed layers, thermal or anodic oxide films, thin deposits and so on. In GDOS, primary particle is ion and secondary particle is photon. It can provide simultaneous elemental analysis of dozens of elements and in-depth sample analysis over large areas. In RBS, both primary and secondary particles are ions. In this technique, the analysis of the energy of backscattered helium ions after elastic collision with the sample surface atoms allows the determination of their mass. In SIMS, both primary and secondary particles are ions. It provides elemental and isotropic analysis of all elements in the periodic table. It can be used to obtain ionic images providing elemental distribution. In IR spectroscopy, both primary and secondary particles are photons. This is primarily used for the analysis of organic materials and allows the determination of functional groups present in the material and the nature of the different bonds.

32.7 MEASUREMENT OF LUBRICANT CHARACTERISTICS

Characterization of a lubricant is an essential part of any study related to lubricated wear and friction. Lubricant characterization requires both chemical and physical parameters. Hence different techniques of rheometry and chemical analysis are used for this purpose. Significance of lubricant parameters varies widely depending on the tribological function for which the same is used. For hydrodynamic and elasto-hydrodynamic lubrication, the parameters of significance are viscosity, Barus pressure–viscosity coefficient, temperature dependence of viscosity, limiting shear stress, glass transition, compressibility, thermal conductivity and so on. For boundary lubricants, the significant parameters are composition and concentration of surfactants and corrosive compounds, temperatures of lubricant failure, corrosivity, solubility, and diffusivity of dissolved gases. Sample of a used

lubricant can provide much information regarding the prevailing wear taking place within the tribological system or the quality of lubrication. Also the oxidation or chemical decomposition of a lubricant can easily be detected from the actual lubricant sample.

32.7.1 Analysis of Chemical Changes

Chemical changes occurring in lubricants are directly analyzed using Infra-Red Spectroscopy (IR Spectra). The only difficulty in this technique is the interpretation of complex spectra obtained from degraded lubricants. Specialized numerical analysis of IR spectra is done to extract proper analytical information from overlapping absorption peaks. The most popular one used to detect small changes in the IR spectra is Fourier Transform IR (FTIR) Spectroscopy. Typical IR spectra are displayed as a standard graph of IR transmittance or absorbance within the sample against wave number of the IR radiation.

32.7.2 Viscosity Measurement

The viscosity of lubricants can be measured by many methods based on different principles. Most commonly used instruments typically fall into four categories depending on geometry: Capillary, rotational, and falling sphere and efflux viscometers.

The capillary viscometer is based on measuring the rate at which a fluid flows through a small diameter tube and this takes the form of measuring the time taken to discharge a given quantity of liquid for an applied pressure difference and tube dimensions. In capillary viscometers a direct measurement of absolute viscosity is not possible and kinematic viscosity is measured. The kinematic viscosity is given as: $\nu = ct$ where c is a constant for a given viscometer and t is the time required for a given volume of liquid to flow through the tube. By using some liquid of known viscosity and density, the time required for a given volume to flow through the capillary is determined and the constant for the instrument is found. The sketch of one such viscometer is shown in Figure 32.15.

In rotational viscometers the absolute viscosity of oil is measured. Three different types of rotational viscometers are used: cylindrical, cone-plate type, and parallel plate type. The cylindrical type is in the form of two concentric cylinders of which one rotates in the oil whose viscosity is to be measured. The viscosity measurements are made either by applying a fixed torque and measuring the speed of rotation, or by driving the rotating element at a constant speed and measuring the torque required. The cone-plate and parallel-plate rotational viscometers are the variations of the same technique. For viscosity measurements of non-Newtonian fluids, rotational viscometers involving shearing of the fluid are used. These instruments are usually calibrated with fluids of known viscosity and the viscosities of test samples are taken from the calibration chart. Basic arrangement of cylindrical and cone-plate rotational viscometers are shown in Figure 32.15.

In falling sphere viscometer, the time taken for a ball to fall through a measured height of fluid in a tube is measured. The time required is a measure of absolute viscosity. Other variations of this type are also available. In rolling-sphere viscometer, a ball rolls down an inclined tube filled with test fluid. The falling co-axial cylinder viscometer consists of two coaxial cylinders with their axes vertical. The outer cylinder is clamped and the clearance space between the cylinders is filled with the test fluid. In operation, the inner cylinder is released and falls under gravity. The viscosity

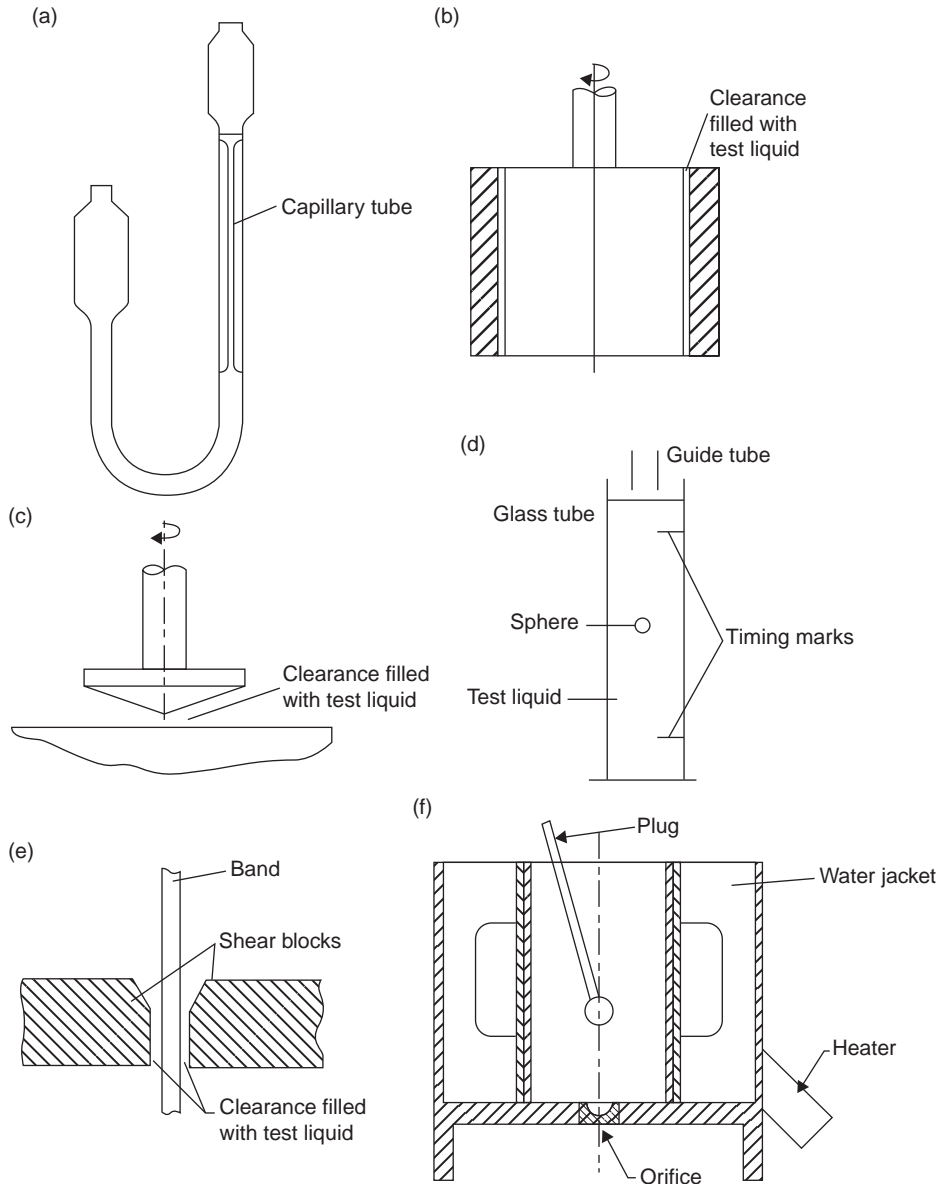


FIGURE 32.15 Different viscometers: (a) u-tube viscometer, (b) rotating-cylinder viscometer, (c) cone and plate viscometer, (d) falling-sphere viscometer, (e) band viscometer, and (f) efflux viscometer.

is determined from the speed of descent of the cylinder whose motion is opposed by the shear stresses induced in the test fluid. Another variety is the band viscometer where a thin band is located in a narrow gap between two parallel shear blocks filled with the test fluid. The band is released to fall under gravity, its motion being opposed by the shear stresses induced in the test fluid. The viscosity is determined from the

speed of descent of band. These types of instruments are useful for non-Newtonian measurements and for high viscosity fluids.

In efflux viscometers the viscosity is measured by the time taken for a given volume of liquid to discharge under gravity through a short tube orifice placed in the base of the instrument. Three types are common: Redwood, Saybolt, and Engler. The first one is commonly used in the U.K. while the second one is mainly used in the Europe and U.S.A. In all these three instruments, the viscosities are quoted in terms of the efflux time, for example, Redwood sec, Saybolt sec or Engler sec. These measurements are then converted to kinematic viscosity using conversion tables.

32.7.3 Lubricant Oxidation Tests

Hydrocarbon lubricants oxidize readily as a result of low temperature acidification. Thus it is important to know the remaining useful life (RUL) of lubricant before the uncontrolled oxidation of the base oil occurs. This is done by conducting laboratory oxidation stability tests where the lubricant is oxidised under various conditions of temperature, pressure, and catalysts. Then lubricant characteristics such as viscosity, antioxidant depletion, total acid number, sludge content and so on are evaluated to measure oil degradation and oxidation stability, and finally to determine the remaining useful life of the lubricant. Lubricant oxidation stability tests can be grouped into three categories: bulk oxidation tests, micro-scale oxidation tests, and non-standard tests. The bulk oxidation test is done to see if excessive oil oxidation takes place when lubricating oil is stored in a storage tank. These methods include open vessel tests, bomb oxidation tests and so on. Micro-scale oxidation tests involve small amounts of lubricant and are carried out to accurately simulate oxidation of oil when it is present as a thin layer so that all parts of the oil are in close contact with atmospheric oxygen. These methods include thin layer oxidation tests, gel permeation chromatography and so on. Non-standard tests are done to reduce the test duration in comparison to standard tests. These methods include differential scanning calorimetry, differential thermal analysis, thermogravimetric balance, chemiluminescence tests and so on.

32.8 WEAR PARTICLE ANALYSIS

Wear particles are produced during almost every tribological operation and are an important source of information about wear processes. The composition, size, and morphology of wear particles are indicative of the wear mechanism involved in their formation. Wear particle analysis helps to assess the condition of the machinery and reliability in mechanical systems. Two basic types of analysis are widely used. One is based on the analysis of the chemical composition of the lubricant to reveal the presence and quantity of metallic wear particles in hydrocarbon based lubricating oil. The other method involves an extraction of wear particles from a used lubricant and then assessment of their number, size, shape, and morphology.

32.8.1 Chemical Analysis of Particles in Lubricant

The common method used to analyze the chemical composition of particles present in lubricating oil is flame spectroscopy. This is based on the principle of emission or

absorption of characteristic colour by individual chemical element when heated in a flame or arc. Systematic analysis of the colours so produced by a lubricant sample provides a complete data of the chemical elements present therein. The concentration of different elements can also be assessed. The technique of flame spectroscopy is commonly known as Spectroscopic Oil Analysis Program (SOAP) which is used in machine condition monitoring. SOAP is used to specify whether to continue operating a machine or to go for any maintenance based on the concentration of contaminant metals and water in the lubricating oil beyond acceptable limits. In tribological experiments on lubricated wear, SOAP is used to determine if oil is functioning at desired conditions. SOAP provides the information about very small wear particles that remain suspended in the oil but fails to capture large wear particles. Thus as the size of the wear particles increases the SOAP accuracy decreases and it becomes insensitive to particles larger than 10 μm .

32.8.2 Analysis Based on Separation of Wear Particles

To avoid the limitations of SOAP analysis, a technique called 'Ferrography' is developed. In this method, wear particles are separated from the lubricating oil according to their size based on the application of a strong magnetic field. A sample of lubricating oil or oil diluted by solvent is allowed to flow over an inclined glass slide suspended above a magnetic field. The fluid sample flows by gravity along the slide. The combination of viscous and magnetic forces acting on the wear particles helps in separation and settlement of the particles on the glass slide. Large particles of iron and steel get deposited near the entry region due to strong attraction of the magnetic field and fine particles are deposited away from the entry region on the slide. Then the slide containing the thin streak of wear particles is subjected to either analytical ferrography or direct reading ferrography. In the former method, the slide is examined under a microscope to study the morphology of wear particles. In the other method, the change in transmission of light through wear particles deposited in a glass tube is measured.

The shape and surface morphology of wear particles provide the information of how the wear particles are formed. Wear particle morphology is in general described by two sets of parameters: particle boundary descriptors like aspect ratio, shape factor, convexity, curl, roundness, fractal dimension and so on and surface topography descriptors like perspective view, colour stereo, contour map image and so on.

32.9 INDUSTRIAL MEASUREMENTS

Measurements in an industrial context are performed to confirm the performance of a component or system. This is usually different from scientific tribological measurement where the basic objective is to develop a model of friction, wear or related phenomena. Industrial measurements are primarily empirical and differ from scientific investigation in respect to specification. Industrial testing usually relates to commercial products, where the test specification is either confidential or incompletely specified. For example, oil additives are marketed without publicly available specification. In many cases, complexity of the phenomena occurring inside the system hinders direct tribological investigation in a simple tribometer. The interaction

between the lubricating oil and combustion cycles of an internal combustion engine is extremely difficult to simulate with the help of a tribometer. Thus direct testing of the engines and gearboxes is the only means to evaluate the performance of the system. Similarly metalworking involves many frictional situations that cannot be simulated in a tribometer and the same requires special tests that are significantly different from common tribological testing with respect to the controlling factors. In metalworking the significant parameters are surface finish, precision, tool life, power consumption and so on. The complexity of service conditions inside a synovial joint in orthopaedic implants is always a challenge to tribologists and designers of tribometers. All these fields of testing ranging from lubricants to orthopaedic implants are known as 'industrial tribology' where the major concern is to overcome the difficulties in identifying the controlling parameters to have a comprehensive understanding of the system.

32.10 SUMMARY

Tribologists for generations have faced a unique problem; predictions and designs are based on continuum and scale independent phenomenon. Though tribological interactions often occur in very small scales and volumes, the observations are related to large-scale dimensions and volumes. As an example, Bowden and Tabor's junction growth model may be cited. It is a very small volume phenomenon at the asperity level that involves molecular level interactions. But due to unavailability of instrumentation that makes it possible to observe directly the phenomenon and the related small volume deformations, the model was based on continuum plasticity theory to provide a phenomenological view of adhesive wear and friction. The situation has undergone a dramatic change in last two decades. With a virtual revolution in materials technology and electronics; reliable measurements down to nano- and Pico metric levels are now possible. This advancement coupled with recent advances in non-continuum modelling at the molecular level has opened up a new horizon for the tribologists and scientists. These advances have opened up the new field of nanotribology. It involves experimental and theoretical investigations of interfacial processes on scales ranging from the atomic/molecular to micro-scale that occur during adhesion, friction, wear, indentation, and thin film lubrication at sliding surfaces.

33

CORROSION MONITORING

PIERRE R. ROBERGE

- 33.1 What is corrosion monitoring?
- 33.2 The role of corrosion monitoring
- 33.3 Corrosion monitoring system considerations
 - 33.3.1 What is the monitoring objective?
 - 33.3.2 Corrosion monitoring locations
 - 33.3.3 Probe design and selection
 - 33.3.4 Corrosion monitoring techniques
 - 33.3.5 Monitoring microbiologically influenced corrosion
 - 33.3.6 Monitoring pipeline CP systems
 - 33.3.7 Atmospheric corrosion monitoring

References

33.1 WHAT IS CORROSION MONITORING?

Corrosion monitoring refers to corrosion measurements performed under industrial or practical operating conditions. In its simplest form, corrosion monitoring may be described as acquiring data on the rate of material degradation. However, such data are generally of limited use and need to be converted into useful information for inclusion in a corrosion management program. This requirement has led to the evolution of corrosion monitoring tools toward real-time data acquisition, process control tools, knowledge-based systems, and smart structures. Corrosion monitoring is more complex than the monitoring of most other process parameters because

- There are a number of different types of corrosion (see Chapter 2 for more details).
- Corrosion may be uniform over an area or concentrated in very small areas (pitting).

- General corrosion rates may vary substantially, even over relatively short distances.
- There is no single measurement technique that will detect all of these various conditions.

Before embarking in a corrosion monitoring program, it is therefore helpful to review historical data and consider the types of corrosion problems that need to be investigated. It is also advisable to use several complementary techniques rather than rely on a single monitoring method.

Real-time monitoring of pipelines, vessels, and other static equipment enables a near instantaneous appraisal of the corrosivity of produced and transported fluids. If the corrosion activity increases as a result of process nonconformities, the corrosion information can be viewed alongside process variables such that cause-and-effect can be determined and rapid action can be taken to stifle the progress of any associated problem. The same approach can be used to demonstrate the effectiveness of remedial actions or preventive treatments.

Some modern corrosion monitoring technologies are particularly apt at revealing the highly time-dependent nature of corrosion processes. The integration of these corrosion monitoring technologies in existing systems can thus provide early warnings of costly corrosion damage.

An extensive range of corrosion monitoring techniques and systems has evolved, particularly in the last two decades, for detecting, measuring, and predicting corrosion damage. The development of efficient corrosion monitoring techniques and user-friendly software have permitted us to field new techniques that were until recently perceived to be mere laboratory curiosities. Noteworthy catalysts to the growth of the corrosion monitoring market has been the expansion of oil and gas production under extremely challenging operational conditions (e.g., the North Sea), cost pressures brought about by global competition, and the public demand for higher safety standards. In several sectors, such as oil and gas production, sophisticated corrosion monitoring systems have gained successful track records and credibility, while in other sectors their application has made only a limited progress.

33.2 THE ROLE OF CORROSION MONITORING

Correct and effective corrosion monitoring strategies should be used as a proactive tool to assist with operating a plant or any other system more effectively, thereby prolonging its life and gaining optimum throughput. Fundamentally, four strategies are available to an organization in its dealings with corrosion:

1. Ignoring it until a failure occurs.
2. Inspection, repairs, and maintenance at scheduled intervals.
3. Using corrosion prevention systems (inhibitors, coatings, resistant materials etc.).
4. Applying corrosion control selectively, when and where it is actually needed.

The first strategy represents corrective maintenance practices, whereby repairs and component replacement are only initiated after a failure has occurred. Corrosion monitoring is completely ignored in this reactive philosophy. Obviously, this practice is unsuitable for safety critical systems and in general is inefficient from maintenance cost considerations, especially during life extension of aging engineering systems.

The second strategy is one of preventive maintenance. The inspection and maintenance intervals and methodologies are designed to prevent corrosion failures, while achieving 'reasonable' system usage. Corrosion monitoring can assist in optimizing these maintenance and inspection schedules. In the absence of information from a corrosion monitoring program, such schedules may be set too conservatively with excessive downtime and associated cost penalties.

Alternatively, if set too infrequently, the inspection and maintenance intervals may represent an excessive corrosion risk with associated cost and safety consequences (cf. Figure 3.10). Furthermore, without input from corrosion monitoring information, preventive inspection and maintenance intervals will be of the "routine" variety, without accounting for the time dependence of critical corrosion variables. In the oil and gas industry, for example, the corrosivity at a well-head can fluctuate significantly over the lifetime of the production system, between benign to highly corrosive. In oil refining plants, the corrosivity can vary with time, depending on the grade or hydrogen sulfide content of crude that is processed.

The application of corrosion prevention systems is obviously crucial in most corrosion control programs. However, without corrosion monitoring information, their application may be excessive and costly. For example, a certain inhibitor dosage level on a particular system may combat corrosion damage, but real-time corrosion monitoring might reveal that a lower dosage would actually suffice. Ideally, the inhibitor feed rate should be continuously adjusted based on real-time corrosion monitoring information.

In an ideal corrosion control program, inspection, and maintenance would be applied only where and when they are actually needed. In principle, the information obtained from corrosion monitoring systems can be of great assistance in reaching this goal. However, it is sometimes difficult for a corrosion engineer to get management's commitment to investing funds in such initiatives. The importance of corrosion monitoring in industrial plants and other engineering systems should be presented as an investment to achieve some of the following goals:

- Improved safety.
- Reduced downtime.
- Production of early warnings before costly serious damage sets in.
- Reduced maintenance costs.
- Reduced pollution and contamination risks.
- Longer intervals between scheduled maintenance.
- Reduced operating costs.
- Life extension.

Experience has shown that the potential cost savings resulting from the implementation of corrosion monitoring programs generally increase with the sophistication level (and cost) of the monitoring system.

33.3 CORROSION MONITORING SYSTEM CONSIDERATIONS

Corrosion monitoring systems vary significantly in complexity, from simple coupon exposures or handheld data loggers (Figure 33.1) to fully integrated plant process surveillance units with remote data access and data management capabilities.

Corrosion sensors (probes) are an essential element of all corrosion monitoring systems. The nature of the sensors depends on the various individual techniques used for monitoring, but often a corrosion sensor can be viewed as an instrumented coupon. A single high-pressure access fitting for insertion of a retrievable corrosion probe can be used to accommodate most types of retrievable probes. With specialized tools (and brave specialist operating crews!) sensor insertion and withdrawal can be possible under pressurized operating conditions.

The signal emanating from a corrosion sensor usually has to be processed and analyzed. Examples of signal processing include filtering, averaging, and unit conversions. Furthermore, in some corrosion sensing techniques, the sensor surface has to be perturbed by an input signal to generate a corrosion signal output. In older systems, electronic sensor leads were usually employed for these purposes and to relay the sensor signals to a signal-processing unit. Advances in microelectronics have facilitated the sensor signal conditioning and processing by the introduction of microchips that have become an integral component of sensor units (Zollars et al., 1997; Kelly, 1997). Wireless data communication with such sensing units is also a product of the microelectronic revolution. Figure 33.2a and b illustrates a recent wireless monitoring system developed to provide an early warning that a protective coating is degrading (Davis et al., 2005). Such a direct



FIGURE 33.1 Field corrosion monitoring using electrochemical noise recorded with a handheld data logger. (Courtesy of Kingston Technical Software.)

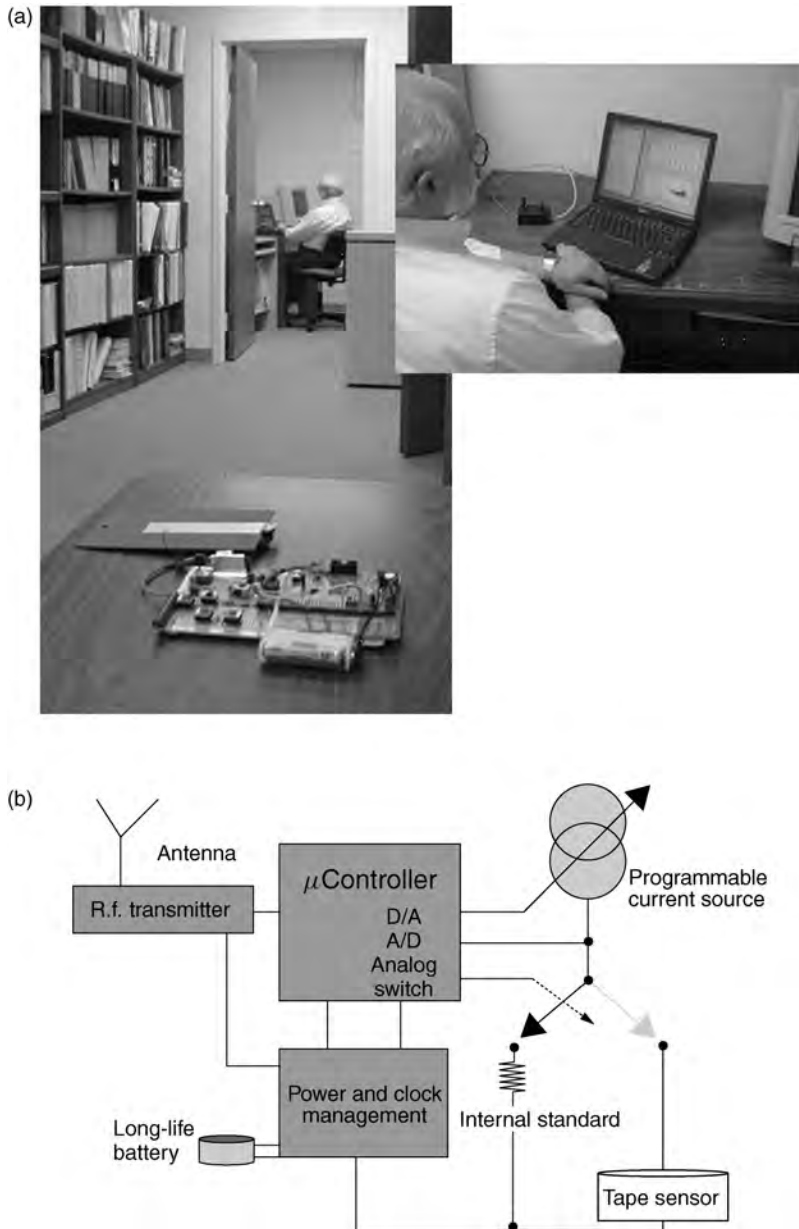


FIGURE 33.2 (a) Test and demonstration of breadboard coating health monitor, including wire-less communication. Inset shows transceiver unit attached to laptop computer. (b) Block diagram of the electronics. (Courtesy of Guy D. Davis, DACCO SCI, Inc.)

measurement contrasts with corrosivity sensors that monitor indirectly the problem by monitoring ambient environments.

Irrespective of the sensor details, a data acquisition system is required for on-line and real-time corrosion monitoring. On several plants, the data acquisition system is housed in mobile laboratories, which can be made intrinsically safe. A computer system often performs a combined role of data acquisition, data processing and information management system. In data processing, a process is initiated to transform corrosion monitoring data (low intrinsic value) into process relevant information (higher intrinsic value). Complementary data from other relevant sources, such as process parameter logging and inspection reports, can be acquired together with the data from corrosion sensors, for use as input to a management information system.

Numerous real-time corrosion monitoring programs in diverse branches of industry have revealed that the severity of corrosion damage is rarely uniform with time. Rather, serious corrosion damage is usually sustained in time frames where operational parameters have suffered upsets. These undesirable operational windows can only be identified with the real-time monitoring approach.

In general, it can be said that no individual technique alone is suitable for monitoring corrosion under complex industrial conditions. Therefore, a multitechnique approach is often preferred. In many cases, this approach does not require a higher number of sensors, but rather only an increased number of sensor elements for a given probe and access fitting.

Another important consideration is that, irrespective of the technique, most instrumented sensors only provide semiquantitative corrosion damage information. It is thus sensible to correlate monitoring data from these sensors to long-term coupon exposure programs and actual plant damage. Unfortunately, nonspecialists may put too much faith in the numerical corrosion rate displayed by a commercial corrosion monitoring device. An example of a widely available technique is linear polarization resistance (LPR), which many commercial monitoring systems use to measure corrosion rates, commonly displayed as millimeters/year or milli-inches/year (mpy).¹ Such systems are used extensively in industry for monitoring the effectiveness of water treatment additives and various other applications. However, from fundamental considerations, the derived LPR corrosion rate is only valid if the following assumptions are met, which is rarely the case in actual operational conditions:

- There is only one simple anodic reaction.
- There is only one simple cathodic reaction.
- The anodic and cathodic Tafel constants are known and invariant with time.
- The corrosion reactions proceed by a simple charge-transfer mechanism under activation control, which essentially implies that the corroding surface is clean without corrosion product build-up, scale deposits, or solids settled out of solution.
- Corrosion proceeds in a uniform manner (whereas the vast majority of industrial corrosion problems are related to localized attack).

¹ Table 2.2 provides the conversion factors between commonly used corrosion rate units for all metals and Table 2.3 describes these conversion factors adapted to iron or steel (Fe) for which $n = 2$, $M = 55.85 \text{ g mol}^{-1}$, and $d = 7.88 \text{ g cm}^{-3}$.

- The solution resistance is negligible (some instruments make a solution resistance compensation but this is not necessarily accurate).
- The corrosion potential has reached a steady state value.

The following sections describe in more specific details the various elements that should be considered when deciding on a corrosion monitoring program.

33.3.1 What Is the Monitoring Objective?

The first and most essential step of a corrosion monitoring program is to define the monitoring objective, a step often forgotten. If corrosion monitoring is done for corrosion control the purpose is to assure that asset life is not jeopardized by too many high corrosion rate events. The main objective of corrosion monitoring is in this case to limit the “corrosion events,” without completely using the corrosion allowance of a system before the end of its design life. The main factors that govern the design of a monitoring system in this case are (Thomas and Terpsta, 2003) available corrosion allowance; uncontrolled corrosion rates; event rates; corrosion rate detection sensitivity and response rate; and required service life.

If corrosion monitoring is used for corrosion control, it is thus essential that the corrosion mechanism and the corrosion rates during times of noncompliance be relatively well known. Corrosion monitoring can also be used in a corrosion control context to establish the corrosion mechanism of a system or optimize the corrosion control, for example, testing the efficiency of corrosion inhibitors, adjusting the corrosion inhibitor injection rate, or studying corrosion mechanisms. Such measurements can either be carried out on instrumented sections of the actual system or in a side stream.

Corrosion monitoring could also be needed in a broader context of integrity management to ensure that the operating envelope of a system is not exceeded. The time horizon of ongoing integrity activities can be much shorter than the plant lifetime, by virtue of the definition of integrity. This can be satisfied by ensuring that integrity is maintained up to the next inspection date and reassessed for another period.

33.3.2 Corrosion Monitoring Locations

An important decision in setting up a corrosion monitoring system is the selection of the monitoring points or sensor locations. As only a finite number of points can be considered for obvious economical reasons, it is usually desirable to monitor the “worst-case” conditions, at points where corrosion damage is expected to be most severe. Often, such locations can be identified by reasoning with basic corrosion principles, analysis of in-service failure records and in consultation with operational personnel. For example, the most corrosive conditions in water tanks are usually found at the water–air interface. Corrosion sensors could be attached to a floating platform to maintain these conditions independently of water level changes in order to monitor corrosion under these conditions.

In any monitoring situation, the ideal probe placement is most often not possible. Invariably, the flush or protruding electrode probes require to be placed in the most corrosive environment (e.g., at a 6 o’clock position in a pipeline, at the bottom of a vessel, or at a solution accumulation point in a separator tower). Almost equally invariably, the

available location is entirely at odds with these requirements. Although certain steps can be taken to provide the application with a modified design (e.g., a protruding electrode probe installed at the 12 o'clock position with a long body to extend to the aqueous phase), this is without question a compromise on the part of the technology provider (Eden et al., 2003).

It is obviously imperative that corrosion sensors be positioned to reflect the state of the actual component or system being monitored. If this requirement is not met, all subsequent signal processing or data analysis is negatively impacted and the value of information greatly diminished or even rendered worthless. For example, if turbulence is induced locally around a protruding corrosion sensor mounted in a pipeline, the sensor will in all likelihood give a very poor indication of the risk of localized corrosion damage to the pipeline wall. In this particular case, a flush mounted sensor should be used instead if the goal is to monitor localized corrosion (Figure 33.3).

In practice, the choice of monitoring points is also dictated by the existence of suitable access points, especially in pressurized systems. It is usually preferable to use existing access points, such as flanges for sensor installations. If it is difficult to install a suitable sensor in a given location, additional by-pass lines with customized sensors and access fittings may represent a practical alternative. One advantage of a by-pass is the

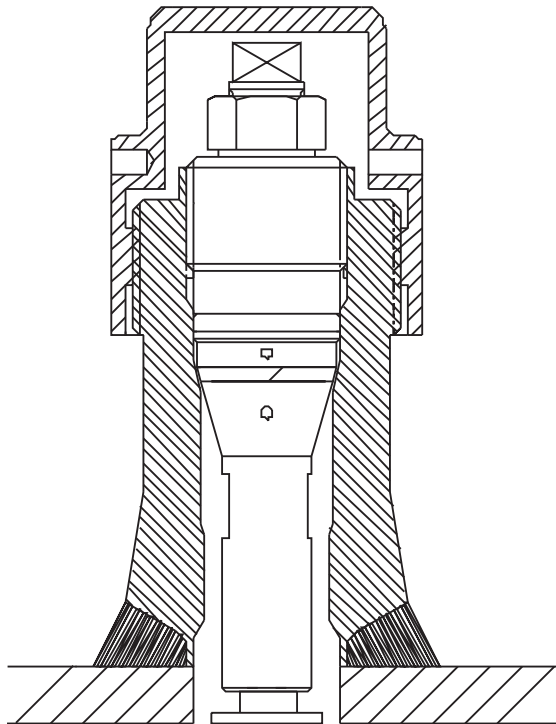


FIGURE 33.3 Flush mounted corrosion sensor in an access fitting. (Courtesy of Metal Samples Company, www.metalsamples.com.)

opportunity of experimenting with local conditions of highly corrosive regimes in a controlled manner, without affecting the actual operating plant.

33.3.2.1 Process Industry The following example illustrates how critical sensor locations have been identified for a distillation column (Dean, 1986). The feed point, overhead product receiver, and bottom product line represent locations of temperature extremes and also points where products with different degrees of volatility concentrate. In many cases however, the highest corrosivity is encountered at an intermediate height in the column where the most corrosive species concentrate. Initially, therefore, several monitoring points would be required in such a column (Figure 33.4). As monitoring progresses and data from these points become available, the number of monitoring points could be optimized further.

33.3.2.2 Oil and Gas Production Gathering Lines Corrosion processes within oil and gas production gathering lines and process piping are usually monitored by the use of metal loss coupons or electrical resistance probes inserted into process fluids through access fittings (Powell et al., 2001). Once exposed to the process fluids for periods ranging from 90 days to 1 year, depending on the system and the corrosivity of the fluids, the coupons are removed and cleaned. A comparison of their initial and final weights is used to determine the general corrosion rate, based on the assumption of uniform corrosion

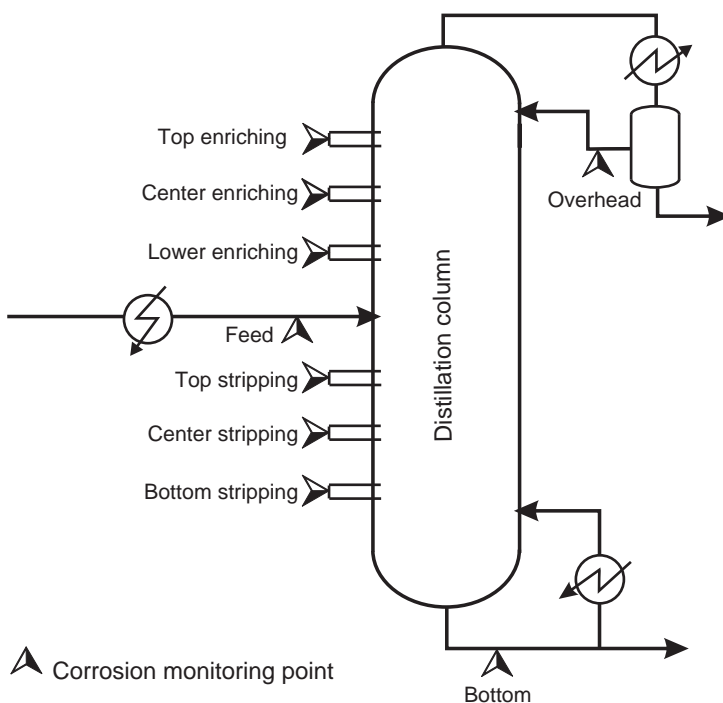


FIGURE 33.4 Corrosion monitoring points in a distillation column.

throughout the exposure period. Corrosion pitting rates can also be determined on the same coupons by measuring the depth of the deepest pit.

In the same environment, electrical resistance (ER) and LPR probes can yield near real-time measurement of the corrosion rate within the process systems. These techniques provide corrosion rate measurements, such that short-term events that affect the rates can be easily identified. For example, the flowback of an acid stimulation treatment can be detected, provided the data readings are sufficiently frequent. Fortunately, remote data collectors and computers greatly reduce the time needed to record and analyze corrosion probe data.

Coupon monitoring and real-time corrosion monitoring techniques complement each other as they focus on different time intervals and should be integral parts of any comprehensive corrosion monitoring system.

The most important consideration when selecting corrosion monitoring locations within crude oil or wet gas production systems is to find locations near the end of the pipeline where the corrosion coupon or probe will be immersed in any produced water. This placement is typically at the 6 o'clock position on horizontal sections of pipeline because produced water is heavier than crude oil or gas condensates. In Figure 33.5a, the monitoring probe installed at 6 o'clock is in an ideal position to sense corrosion processes.

Figure 33.5b illustrates a gas production line that contains a small quantity of condensate and produced water. The water is swept along the bottom of the pipe for horizontal runs. (Only at high fluid velocity is water swept along the circumference of the pipe.) Hence, monitoring locations on the side of the pipeline (3 or 9 o'clock) as shown in Figure 33.5b cannot accurately measure the corrosion rates associated with the aqueous phase at the bottom of the pipeline.

Unfortunately, many pipeline and facility designers have installed monitoring locations on the sides of the pipelines rather than the bottom. Although the side locations may provide easier access for coupon crews, these coupons or probes cannot provide accurate data unless the pipelines are essentially full of water.

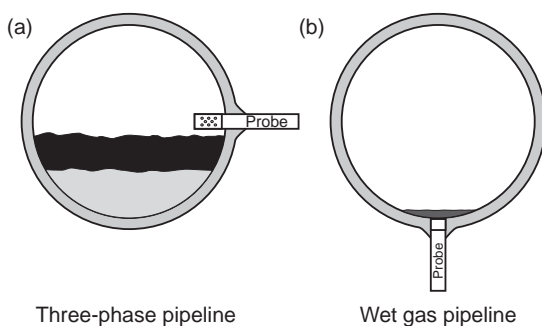


FIGURE 33.5 (a) Probe at 9 o'clock. The intrusive probe is incorrectly positioned and cannot monitor corrosion associated with the water phase. This probe would yield invalid results because it is in the gas or oil phase. (b) Probe at 6 o'clock. The flush-mounted probe is best positioned to monitor corrosion, even when there are small volumes of produced water, as in wet gas or some three-phase pipelines.

Hence, operator convenience must not come at the expense of obtaining valid results when selecting coupon or probe monitoring locations.

It is best to design and install the coupon or probe access fittings during the initial fabrication and installation of pipeline systems rather than retrofitting. Although “hot taps” can be conducted safely, they require additional cost, preparation, and integrity checks.

Another consideration is the quantity of water within the pipeline. This determines whether intrusive or flush mounted coupons or probes are the better choice. For example, if an intrusive probe is installed within a wet gas pipeline, the sensing element extends well above the bottom pipe wall and cannot give valid readings. If, however, the pipeline contains three-phase production with a large water cut, inserting an intrusive probe to position the sensing element within the produced water might work. However, a flush mounted probe is generally the better choice because it is less dependent on the quantity of produced water. For pipelines carrying water, linear polarization probes may be used. If hydrocarbons are present, however, they will coat the sensing element and increase the noise/signal ratio. Thus, electrical resistance probes are recommended whenever hydrocarbons are present.

When natural gas is produced, small quantities of water and condensates may typically be produced. Quantities vary, but volumes are usually 160 L of water and up to 16,000 L of hydrocarbon condensate per 30,000 standard cubic meters. The produced water, being heavier than the gas or condensates, is generally swept along the bottom of the pipelines by the flow of gas. The water may accumulate in dead legs or low spots along the pipelines until the volumes are such that the gas will be able to push the fluids further down the pipelines.

Intrusive probes should be located where they can remain in place for extended periods, rather than having to be removed periodically to support pigging and other routine operations. Thus, an intrusive probe should be installed upstream of any pig launcher and downstream of any pig receiver so as not to block the path of the pig. Where this is not possible, flush-mounted probes should be used.

33.3.3 Probe Design and Selection

The probe design is another important consideration in corrosion monitoring since the probe element interfaces directly with the process environment and must be both suitable for the installation location and enable to make representative corrosion measurements. All too often, the quality and relevance of the corrosion data measured can be severely compromised by inappropriate probe design. In this context, knowledge of the probe surface condition is particularly crucial during the initial design and obviously remains important for the duration of the exposure period.

Other factors to consider relate to surface roughness, residual stresses, corrosion products, surface deposits, preexisting corrosion damage, and temperature that can all have an important influence on corrosion damage and need to be taken into account for making representative probes. By considering these factors, it can be desirable to manufacture corrosion sensors from a precorroded material that has experienced actual operational conditions. Heating and cooling may also be applied to corrosion sensors, using special devices, for their surface conditions to reflect certain plant operating domains. Sensor designs, such as spool pieces in pipes and heat exchanger tubes, flanged sections of

candidate materials or test paddles bolted to agitators, also represent efforts to make these sensors representative of actual operational conditions.

The choice of a specific monitoring probe should also be based on the anticipated corrosion rates within the system, as well as on the required sensitivity. When conducting a short-term corrosion test, probes with high sensitivity are desired. For long-term monitoring, however, a thicker probe element with a longer measurement lifetime may be desired.

For rapid flow conditions or if there are concerns related to suspended solids, the sensing elements should be protected with a velocity shield. An ER probe can also be used to measure an “erosion rate” associated with production of sand or other solids. For this purpose, a noncorroding metal element should be selected (Powell et al., 2001).

33.3.3.1 Flush-Mounted Electrode Design The flush-mounted electrode design is most appropriate for use in applications, such as oil and gas flowlines where pigging operations are necessary. While the design is suited to this application operational need, it greatly limits the electrode exposed surface area and the accuracy of the measurements, particularly in low conductivity environments or with low sensitivity instrumentation. As with most measurement processes there is a trade-off between available area for measurement and the opportunity to actually measure corrosion events of low statistical probability (Eden et al., 2003).

Great care must also be taken in the manufacture of this type of probe as the opportunity exists to artificially create unwanted physical phenomena, such as crevices. A crevice created between the outer circumference of an electrode and the surrounding insulating material could provide a focal point for localized corrosion activity and introduce a significant error in the measured data. This effect would be reduced by using larger surface area electrodes.

While this type of probe can be supplied with electrodes of sufficient material to last the lifetime of the component into which it is installed, the monitoring program undoubtedly will benefit from the ability to replace the probe on a regular basis for visual inspection and to confirm the measured data.

33.3.3.2 Protruding Electrode Design The protruding electrode design has more broad-ranging applications than the flush mounted design. A major benefit of this design is the possibility to use replaceable electrodes, as this provides a cost-effective solution. The possibility of crevice corrosion is also less important as the exposed length of an electrode increases the ratio of exposed surface area to the region where a crevice might occur, that is, the circumference of the electrode. However, this design relies on the complete exposure of the electrode surface to the corrosive environment, and so issues may arise in situations where the flow regime within the monitored system becomes turbulent or if the water cut is reduced significantly during operation (Eden et al., 2003).

33.3.3.3 Probes to Suit the Application Each corrosion monitoring application has its specific needs and requirements. The following sections describe a few probe designs that have been developed for specific degradation mechanisms and environments.

Stress Corrosion Cracking Probe Corrosion probes have been developed that enable the working electrode of a three-electrode arrangement to be prestressed to yield stress in

order to match the operating condition of a pipe or vessel. In the following example, corrosion monitoring and control of the double-shell tanks (DSTs) at Hanford had historically been provided through a waste chemistry sampling and analysis program. In this program, waste tank corrosion was inferred by comparing waste chemistry samples taken periodically from the DSTs with the results from a series of laboratory tests done on tank steels immersed in a wide range of normal and off-normal waste chemistries (Edgemon, 2005).

This method has been effective, but is expensive, time consuming, and does not yield real-time data. The Hanford Site near Richland, Washington, has 177 underground waste tanks that store ~253 million liters of radioactive waste from 50 years of plutonium production. In 1996, the Department of Energy Tanks Focus Area launched an effort to improve Hanford's DST corrosion monitoring strategy and to help address questions concerning the remaining useful life of these tanks. Several new methods of on-line localized corrosion monitoring were evaluated. The electrochemical noise (EN) technique was selected for further study based on numerous reports that showed this technique to be the most appropriate for monitoring and identifying the onset of localized corrosion.

Based on a series of studies, a three-channel prototype field probe was designed, constructed and deployed in August 1996. Following the demonstration of the prototype for approximately a year, a longer more advanced eight-channel system was designed and installed in September 1997. Figure 33.6 shows the installation of this system. Unlike the previous prototype, the in-tank probe on this system reached from tank top to tank bottom exposing two channels of EN electrodes in the sludge at tank bottom, four channels in the tank supernate, and two channels in the tank vapor space. Four additional systems of similar design have been installed into other DSTs.

Like most EN based corrosion monitoring systems, the active Hanford systems monitor EN on channels composed of three nominally identical electrodes immersed in the tank waste. Each system is composed of an in-tank probe and ex-tank data collection hardware. The in-tank probe is fabricated from a ~17-m long piece of 2.5-cm diameter



FIGURE 33.6 Installation of first full-scale probe into a double-shell tank (DST) at the Hanford site. (Courtesy of HiLine Engineering & Fabrication.)



FIGURE 33.7 Detail of the Hanford site 25-cm² bullet and 47-cm² C-ring channel electrodes. (Courtesy of Glenn Hedgemon, HiLine Engineering & Fabrication.)

stainless steel tubing. Eight three-electrode channels are distributed along the probe body. Electrodes are fabricated from UNS K02400 steel that has been heat treated to match the tank wall heat treatment. Four channels on each probe are formed from sets of bullet-shaped electrodes (25 cm² per electrode). Four channels are formed from sets of thick-walled C-rings (44 cm² per electrode). Figure 33.7 shows two channels on the most recent probe. The unstressed bullet-shaped electrodes are used for pitting and uniform corrosion detection. The working electrode on each C-ring channel is notched, precracked and stressed to yield prior to installation to facilitate the monitoring of SCC should tank chemistry conditions change to allow the onset of cracking. The other two C-rings on each C-ring channel are not stressed to match the operating conditions of the vessel. Bullet and C-ring channels alternate up the length of the probe. Current DST waste levels in monitored tanks immerses three channels of bullet-shaped electrodes and three channels of C-ring electrodes.

In this way, the working electrode is allowed to behave in a manner most representative of the material in-service, thus providing corrosion information reflecting the real-life situation of the plant equipment. The exposure of single or multiple corrosion probes can enable informed decisions to be made regarding the choice of a material or of a stress relief process.

Corrosion in Hydrocarbon Environments In hydrocarbon environments there must be an electrolytically conductive phase present, which is generally provided by an aqueous phase or by polar solvents, for corrosion to occur. Examples may be flowlines in oil and gas applications or pipelines in chemical process environments where it may not be straightforward to introduce a complex probe system.

A circumferential spool probe has been developed specially for this application to allow a maximum contact of the electrodes with the process environment in “true” flowing conditions along flowlines or pipelines, especially in multiphase oil and gas applications.

The basic ring principle of these sensors allows elements to be made by “salami” slicing a pipe and reassembling pairs of the resulting rings, separated by insulation, to remake

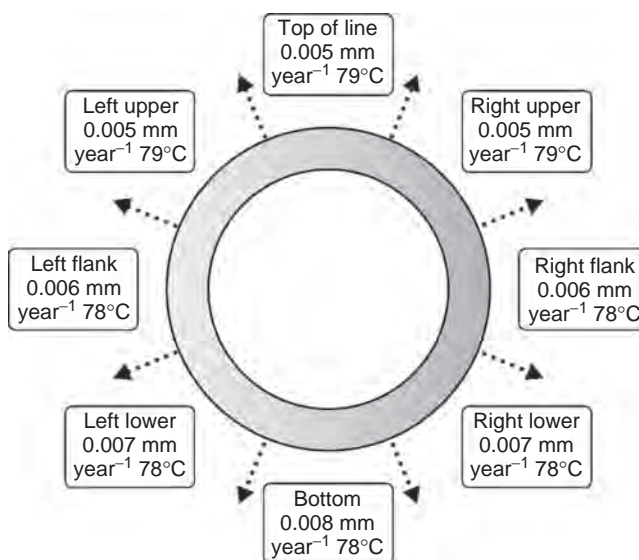


FIGURE 33.8 The principle of ring pair corrosion monitoring. (Courtesy of Cormon Ltd.)

a pipe section capable of retaining line pressure. Each electrically isolated ring is measured using pick up wires attached to the outer face. If wires are attached at equally spaced intervals around the ring the instrumentation can be configured to measure the overall metal loss and the loss in the segment between each pick-up point Figure 33.8.

One of each pair of rings is kept from contact with the process stream by means of a high integrity, thin-film, ceramic coating. The coated ring acts as the reference to the exposed sample ring.

A number of pairs of rings may be used enabling the study of different materials including weldment and HAZ material if preferential weld corrosion is an issue. In addition to the temperature data from the elements, it is a simple matter to include pressure measurement in the device. Together these may add considerably to the understanding of the behavior of the fluid in the line.

Standard rings have the same wall thickness as the original line so no problem should arise with element life in relation to the service life of the line. The thicker wall rings do have a lower speed of response that may not always meet the requirement, especially if real-time adjustment to chemical treatment is proposed. In this case, a combination of two concentric rings may be used to provide a fast response from a thin element inside a thicker supporting ring.

The potentially limited life of the element is balanced by the ability to maintain low corrosion rates through active control, thereby extending the life of the asset and the sensor. The spool sensor is, therefore, a very versatile measurement tool for looking at the character and degree of corrosion of various materials in conditions that are a true representation of line flow. It is capable of being finely tuned to perform the required task with great precision.

The sensor housing uses a double pressure barrier principle. The sensor rings, spacers, and isolators are held in compression by a clamp arrangement producing an inner-pressure tight cylinder. This cylinder is mounted in the outer housing using a pair of



FIGURE 33.9 A ring pair corrosion monitoring system being deployed at sea during the installation of a submerged pipeline. (Courtesy of Cormon Ltd.)

elastomer seal rings to complete the primary containment. The outer housing is a pressure tight assembly in its own right, sealed using flanges, ring type joints, and a spacer ring. Electronics, housings, power consumption, and telemetry are similar in most respects to those for intrusive probes. Figure 33.9 illustrates a monitoring system being deployed with a pipeline as it is submerged in sea.

Coupled Multielectrode Array Systems and Sensors The use of multielectrode array systems (CMAS) for corrosion monitoring is relatively new. The advantages of using multiple electrodes include the ability to obtain greater statistical sampling of current fluctuations, a greater ratio of cathode-to-anode areas that enhances the growth of localized corrosion once initiated, and, depending on the design, the ability to estimate the pit penetration rate and to obtain macroscopic spatial distribution of localized corrosion (Yang et al., 2002).

Figure 33.10 shows the principle of the CMAS in which a resistor is positioned between each electrode and the common coupling point. Electrons from a corroding or a relatively more corroding electrode flow through the resistor connected to the electrode and produce a small potential drop usually of the order of a few microvolts. This potential drop is measured by the high-resolution voltage-measuring instrument and used to derive the current of each electrode. The CMAS probes can be made in several configurations and sizes, depending on the applications. Figure 33.11 shows some of the typical probes that were reported for real-time corrosion monitoring.

The data from these CMAS probes are the large number of current values measured at a given time interval from all the electrodes. In a CMAS probe system, these data are reduced to a single parameter so that the probe can be conveniently used for real-time and online monitoring purposes. The most anodic current has been used as a one-parameter signal for the CMAS probes. Because the anodic electrodes in a CMAS probe simulate the anodic sites on a metal surface, the most anodic current may be considered as the corrosion current from the most corroding site on the metal.

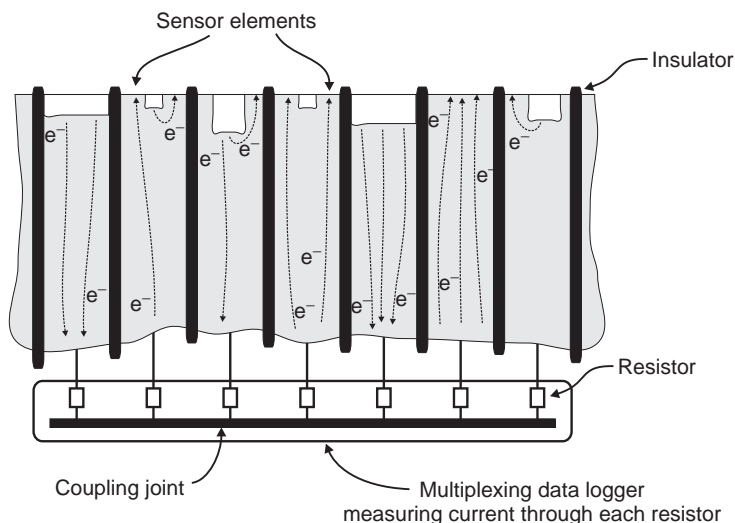


FIGURE 33.10 Multielectrode array multiplexed current and/or potential measurement. (Adapted from Yang and Sridhar (2003).)

The value based on three times of the standard deviation of currents is another way to represent the corrosion current from the most corroding site on the metal. Because the number of electrodes in a CMAS probe is always limited and usually far fewer than the number of corroding sites on the surface of a metal coupon, the value based on the statistical parameter, such as three times the standard deviation of current, was considered to be more appropriate than the single value of the most anodic current. The standard deviation value may be from the anodic currents or from both the anodic and cathodic currents.

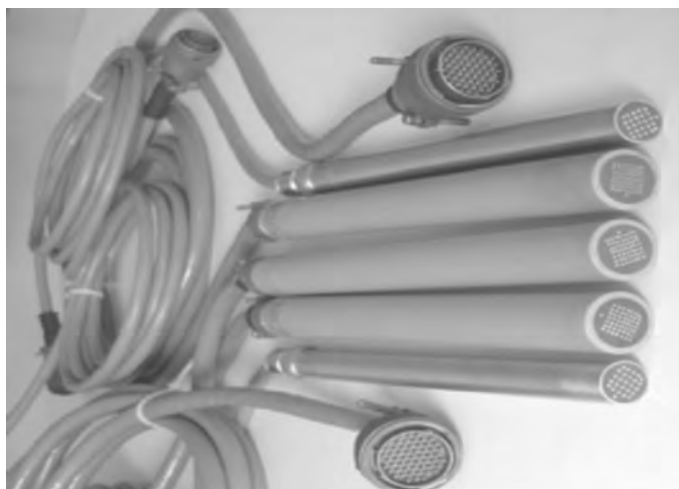


FIGURE 33.11 Typical CMAS probes used for real-time corrosion monitoring. (Courtesy of Corr Instruments, LLC.)

In a less corrosive environment or with a more corrosion-resistant alloy, the most anodic electrode may not be fully covered by anodic sites until the electrode is fully corroded. Therefore, the most anodic electrode may still have cathodic sites available, and the electrons from the anodic sites may flow internally to the cathodic sites within the same electrode. The total anodic corrosion current, I_{corr} , and the measured anodic current, I_{a}^{ex} may be related by Equation (33.1).

$$I_{\text{a}}^{\text{ex}} = \varepsilon I_{\text{corr}} \quad (33.1)$$

where ε is a current distribution factor that represents the fraction of electrons resulting from corrosion that flows through the external circuit. The value of ε may vary between 0 and 1, depending on parameters such as surface heterogeneities on the metal, the environment, the electrode size, and the number of sensing electrodes. If an electrode is severely corroded and significantly more anodic than the other electrodes in the probe, the ε value for this corroding electrode would be close to 1, and the measured external current would be equal to the localized corrosion current.

Because the electrode surface area is usually between 1 and 0.03 mm², which is ~ 2 to 4 orders of magnitude less than that of a typical LPR probe or a typical electrochemical noise (EN) probe, the prediction of penetration rate or localized corrosion rate by assuming uniform corrosion on the small electrode is realistic in most applications. CMAS probes have been used for monitoring localized corrosion of a variety of metals and alloys in the following environments and conditions:

- Deposits of sulfate-reducing bacteria.
- Deposits of salt in air.
- High-pressure simulated natural gas systems.
- H₂S systems.
- Oil–water mixtures.
- Cathodically protected systems.
- Cooling water.
- Simulated crevices in seawater.
- Salt-saturated aqueous solutions.
- Concentrated chloride solutions.
- Concrete.
- Soil.
- Low-conductivity drinking water.
- Process streams of chemical plants at elevated temperatures.
- Coatings.

Atmospheric Corrosion Monitoring For atmospheric corrosion monitoring, the required electrolyte for a probe to generate an electrochemical signal may either be a fine mist or a discontinuous film made by condensing humid air. The high resistivity of such an electrolyte means that the probe electrodes must be close while still electrically insulated to function properly. In addition, a further complication is the extremely low

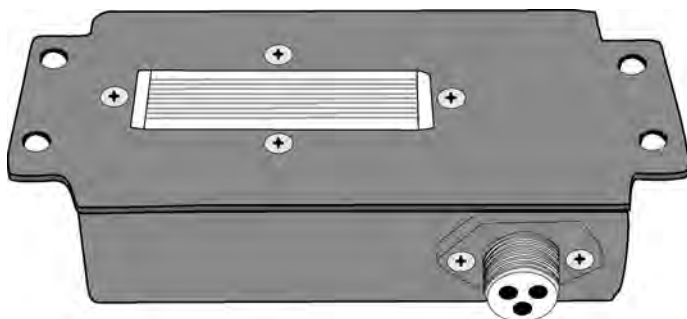


FIGURE 33.12 Electrochemical probe manufactured from an uncoated aluminum alloy in the form of closely spaced elements.

electrochemical activity normally produced by atmospheric corrosion, which means that the following requirements must be met to produce useful results:

- High sensitivity of the measuring instrumentation.
- Minimal IR drop throughout the monitoring system.
- Relatively large electrodes to maximize the opportunity to record corrosion signals.

Figure 33.12 shows schematically a probe configuration to achieve such measurements in a low conductivity condensing environment for monitoring aircraft corrosion. One such corrosion surveillance system was installed on an unpressurized transport aircraft. Electrochemical probes in the form of closely spaced probe elements were manufactured from an uncoated aluminum alloy. All but one of the probes were located inside the aircraft, in the areas that were most prone to corrosion attack and difficult to access. Another probe was located outside the aircraft, in its wheel bay (Roberge et al., 1996). In flights from inland to marine atmospheres, a distinct increase in corrosivity was recorded by potential noise surveillance signals during the landing phase in the marine environment (Figure 33.13). However, the strongest localized corrosion signals were recorded at ground level in a humid environment (Figure 33.14).

Another example of an atmospheric corrosion probe is shown in Figure 33.15. This sensor was fabricated using microcircuit technology in which a thin polyamide film was electroplated with two different metals (gold and cadmium) in a pattern to maximize the galvanic current produced in the presence of an even moderately corrosive environment. With this sensor the galvanic current produced by the thin-film bimetallic elements is integrated with a coulometer as a function of time. The data are then stored in a memory chip for future download when queried through a radio frequency (rf) data-gathering transceiver (DGT) or transponder to a laptop computer.

33.3.3.4 Location of Monitoring Hardware Many industrial plants have intrinsic safety requirements that impose important restrictions on corrosion monitoring systems. To ensure flexibility on large plants, some organizations have adopted the strategy of using a “mobile” corrosion monitoring laboratory that meets their safety regulations. Such a laboratory housing the corrosion monitoring instrumentation can be conveniently moved to

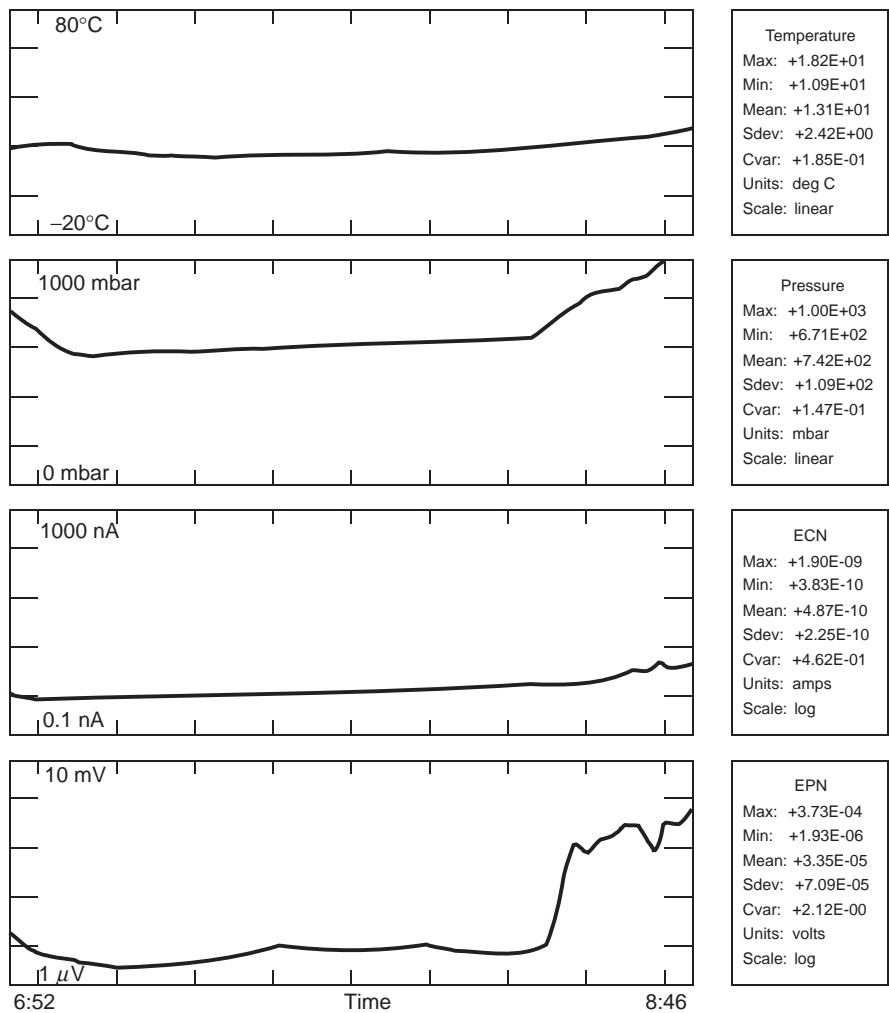


FIGURE 33.13 Temperature, pressure, and electrochemical signals as a function of time during a flight to a marine environment in South Africa.

different locations as required in order to overcome the problems associated with excessive lengths of sensor leads. Such an arrangement additionally provides a protective environment for measuring equipment and data storage hardware, which could otherwise be damaged in corrosive atmospheres.

Mobile laboratories may also be used for corrosion measurements on treated water circuits. A schematic of the water handling facility to carry out microbiologically influenced corrosion (MIC) field tests is shown as an example in Figure 33.16 (Roberge and Sastri, 1994). This facility was used for the growth and monitoring of naturally forming biofilms in five separate slip-streams. Each line accommodated 40 sample coupons and was equipped with an individual injector system that allowed testing of the effect of biocides on active biofilms in a fully equipped trailer (Figure 33.17).

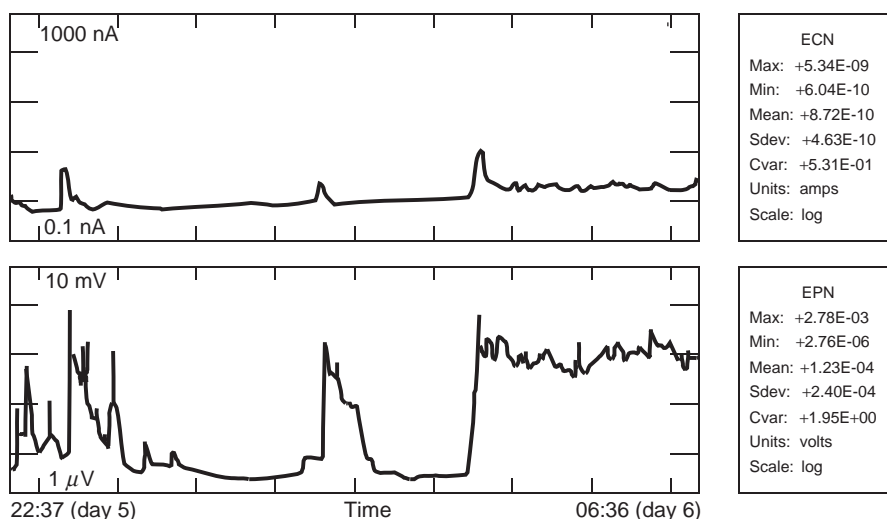


FIGURE 33.14 Electrochemical signals as a function of time in a marine environment in South Africa.

33.3.3.5 Sensitivity and Response Time The usefulness of a corrosion monitoring system strongly depends on how well it can deliver warnings of unwanted corrosion conditions. For measuring techniques this translates into two closely related properties:

1. The sensitivity to detect a change in corrosion rate.
2. The time it requires to detect such a change (i.e., the response time).

By virtue of the measuring principle of many systems, sensitivity and response time have an inverse relation to each other. In order to compare corrosion monitoring systems on the basis of their sensitivity, it is important to distinguish between the accuracy of the measurement and the sensitivity to measure a change in the corrosion rate. The sensitivity to

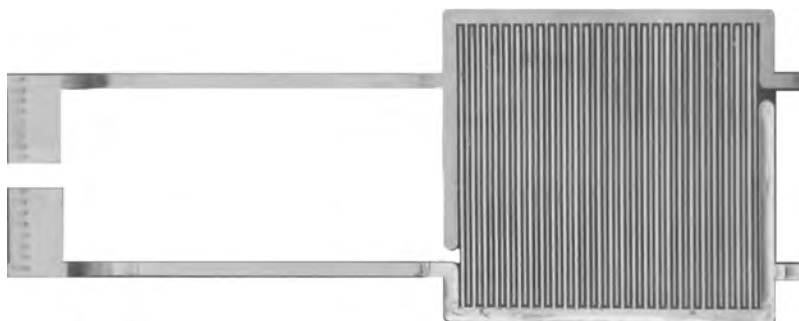


FIGURE 33.15 Thin-film sensor for atmospheric corrosion monitoring. (Courtesy of Kingston Technical Software.)

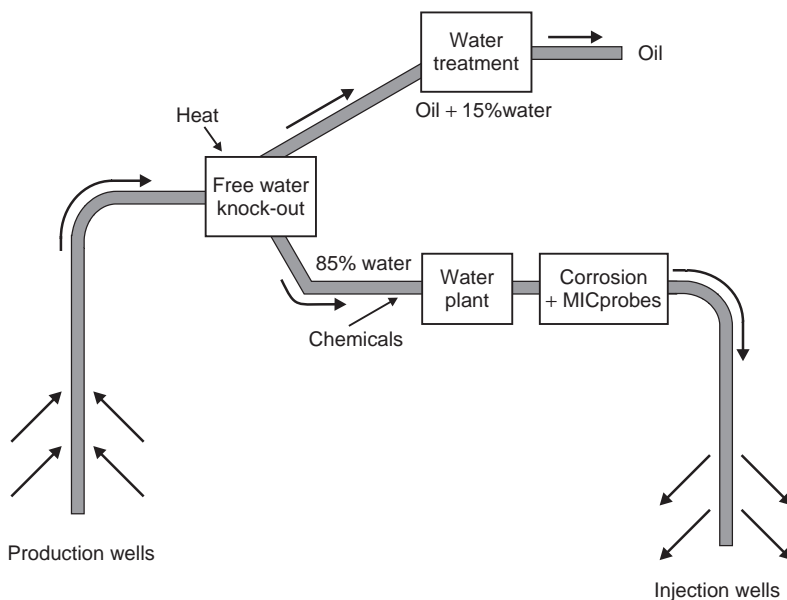


FIGURE 33.16 Schematic of the water handling facility at the MIC field testing facility.

measure corrosion rate is the combined result of measurement accuracy and the elapsed time (Thomas and Terpsta, 2003).

Sensitivity (S) and response time (R) of a certain technique are closely related, and can be conveniently displayed in a single graph, as shown in Figure 33.18. It is most convenient to display such S – R curves on a log–log graph. The example in this figure requires a 1–7 days response time while requiring sensitivity in the order of $1\text{--}20\text{ mm/year}^{-1}$



FIGURE 33.17 Oil recovery microbial test lines to evaluate biocide programs. (Courtesy Kingston Technical Software.)

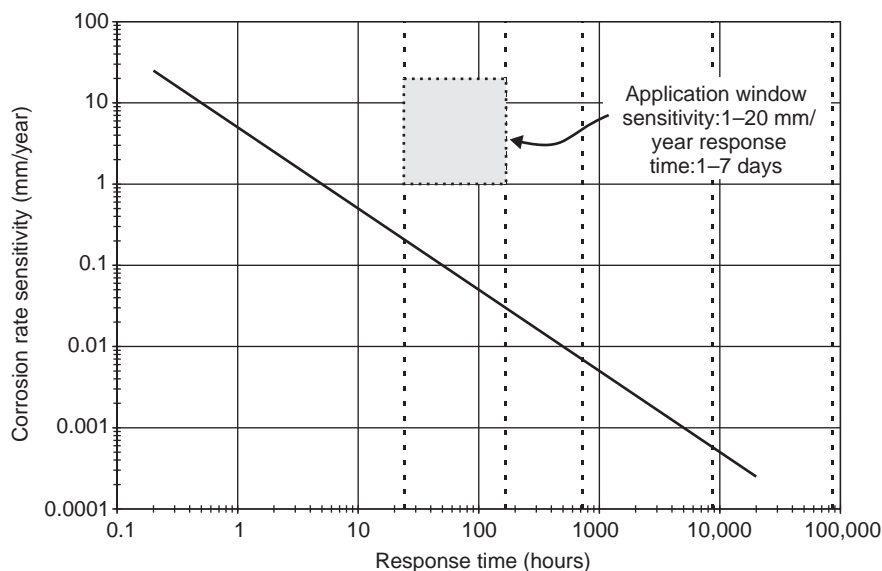


FIGURE 33.18 A corrosion monitoring system sensitivity (1–20 mm/year)/response time (1–7 days) application window for a given system performance threshold (solid line).

corrosion rate measurement. The S – R -curve of a specific technique lies below the window, hence will satisfy each requirement for this application. In Table 33.1, typical applications are given with their characteristics. These applications are depicted in Figure 33.19 for their respective S – R windows.

A monitoring system can also be limited to measure accurately over long time scales because of inherent instability of the monitoring system. This might be the result of deterioration of the sensor, or drift in the recording instrumentation. For systems that have to measure with high sensitivity and over a long interval it is recommended to perform routine verifications and calibrations of the monitoring equipment.

TABLE 33.1 Sensitivity and Response Time of Typical Corrosion Monitoring Applications

Application	Sensitivity Range (mm year ⁻¹)	Response Time	System Characteristics
Corrosion tests	0.1–100	1 h–5 days	Continuous
Inhibition control	0.1–20	0.5–2 days	Continuous optimization
Corrosion control (upsets)	1–100	1 h–2 days	Continuous monitoring (upsets)
Corrosion control performance demonstration	1–10	1 week– 1 month	Continuous/interval measurement
Inspection planning	0.2–10	1 month– 0.5 year	Interval
Inspection	1–20	3 months– 10 year	Interval

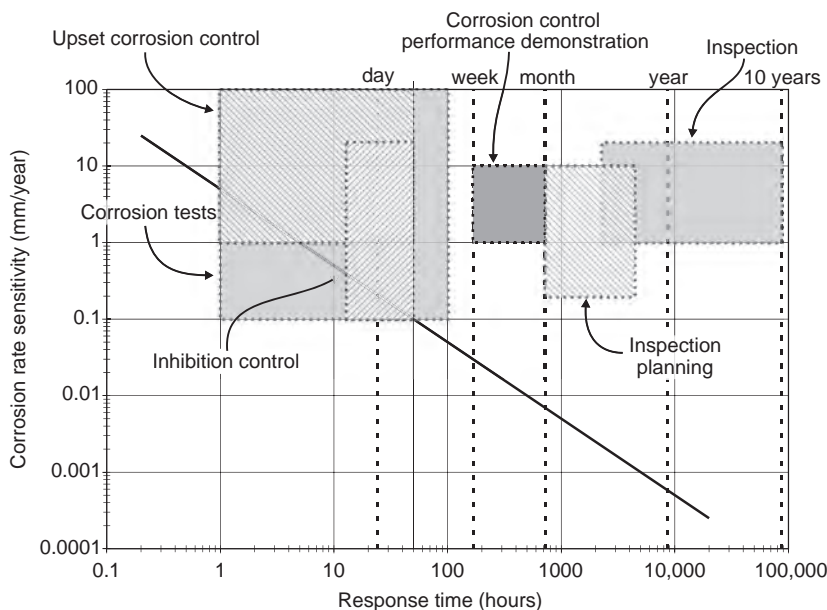


FIGURE 33.19 Application windows depicted in the S-R plot. The size range of the application windows is given in Table 33.1.

33.3.3.6 Data Communication and Analysis Requirements It is important, at the onset of a corrosion monitoring program, to define the full data communication chain from signaling unacceptable corrosion to the implementation of a remedial action. The times for each step in the chain should be in balance, that is, it is obviously not very useful to invest in a system with a response time of 1 day if it requires weeks to process the information and/or months to implement follow-up remedial measures. The following individuals may be involved in the communication process (Thomas and Terpsta, 2003):

1. Process plant operator, to collect data.
2. Corrosion monitoring specialist (corrosion or inspection engineer), to process data.
3. Corrosion engineer, to assess the information and determine follow up.
4. Operations or maintenance engineer, to plan and implement remedial action.

The response time from “sensor-to-desk” for the steps 1–3 determine the actual response time obtained from a corrosion monitoring system. For a highly critical monitoring task, the data might go directly to the party responsible for remedial action, for example, to the control room, for action by an operator.

The perceived importance of the monitoring system and strategy has to be mirrored by commitment of all individuals involved in integrity management, that is, the asset holder, usually operations, but also maintenance and inspection staff, corrosion engineering, production chemists, and frequently the chemical treating contractor. It is essential that the approach is agreed and implemented by a team that includes these individuals, who

together decide not only how corrosion should be controlled, but also how the corrosion monitoring should be implemented.

33.3.3.7 Define System Reliability Note that if the monitoring system is an active operational component, a high demand is put on its reliability. For example, large economic losses could be incurred if the monitoring system used to control the inhibitor concentration of a system with a high-uninhibited corrosion rate would fail. For such an application, a quantitative reliability study of the monitoring system may be warranted (Thomas and Terpsta, 2003).

If a monitoring system has a passive role, for example, if it is used to verify that corrosion conditions have been within expected limits, a more qualitative analysis of possible failures might be adequate. It may well be that there are other sources of information to estimate corrosion progress besides monitoring, or it may be necessary to have more than one system in place to warn if extreme degradation is possible.

33.3.3.8 Availability and Cost The final choice of the corrosion monitoring system will be influenced by availability of the tools and their cost. Some methods and in particular the installation and implementation of the overall corrosion monitoring system can be quite costly. In view of the implications of excessive corrosion and loss of integrity, the choice should be based on the total cost of ownership of the equipment (Thomas and Terpsta, 2003).

33.3.3.9 In a Nutshell An effective corrosion monitoring system or device should exhibit the following characteristics (Zintel et al., 2001):

- *User Friendly*: The monitoring system must be simple to install, simple to use, and simple to interpret by system operators. At least some interpretation functions must be sufficiently developed so that the system can be interfaced to alarms and controllers for chemical treatment additions or on-line cleaning systems.
- *Rugged*: The monitoring system must be able to withstand the normal use and abuse if it is deployed in an industrial environment.
- *Sensitive*: The monitoring element or probe must be sensitive to the onset of a corrosion problem and provide a definitive indication in real-time that may be used as a process control variable or to evaluate the effectiveness of a control measure.
- *Accurate*: False positives and negatives or any indications caused by interferences from effects, such as flow, erosion, and fouling, can be detrimental in many ways. Erroneous readings may seriously affect the credibility and straightforward usefulness of a corrosion monitoring program.
- *Maintainable*: Probes are expected to foul in service. A minimum time between servicing operations of several months to several years may be required for most applications. Periodic servicing and calibration should be simple and easy to perform.
- *Cost Effective*: The cost of the monitoring system must be significantly less than the cost of the downtime that is avoided or the treatment costs that are saved. The speed and accuracy of the technique are also factors in the cost effectiveness of the monitoring system.

33.3.4 Corrosion Monitoring Techniques

Assessment of corrosion in field conditions is complex due to the wide variety of applications, process conditions, and fluid phases that exist in industrial systems. As discussed in Chapter 3, the expectations of a corrosion monitoring program will vary greatly between organizations that have well-established proactive corrosion management programs and other organizations where corrosion damage is simply a nuisance. Many of the possible corrosion monitoring and inspection techniques available have been recently organized by a group of experts and interested users in different categories as shown in Tables 33.2 and 33.3 (Techniques for Monitoring Corrosion and Related Parameters in Field Applications, 1999).

In the report produced by this group, a direct technique is one that measures parameters directly affected by the corrosion processes while an indirect technique provides data on parameters that either affect, or are affected by the corrosivity of the environment or by the products of the corrosion processes. Additionally, a technique can be described as being intrusive if it requires access through a pipe or vessel wall in order to make the measurements. Most commonly intrusive techniques make use of some form of probe or test specimen, which include flush mounted probe designs. Some indirect techniques can

TABLE 33.2 Direct Corrosion Measurement Techniques

Intrusive Techniques
<i>Physical techniques</i>
Mass-loss coupons
Electrical resistance (ER)
Visual inspection
<i>Electrochemical dc techniques</i>
Linear polarization resistance (LPR)
Zero-resistance ammeter (ZRA) between dissimilar alloy electrodes: galvanic
Zero-resistance ammeter (ZRA) between the same alloy electrodes
Potentiodynamic–galvanodynamic polarization
Electrochemical noise (ECN)
<i>Electrochemical ac techniques</i>
Electrochemical impedance spectroscopy (EIS)
Harmonic distortion analysis.
Nonintrusive Techniques
<i>Physical techniques for metal loss</i>
Ultrasonics
Magnetic flux leakage (MFL)
Electromagnetic: eddy current
Electromagnetic: remote field technique (RFT)
Radiography
Surface activation and gamma radiometry
Electrical field mapping
<i>Physical techniques for crack detection and propagation</i>
Acoustic emission
Ultrasonics (flaw detection)
Ultrasonics (flaw sizing)

TABLE 33.3 Indirect Corrosion Measurement Techniques

On-Line Techniques
<i>Corrosion products</i>
Hydrogen monitoring
<i>Electrochemical techniques</i>
Corrosion Potential (E_{corr})
<i>Water chemistry parameters</i>
pH
Conductivity
Dissolved oxygen
Oxidation reduction (redox) potential
<i>Fluid detection</i>
Flow regime
Flow velocity
<i>Process parameters</i>
Pressure
Temperature
Dewpoint
<i>Deposition monitoring</i>
Fouling
<i>External monitoring</i>
Thermography
Off-Line Techniques
<i>Water chemistry parameters</i>
Alkalinity
Metal ion analysis (iron, copper, nickel, zinc, and manganese)
Concentration of dissolved solids
Gas analysis (hydrogen, H_2S , and other dissolved gases)
Residual oxidant (halogen, halides, and redox potential)
Microbiological analysis (sulfide ion analysis)
<i>Residual inhibitor</i>
Filming corrosion inhibitors
Reactant corrosion inhibitors
<i>Chemical analysis of process samples</i>
Total acid number
Sulfur content
Nitrogen content
Salt content in crude oil

serve to monitor various parameters on-line in real-time while others provide information off-line after samples collected from process streams or other operational locations are further analyzed following an established method.

The following sections will generally follow the organization used in that report with the notable exceptions of visual inspection and the direct nonintrusive techniques that are usually described as either nondestructive evaluation (NDE), nondestructive testing (NDT), or nondestructive inspection (NDI) techniques.

33.3.4.1 Direct Intrusive Techniques

Physical Techniques Physical corrosion monitoring techniques determine corrosion damage by measuring changes in the geometry of exposed coupons or test specimens. There are many properties of a test specimen that may change to some degree as a result of corrosion, such as its mass, electrical resistance, magnetic flux, reflectivity, stiffness, or any other mechanical properties. When the physical property is measured by electronic means, the test specimen can remain *in situ* and frequent readings are possible. However, frequent readings are less easy to implement if the test specimen needs to be removed from the process to have its physical property measured.

MASS LOSS COUPONS Mass loss coupons are usually designed to monitor the damage rate occurring on existing equipment, to evaluate alternative materials of construction, and sometimes to determine the effects of process conditions that cannot be reproduced in the laboratory. This simple and low cost method of corrosion monitoring consists in exposing small specimens in the environment of interest for a specific period of time and subsequently removing them for weight loss and more detailed examination. Even though the principle is relatively simple, numerous potential pitfalls exist, which can be avoided by following the recommendations of a comprehensive standard guide, such as ASTM G-4 (Standard Guide for Conducting Corrosion Tests in Field Applications, 2001).

Various means of introducing coupons of materials of construction into operating equipment have been devised. For equipment operating under considerable pressure, special attachments are available from corrosion equipment suppliers. A single high-pressure access fitting for insertion of a retrievable corrosion probe (Figure 33.20) can be used to accommodate most types of retrievable probes (Figure 33.21). As mentioned earlier, specialized tools exist for sensor insertion and withdrawal under pressurized operating conditions (Figure 33.22).

Mass loss coupon testing provides several specific advantages over laboratory coupon testing. A large number of materials can be exposed simultaneously and ranked in actual process streams with actual process conditions. A slip-in rack, for example, can be inserted into and removed from equipment with a retractable coupon holder full (Figure 33.23). A rod-shaped coupon holder is contained in the retraction chamber, which is flanged to a full-port gate valve. The other end of the retraction chamber contains a packing gland through which the coupon holder can pass. Coupons mounted on the rod in the extended position are drawn into the retraction chamber. The chamber is bolted to the gate valve, which is then opened to allow the coupons to be slid into the process stream. The sequence is reversed to remove the test coupons. It is essential that the rod or handle be equipped with a restraining chain or other device to prevent the blowout of the specimen holder when retracting the specimens in a system under pressure.

Because many coupons can be exposed simultaneously, they can be tested in duplicate or triplicate (to measure scatter), and be fabricated to simulate such conditions as welding, residual stresses, or crevices. These variations can provide the engineer with increased confidence in selecting materials for new equipment, maintenance, or repair (Dean, 2003).

Coupons can be designed to detect such phenomena as crevice corrosion, pitting, and dealloying corrosion. The environment of interest can be the full process flow at a location where the conditions are deemed to be suitably severe to give a meaningful representation. Alternatively, coupons can be exposed in a side stream that can be

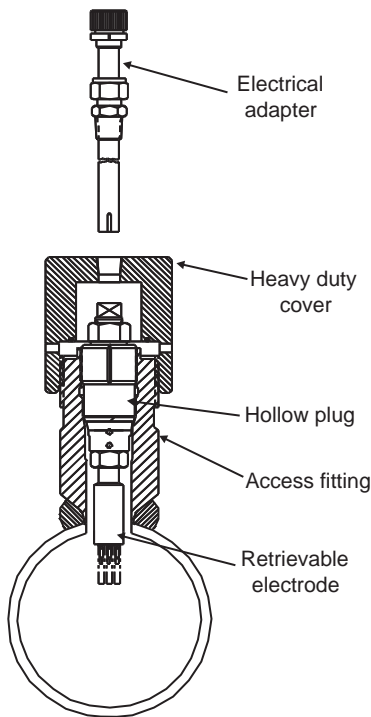


FIGURE 33.20 High-pressure access fitting for insertion of a retrievable corrosion probe. (Courtesy of Metal Samples Company, www.metalsamples.com.)

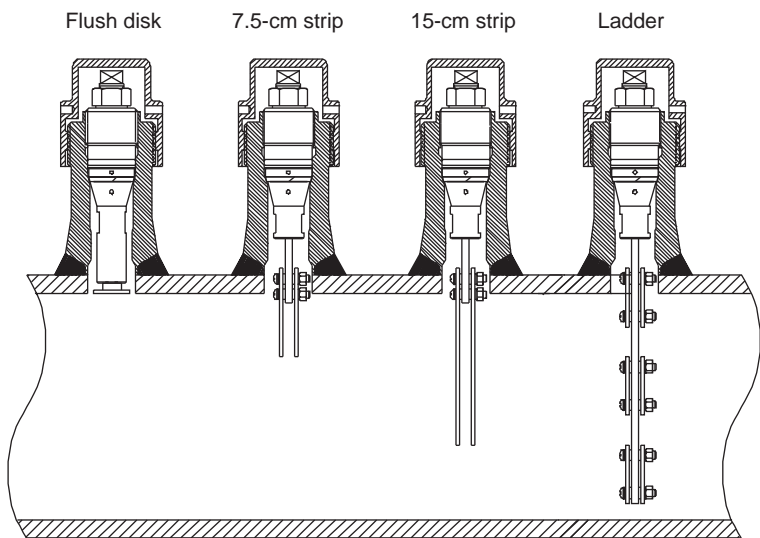


FIGURE 33.21 A single high-pressure access fitting can be fitted with different types of retrievable corrosion probes. (Courtesy of Metal Samples Company, www.metalsamples.com.)



FIGURE 33.22 Retrieval tool for removing corrosion probes under pressure. (Courtesy of Metal Samples Company, www.metalsamples.com.)

isolated from the main process stream. The design of the coupon should match the objective of the test, that is, simple flat sheets for general corrosion or pitting, welded coupons for local corrosion in weldments, stressed, or precracked test specimens for SCC problems (Techniques for Monitoring Corrosion and Related Parameters in Field Application, 1999).

Length of Exposure and Limitations In general, the length of exposure is typically as long as possible to allow for initiation of localized corrosion and adequate evaluation of service conditions. A minimum exposure of 3 months has normally been used for evaluation of pitting and crevice corrosion. For general corrosion, ASTM G-31 describes a minimum test duration in hours that is ~ 50 divided by the expected corrosion rate in millimeter per year (Standard Practice for Laboratory Immersion Corrosion Testing of Metals, 2004).

For example, when the corrosion rate is estimated to be $\sim 0.05 \text{ mm year}^{-1}$, the exposure period should be at least 1000 h. This can vary according to whether the coupons are exposed in well-controlled situations, such as laboratory tests or in the field. The use of

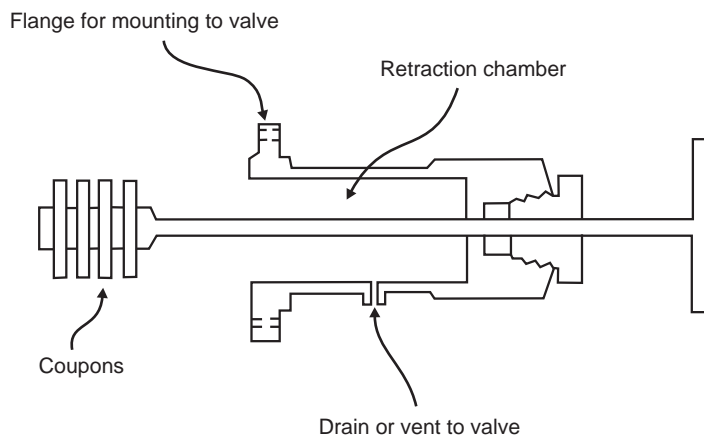


FIGURE 33.23 Illustration of retractable coupon holder (Dean, 2003).

coupons is, however, subject to the following limitations (Techniques for Monitoring Corrosion and Related Parameters in Field Applications, 1999; Dean, 2003):

- The technique only determines average rate of metal loss over the period of exposure.
- Coupon testing cannot be used to detect rapid changes in the corrosivity of a process.
- Localized corrosion cannot be guaranteed to initiate before the coupons are removed even with extended test durations.
- Reinsertion of a used coupon is generally not recommended.
- Short exposure periods normally yield unrepresentative average rates of metal loss. This is often the result of higher metal loss during initial acclimation to the process environment.
- Corrosion rate calculated from coupons may not reflect the corrosion of the plant equipment due to factors, such as multiphase flow, where the aqueous phase is much more corrosive than the organic or vapor phase, or turbulence from mixers, elbows, valves, pumps, and other items that accelerate the corrosion in a specific location in the equipment removed from where coupons were exposed.
- Procedures for mass-loss coupon analysis can be quite labor intensive.
- Another limitation of coupon testing is the simulation of erosion–corrosion and heat-transfer effects. Careful placement of the coupons in the process equipment can slightly offset these weaknesses.
- Coupons may be misleading in situations where the corrosion rate varies significantly over time because of unrealized process factors.

Contamination of the process fluid can also be an issue for corrosion monitoring for particular industries (e.g., food processing, medical, and electronic equipment manufacturing industries).

Uniform Corrosion Coupons The most common coupon shape for the evaluation of uniform corrosion is rectangular because most alloys are available in plate or sheet form.

Other shapes are used when there are restrictions on available product forms or when a specific material condition is required. Coupon identification must be legible and permanent. The simplest, and preferred, method of identification is stencil stamping (Dean, 2003).

Coupon finish represents a significant contribution to the overall cost. The least expensive finish that is consistent with the monitoring requirements should be selected. For example, an inexpensive surface finish is acceptable where carbon steel coupons are used routinely to monitor additions of inhibitor in water treatment programs. This may be achieved by punching or shearing, followed by glass bead blasting. On the other hand, when it is necessary to rank alloys in a process environment, the coupons must be finished with ground or machined parallel edges and sanded faces. Ideally, the surface finish of the coupon should match the finish of the equipment. However, this is hardly achievable for several reasons including the aging and scaling of real surfaces when exposed to process conditions.

Coupons that are cut by punching or shearing have cold-worked edges. The effects of cold work extend back from the cut edge a distance equal to the material thickness. These affected areas can be removed by grinding or machining. Cold-worked edges may affect the corrosion rate in some cases, and the residual stresses it induces may cause SCC in some materials. Extensive edge preparation can be a major contributor to the cost of a coupon.

Galvanic-Corrosion Coupons Pairs of test coupons can be coupled electrically to study the effects of galvanic corrosion. The relative areas exposed usually vary from 1:1 to 10:1 or greater. A major concern with electrically coupled coupons is maintaining electrical continuity for the entire exposure period. Corrosion product films can wedge mechanically joined coupons apart, thereby eliminating the electrical contact and any galvanic corrosion effect. ASTM G-71 provides valuable information on galvanic corrosion testing in both field and laboratory environments (Dean, 2003; Standard Guide for Conducting and Evaluating Galvanic Corrosion Tests in Electrolytes, 2003).

With metals that can become embrittled by hydrogen absorption (e.g., titanium, zirconium, tantalum, and hardenable steels) the cathodic or protected member of the galvanic couple may be subject to the greater damage. However, the typical mass loss measurements would not reveal such damage.

Crevice-Corrosion Coupons Equipment crevices are quite common in complex systems. These crevice sites can easily trigger the onset of localized corrosion in even benign process environments. Many metals perform differently in crevices as opposed to unshielded areas. The various techniques that can be used for crevice corrosion testing include rubber bands, spot-welded lap joints, and wire wrapped around threaded bolts. Each crevice test creates particular crevice geometry between specific materials and has a particular anode/cathode area ratio.

The two most widely used crevice geometries in field coupon testing employ insulating spacers to separate and electrically insulate the coupons. Spacers are usually either flat washers or multiple-crevice washers. Either type of spacer can be made of materials ranging from hard ceramics to soft thermoplastic resins. ASTM G-78 and G-48 provide valuable information on crevice corrosion specimen design and should be consulted before attempting a crevice corrosion test (Standard Guide for Crevice Corrosion Testing of Iron-Base and Nickel-Base Stainless Alloys in Seawater and Other Chloride-Containing Aqueous Environments, 2001; Standard Test Methods for Pitting and Crevice Corrosion

Resistance of Stainless Steels and Related Alloys by Use of Ferric Chloride Solution, 2003). In this ASTM G-78 test, washers make a number of contact sites on either side of the specimens (Figures 33.24 and 33.25a and b). The number of sites showing attack in a given time can be related to the resistance of a material to initiation of localized corrosion, and the average or maximum depth of attack can be related to the rate of propagation.

Stress-Corrosion Cracking Coupons Typical sources of sustained tensile stress that cause SCC of equipment in service are the residual stresses resulting from forming and welding operations and the assembly stresses associated with interference fitted parts, especially in the case of tapered, threaded connections. Therefore, the most suitable coupons for plant tests are the self-stressed bending and residual-stress specimens. Convenient coupons are the cup impression, U-bend (Standard Practice for Making and Using U-Bend Stress-Corrosion Test Specimens, 2003), C-ring (Standard Practice for Making and Using C-Ring Stress-Corrosion Test Specimens, 2001), tuning fork, and welded panel (Standard Practice for Preparation of Stress-Corrosion Test Specimens for Weldments, 1999). All these methods of stressing coupons produce a decreasing load as the cracks form and begin to propagate. Therefore, complete fracture is seldom observed, and careful examination is required to detect cracking.

Heat-Transfer Coupons For heat-transfer effects, specially designed coupons are required that simulate effects, such as those found in heating elements or condenser tubes.

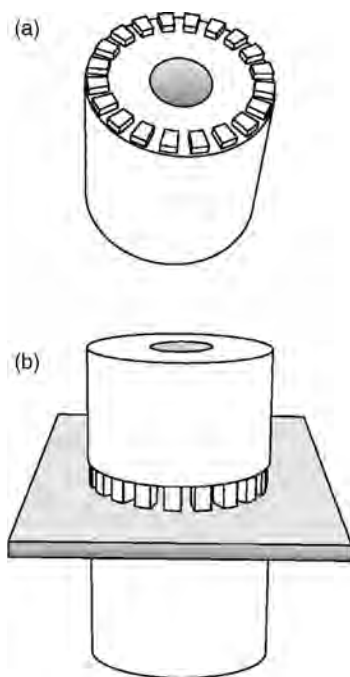


FIGURE 33.24 A schematic representation of the washer (a) and a washer assembly (b) for conducting an ASTM G 78 crevice susceptibility test.

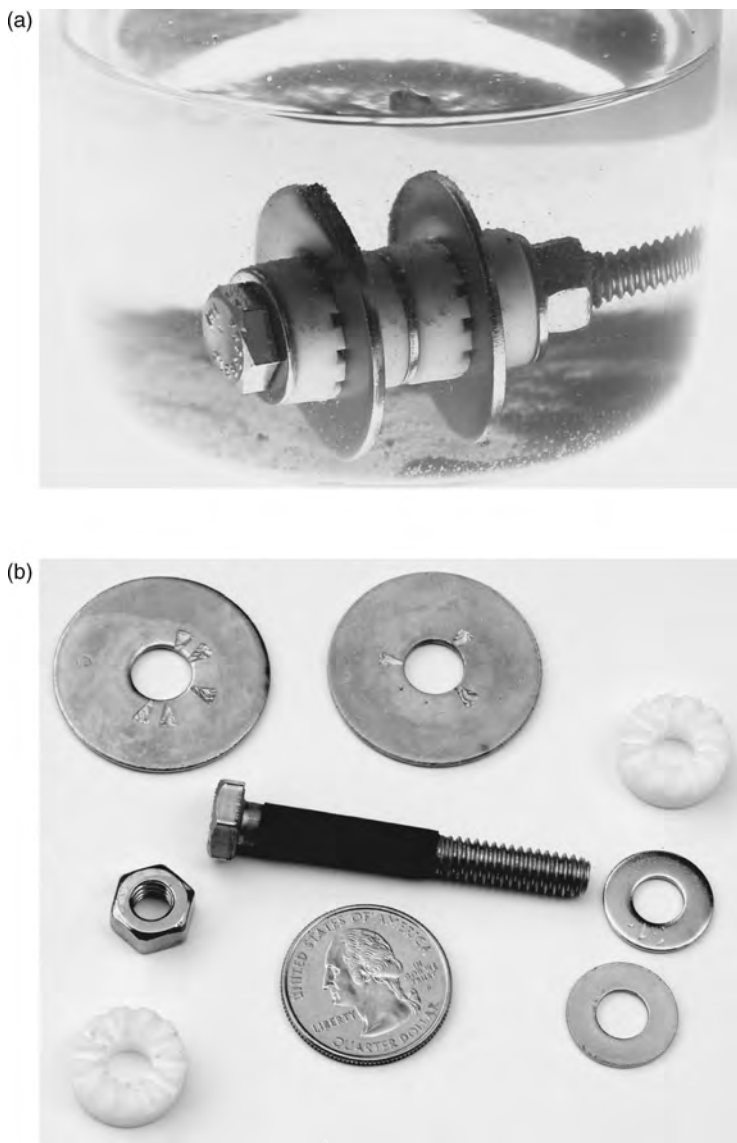


FIGURE 33.25 Crevice forming assembly in a beaker immersion test (a); components of the assembly and S30400 washer coupons after thirty days in a solution containing 4% NaCl and 8% FeCl₃ (b).

Coupons range in design from thermowell-shaped devices to sample tubes in a test heat exchanger. Thermowell-shaped devices are heated or cooled on the inside and project into the process stream (Standard Guide for Conducting Corrosion Tests in Field Applications, 2001). Heat-transfer tests can also be conducted in the laboratory. In this environment, the coupon forms part of the wall of the test vessel, and can therefore be heated or cooled

from one side. Because of the cost involved, heat-transfer coupon tests are usually carried out on only one (or perhaps two) alloys that have been selected from a larger group (Dean, 2003).

Welded Coupons Because welding is a principal method of fabricating equipment, the use of welded coupons is often desirable. Aside from the effects of residual stresses, the primary concern is the behavior of the weld bead and the heat-affected zone (HAZ). In some alloys, the HAZ becomes sensitized to severe intergranular (sometimes called knife-line) attack, and in certain other alloys, the HAZ is anodic to the parent metal. When possible, it is more realistic to remove welded coupons from production-sized weldments than to weld the small coupons (Dean, 2003).

Because the thermal conditions in both the weld metal and the HAZ are a function of the number of weld passes, the metal thickness, the weld position, and the welding technique, it is not usually considered good practice to use welded coupons to assess the possibility of sensitization from welding. It is generally better to carry out a sensitizing heat treatment on an unwelded coupon before it is tested and then look for evidence of intergranular corrosion or cracking.

Sensitized Metal Sensitization is a metallurgical change that occurs when certain noble alloys (e.g., austenitic and ferritic stainless steels, nickel base alloys) are heated under specific conditions. This may result in the precipitation of carbides or other intermetallic phases at grain boundaries that may reduce the corrosion resistance. Any heat-inducing process may cause sensitization, a process that is time and temperature dependent. There is a specific temperature range over which each particular alloy sensitizes rapidly.

Welding is the most common cause of sensitization. However, welded coupons may not exhibit sensitization because they may have been given insufficient weld passes when compared to actual process equipment. As a result, they spend insufficient time in the sensitizing temperature range, and susceptibility to intergranular corrosion may not be detected.

An appropriate sensitizing heat treatment guarantees that any corrosion susceptibility induced by welding or heat treatment is detected. The optimal temperature and time ranges for sensitization vary for different alloys. For example, 30 min at 650°C is usually sufficient to sensitize S31600 stainless steel. Some of the corrosion-resistant aluminum magnesium alloys containing 3–6% Mg, (e.g., 5××× series) are also subject to sensitization when heated at temperatures in the range of 65–175°C (Standard Test Method for Determining the Susceptibility to Intergranular Corrosion of 5××× Series Aluminum Alloys by Mass Loss After Exposure to Nitric Acid (NAML Test), 2004).

Coupon Cleaning and Evaluation The test coupons should be cleaned as soon as possible after removal from a test. The procedures for cleaning and weighing, which depend on the test material and the extent of corrosion, are described in ASTM standard G-1 (Standard Practice for Preparing, Cleaning, and Evaluating Corrosion Test Specimens, 2003). Examination of corrosion coupons after cleaning and weighing should reveal the forms of corrosion that may be expected in equipment made of the coupon material. Coupons are examined with the unaided eye, and then at increasing magnifications, up to 30–50×, with a binocular microscope. A scanning electron microscope (SEM) has often proved to be an extremely useful tool for detecting superficial localized defects (Dean, 2003).

In some cases, coupons must be bent and/or sectioned and metallographically examined to reveal certain types of corrosion damage. There are special localized corrosion effects that may not only jeopardize the determination of realistic corrosion rates, but also signal other serious types of behavior. Once the coupons have been cleaned thoroughly through repetitive cleaning processes, the corrosion or penetration rate can be estimated from a mass loss plot (Figure 33.26) (Standard Practice for Preparing, Cleaning, and Evaluating Corrosion Test Specimens, 2003; Hausler, 2005). The rate is estimated by Equation (33.2).

$$R = \frac{K(m_1 - m_2)}{A(t_1 - t_2)\rho} \quad (33.2)$$

where R is the penetration (corrosion) rate (millimeter per year); A is the exposed area (centimeter square); m_1 and m_2 are the initial and final masses (g), with m_2 being the intercept made by extrapolating line BC to the y axis in Figure 33.26; t_1 and t_2 are the starting and ending times (hours); ρ is the density (grams per centimeter cube); and K is a constant for unit conversion.

One question that arises when estimating the reproducibility of immersion test results is the amount of uncertainty that each measurement of the observables (e.g., time, mass loss, and area) contributes to the total uncertainty in Equation (33.2). This error defines the minimum uncertainty in the penetration rate that is possible during a given experiment. Such minimum uncertainty would be possible when (Freeman and Silverman, 1992):

- There is no localized corrosion.
- The penetration is uniform across the coupon surface.
- The projected and actual surface areas are the same.

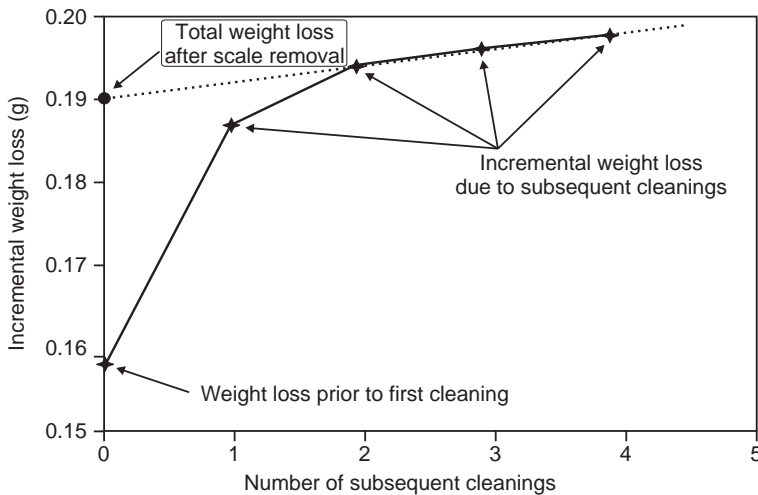


FIGURE 33.26 Cleaning procedure of corrosion coupons for weight loss determination yields scale weight, total corrosion weight loss, and error due to cleaning procedure (Standard Practice for Preparing, Cleaning, and Evaluating Corrosion Test Specimens, 2003; Hausler, 2005).

- Weights are unaffected by corrosion product removal.
- Areas have not changed during the exposure.
- The penetration rate is independent of time.

Accurate weight determination is essential to minimize the uncertainty. The balance should have an accuracy of at least 0.1 mg and weighing each coupon at least three times to obtain an average would decrease the uncertainty somewhat.

Assuming that the environment remains unchanged, the longer the coupons are exposed to the environment the smaller will be the error. As discussed earlier, adhering to the ASTM Standard G-31 recommendation is a good rule-of-thumb to minimize the error. However, the ability to control the environment may limit the test duration so that the long test times needed for accurate measurement of low corrosion rates may not be achievable.

ELECTRICAL RESISTANCE Acceptance of the electrical resistance (ER) corrosion monitoring method grew quickly after the correlation with corrosion rates was established in the 1950s (Dravnieks and Cataldi, 1954; Freedman et al., 1958). The principle of the widely used ER technique is quite simple, that is, the electrical resistance of a sensing element increases as its cross sectional area is reduced by corrosion damage. The electrical resistance of a metal or alloy element is given by Equation (33.3):

$$R = r \frac{L}{A} \quad (33.3)$$

where L is the probe element length (cm); A is the cross-sectional area (centimeter square); and r is the specific resistance of the probe metal (ohm centimeter).

Reduction or metal loss in the element cross section A due to corrosion will be accompanied by a proportionate increase in the element electrical resistance (R). Since temperature influences the electrical resistance of the probe element, ER sensors usually measure the resistance of a corroding sensor element relative to that of an identical shielded element (Figure 33.27). Commercial sensor elements are in the form of plates, tubes, plates, or wires (Figure 33.28).

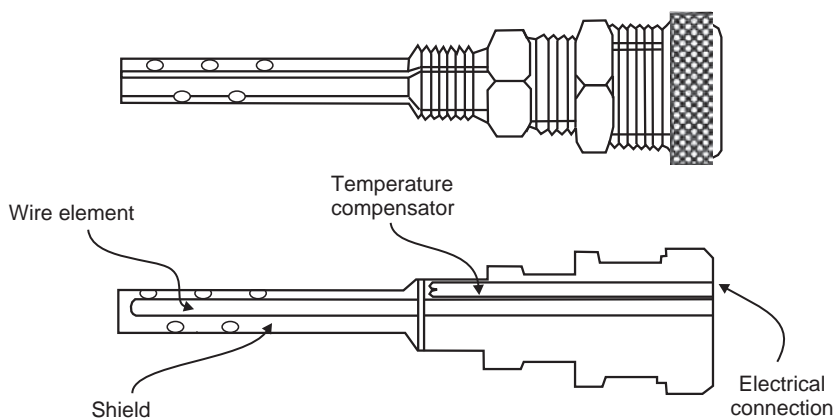


FIGURE 33.27 Illustration of an electrical resistance probe (Dean, 2003).



FIGURE 33.28 Commercial sensor elements to carry out ER measurements. (Courtesy of Metal Samples Company, www.metalsamples.com.)

Reducing the thickness of the sensor elements can increase the sensitivity of these sensors. However, improved sensitivity involves a trade-off against reduced sensor lifetime. The ER probe manufacturers provide guidelines showing this trade-off for different sensor geometries (Figure 33.29). These probes usually have a useful life up to the point where their original thickness has been halved with the exception of wire sensors. For ER wire sensors the lifetime is lower, corresponding to a quarter original thickness loss.

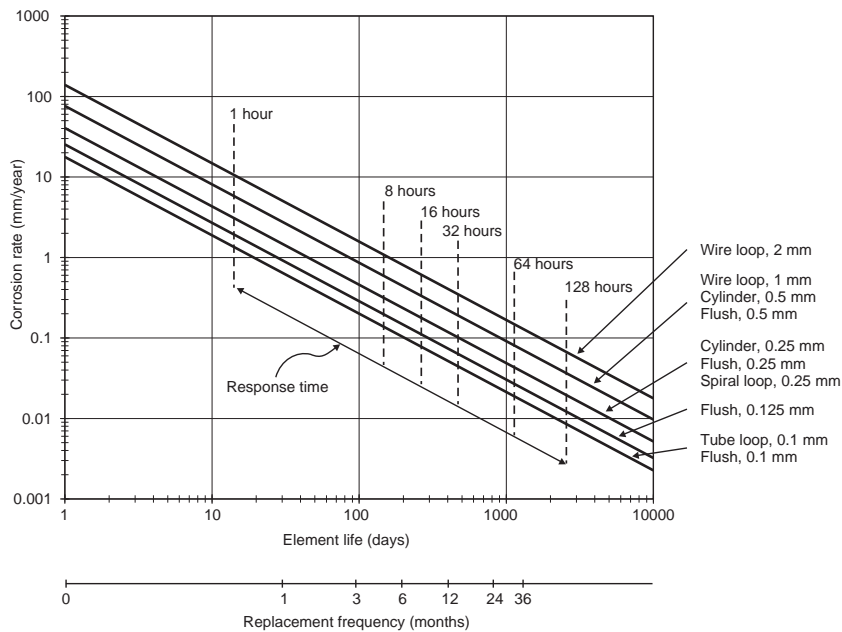


FIGURE 33.29 The ER probe element selection guide. (Adapted from Metal Samples.)

It is obvious that erroneous ER corrosion results will be obtained if conductive corrosion products or surface deposits form on the sensing element. Iron sulfide formed in sour oil-gas systems or in microbial corrosion and carbonaceous deposits in atmospheric corrosion are relevant examples. The same restriction also applies to electrically conductive environments, (e.g., molten salts or liquid metals).

To obtain the corrosion rate, a series of measurements are made over a period of time, and the results are plotted as a function of exposure time. The corrosion rate can be determined from the slope of the resulting plot (Standard Guide for On-Line Monitoring of Corrosion in Plant Equipment (Electrical and Electrochemical Methods), 2001).

There are several advantages to the ER corrosion monitoring method. Because probes are relatively small, they can be installed easily and the system wired directly to a control room location or to a portable resistance bridge at the probe location. For systems that are wired directly to control rooms, a computer system can be used to obtain the data and to transform the results in corrosion rate values. On the other hand, it is time consuming and sometimes impossible to take measurements at the probe site with a portable bridge. The temperature-compensation device reacts slowly, and it can be a source of error if the temperature varies when the measurement is taken (Dean, 2003).

Corrosion rate measurements obtained in short periods of time can also be inaccurate because the method measures only the remaining metal, which produces significant errors by estimating small differences between large numbers. Measurement resolution is typically 1 part in 1000 of the total measuring range of the probe, and the probe range is typically from 0.05 to 0.64 mm. Measured resistance changes are small so that thermal, stress, or electrical noise can affect the signal, necessitating hardware and software filtering.

The ER results provide a good measure of metal loss by general corrosion. However, the probes are less sensitive to effects of localized attack, which increase the element resistance on only a small area of the element, except near the end of probe life on loop element probes, where the localized attack completely corrodes through the element, increasing its resistance to infinity. Special probes have been prepared for sensitivity to crevice corrosion by creating multiple crevices on the measurement element, such as beads on wire loop probes.

INDUCTIVE RESISTANCE PROBES Following advances in electronics, signal processing, and measuring techniques, this new metal loss monitoring technology is a derivative of ER corrosion sensing. This corrosion monitoring technology has been developed to combine very high-resolution measurement with long probe life and the capability of intrinsically safe operation in hydrocarbon process plant environments (Denzine and Reading, 1998).

The thickness reduction of a sensing element is measured by changes in the inductive resistance of a coil embedded in the sensor (Figure 33.30). Sensing elements with high magnetic permeability intensify the magnetic field around the coil; therefore thickness changes affect the inductive resistance of the coil. Sensitivity has been claimed to be several orders of magnitude higher than with comparable ER probes.

The measurements are virtually unaffected by other process variables, such as temperature, hydrostatic pressure, impact loading (slugging), or flow regimes. The system is also immune to extraneous industrial noise, specifically electromagnetic induction, and thermally induced electromotive force voltages. The inductive resistance sensor elements have very high geometric and physical symmetry providing sensor surfaces with identical metallurgy and microstructures.

Electrochemical Techniques In view of the electrochemical nature of corrosion, it is not surprising that measurements of the electrical properties of the metal solution interface are so extensively used across a wide spectrum of corrosion science and engineering activities, from fundamental studies to monitoring and control in service. Electrochemical monitoring methods involve the determination of specific interface properties that can be divided into three broad categories:

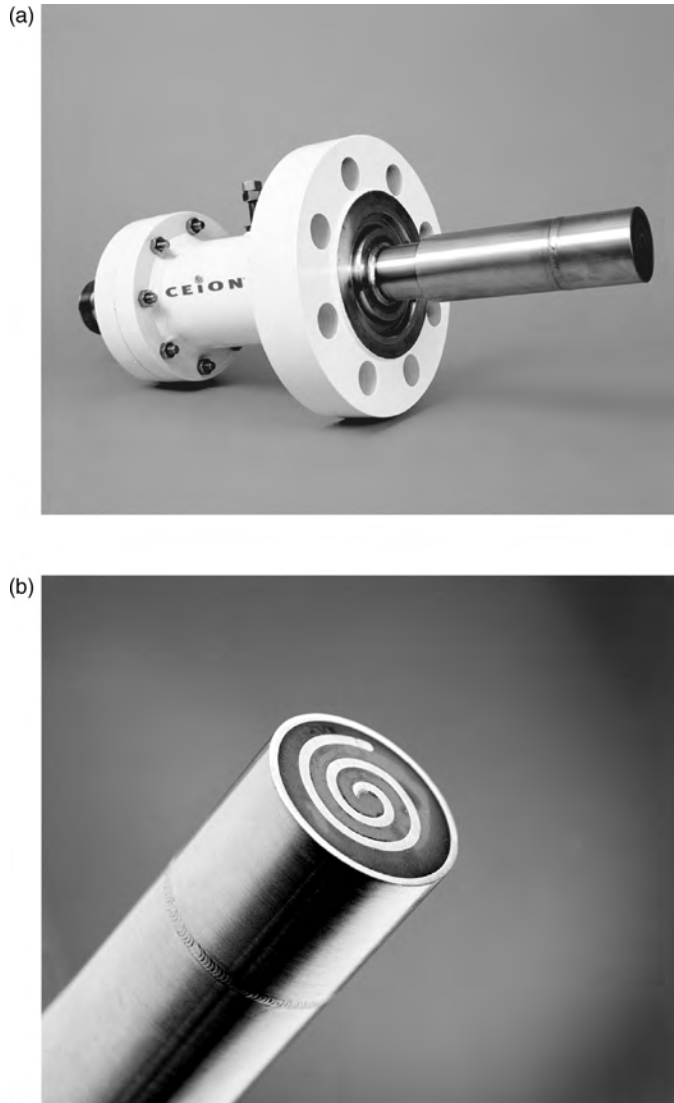


FIGURE 33.30 Subsea inductive resistance probe (a); probe element (b); spool instrument package for subsea (c). (Courtesy of Cormon Ltd.)



FIGURE 33.30 (Continued)

- *Corrosion Potential Measurements:* The potential at a corroding surface arises from the mutual polarization of the anodic and cathodic half-reactions constituting the overall corrosion reaction. Corrosion potential is intrinsically the most readily observable parameter and understanding its relation to the thermodynamics of a system can provide very useful information on the state of a system;
- *Reaction Rate as Current Density:* Partial anodic and cathodic current densities cannot be measured directly unless they are purposefully separated into a bimetallic couple. By polarizing a metal immersed in an aqueous environment, it is possible, with the use of simple assumptions and models of the underlying electrochemical behavior, to estimate net currents for both the anodic and cathodic polarizations from which a corrosion current density can be deduced;
- *Surface Impedance:* A corroding interface can also be modeled for all its impedance characteristics, therefore revealing subtle mechanisms not visible by other means. Electrochemical impedance spectroscopy (EIS) is now well established as a powerful technique for investigating corrosion processes and other electrochemical systems.

Corrosion potential or current produced by naturally occurring or externally imposed conditions can be measured with a variety of electrochemical techniques. Conversions of the measurements into corrosion rates or other meaningful data use equations or algorithms that are specific to each technique. The conversion of the measured current density into a corrosion rate, for example, can be made by using Faraday's law (Tables 33.4 and 33.5) and other empirically determined constants and factors specific to the system. Some techniques use analysis of the interface with direct current (dc) methods. Other techniques use alternating current (ac) methods to provide further characterization of the corrosion interface and the conductivity of the process fluids.

TABLE 33.4 Conversion Between Current, Mass Loss, and Penetration Rates for All Metals^{a,b}

	mA/cm ²	mm/year	mpy	g m ⁻² day ⁻¹
mA/cm ²	1	3.28 M/nday	129 M/nday	8.95 M/n
mm/year	0.306 nday/M	1	39.4	2.74 day
mpy	0.00777 nday/M	0.0254	1	0.0694 day
g m ⁻² day ⁻¹	0.112 n/M	0.365/day	14.4/day	1

^ampy, milliinch per year; *n*, number of electrons freed by the corrosion reaction; *M*, atomic mass; *d*, density.

^bNote: The table should be read from left to right, that is, 1 mA/cm², (3.28 M/nday) mm/year, (129 M/nday) mpy, (8.95 M/n)/g m² day.

Limits of operation for field work are more serious than those experienced in a laboratory environment, mostly for reasons of practical probe geometry. For example, capillary salt bridges (e.g., Luggin capillary) commonly used in laboratory setups to reduce the interference of solution resistance are definitively too delicate or cumbersome for field use (Techniques for Monitoring Corrosion and Related Parameters in Field Applications, 1999).

The widespread use of electrochemical polarization techniques, dc or ac, does not mean that they are without complications. The main complications or obstacles in performing polarization measurements can be summarized in the following categories:

- *Effect of Scan Rate:* The rate at which a potential or current is scanned may have a significant effect on the resulting polarization (Van Orden, 1998). The goal is for the polarization scan rate to be slow enough to minimize surface capacitance charging, otherwise some of the current being generated may serve to charge the surface capacitance with the net result that the measured current can be greater than the current actually generated by the corrosion reactions alone.
- *Effect of Solution Resistance:* The distance between the reference electrode and the working electrode is purposely minimized in most measurements to limit the effect of the solution resistance. In solutions that have extremely high resistivity (e.g., concrete, soils, and organic solutions), this effect can be an extremely significant.
- *Changing Surface Conditions:* Corrosion reactions take place on the metallic surface exposed to the environment and that surface can be modified by changing process conditions. This can have a strong effect on the polarization curves (Van Orden, 1998).
- *Determination of Pitting Potential:* In analyzing polarization curves the presence of a hysteresis loop between the forward and reverse scans often indicates that localized corrosion (e.g., pitting or crevice corrosion) is in progress. This observation has led to the

TABLE 33.5 Conversion Between Current, Mass Loss, and Penetration Rates for Steel

	mA/cm ²	mm/year	mpy	g m ⁻² day ⁻¹
mA/cm ²	1	11.6	456	249
mm/year ¹	0.0863	1	39.4	21.6
mp/year	0.00219	0.0254	1	0.547
g m ⁻² day ⁻¹	0.00401	0.0463	1.83	1

Note: The table should be read from left to right, that is, 1 mA/cm², 11.6 mm/year⁻¹, 456 mpy, 249 gm⁻²/day⁻¹.

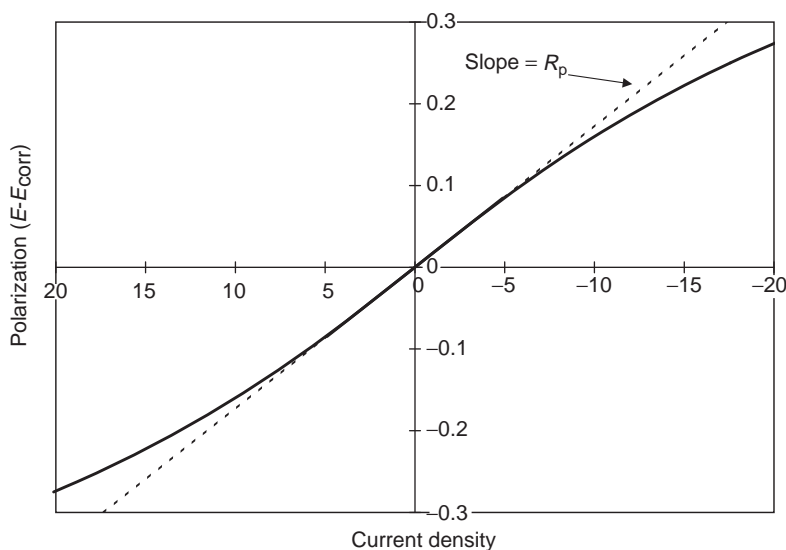


FIGURE 33.31 Hypothetical linear polarization plot.

creation of special potentiodynamic techniques to reveal the severity of localized corrosion problems. However, it can be a serious hindrance for monitoring uniform corrosion.

LINEAR POLARIZATION RESISTANCE (LPR) In this popular electrochemical technique, a small potential perturbation (typically 10–20 or even 30 mV) is applied to the sensor electrode of interest and the resulting current is measured. The ratio of the potential to current perturbations is known as the polarization resistance, which, according to Equation (33.4), is inversely proportional to the uniform corrosion rate. More specifically, the polarization resistance of a metal is defined as the slope of the potential-current density ($\Delta E/\Delta i$) curve at the free corrosion potential (Figure 33.31), yielding the polarization resistance R_p that can be itself related to the corrosion current (i_{corr}) with the help of the following Stern–Geary Eqs. (32–34):

$$R_p = \frac{B}{i_{\text{corr}}} = \frac{(\Delta E)}{(\Delta i)_{\Delta E \rightarrow 0}} \quad (33.4)$$

where R_p is the polarization resistance²; i_{corr} is the uniform corrosion current; and B is an empirical polarization resistance constant that can be related to the anodic (b_a) and cathodic (b_c) Tafel slopes with Equation (33.5).

$$B = \frac{b_a \cdot b_c}{2.3(b_a + b_c)} \quad (33.5)$$

The Tafel slopes required to perform these calculations can be either determined empirically from polarization plots such as shown in Figure 33.32 or obtained from the literature

²The accuracy of the technique can be improved by measuring the solution resistance independently and subtracting it from the apparent polarization resistance value.

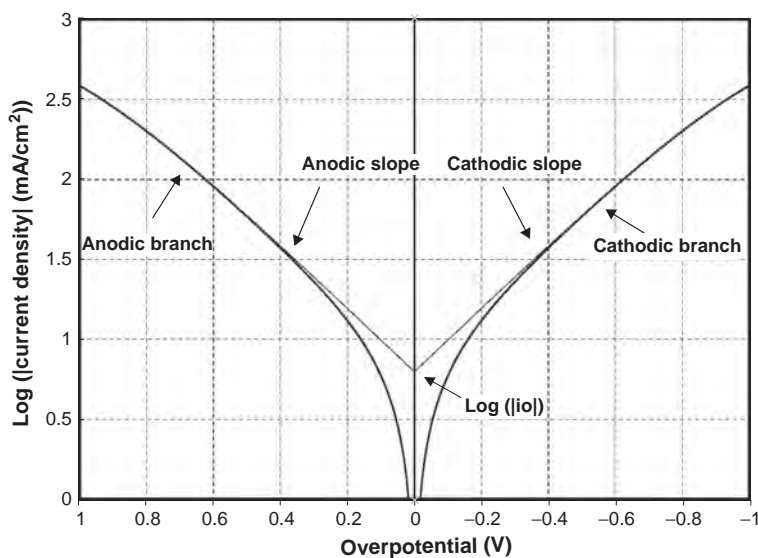


FIGURE 33.32 E of a plot of η against $\log |i|$ or Tafel plot showing the exchange current density can be obtained with the intercept.

(Grauer et al., 1982). Tafel slopes can also be determined by other techniques, such as curve fitting of polarization resistance curves (i.e., nonlinear polarization resistance), by potentiodynamic polarization scans, or by harmonic distortion analysis.

In a plant situation, it is necessary to use a probe as one of those shown in Figure 33.33 such that it enters the vessel in the area where the corrosion rate is desired (Figure 33.34).



FIGURE 33.33 Commercial sensor elements to carry out linear polarization resistance (LPR) measurements. (Courtesy of Metal Samples Company, www.metalsamples.com.)

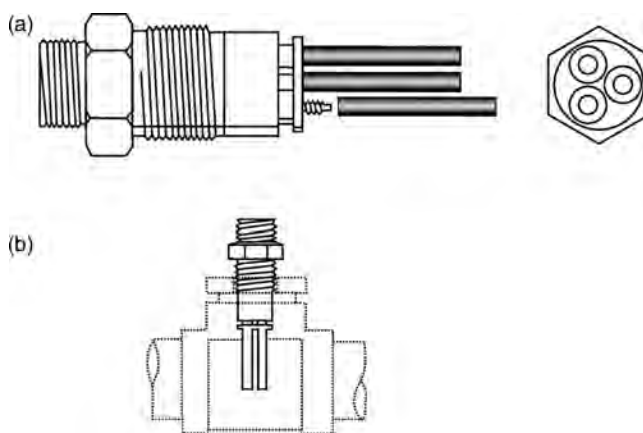


FIGURE 33.34 Typical linear polarization resistance probe (a) and probe in pipe tee (b) (Dean, 2003).

An electronic power supply polarizes the specimen from the corrosion potential. The resulting current is recorded as a measure of the corrosion rate. Several commercially available probes and analyzing systems can be directly interfaced with remote computer data-acquisition systems. Alarms can also be used to signal plant operators when high corrosion rates are experienced (Dean, 2003; Standard Guide for On-Line Monitoring of Corrosion in Plant Equipment (Electrical and Electrochemical Methods), 2001).

The LPR probes are typically a two or three electrode configuration with either flush or projecting electrodes. With a three electrode system, corrosion measurements are made on the test electrode. Because these measurements take only a few minutes, the need for a stable reference electrode is minimized. For field monitoring, the reference electrode is typically made of stainless steel or of the same alloy as that being monitored on the test electrode. The auxiliary electrode is also normally of the alloy being monitored. The proximity of the reference electrode to the test electrode governs the degree to which compensation for solution resistance is effective. With a two-electrode system, the corrosion measurement is an average of the rate for both electrodes. Both electrodes would then be made of the alloy being monitored (Techniques for Monitoring Corrosion and Related Parameters in Field Applications, 1999).

A combination of LPR and zero resistance ammeter (ZRA) measurements has been used in a special study to evaluate the rate of localized corrosion in a flowing environment by placing a large-area electrode in fast flow conditions and one small electrode in slow flow conditions in a side-stream differential flow cell schematically shown in Figure 33.35 (Yang, 1995, 1998). When the large area electrode and the small electrode were connected together through a ZRA, the large electrode became a cathode and the small electrode an anode, due to differential aeration forcing the small electrode to experience preferential corrosion. As immersion time increased, the small electrode became covered with deposits and its corrosion rate a good representation of an underdeposit or localized corrosion situation.

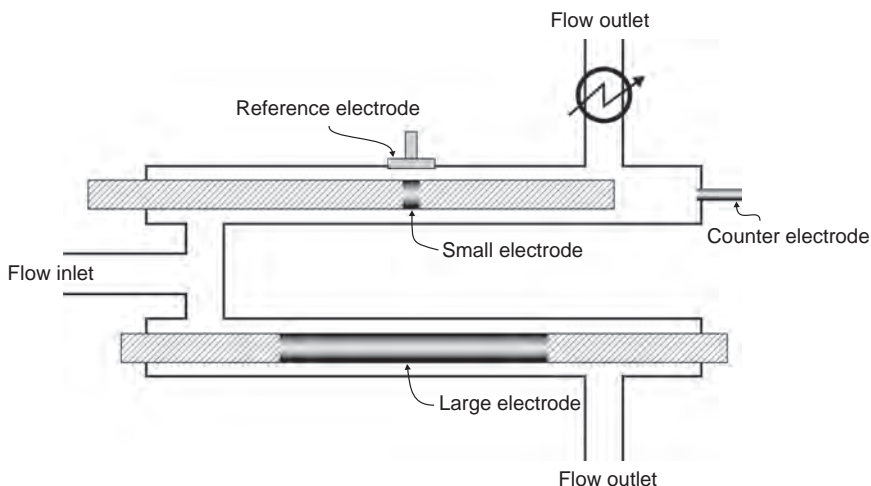


FIGURE 33.35 Schematic of differential flow cell with a fast flow electrode (FE), a slow flow electrode (SE), a reference electrode (RE), and an inert counter electrode (CE) (Yang, 1995).

There are several limitations to corrosion monitoring with LPR. The corroding environment must be an electrolyte with reasonably low-resistivity. High-resistivity electrolytes produce erroneously low corrosion rates. This need for a sufficiently conductive medium precludes the use of this technique for many applications in oil and gas, refinery, chemical, and other low-conductivity applications.

Another problem with the technique is that the vessel or pipe wall must be penetrated, and this involves concerns regarding leaks, personnel safety, and other problems. The ability to use direct wiring from the probe location to a remote control room is desirable, but the installation of these wiring systems is costly. In addition, LPR measurements do not provide information on localized corrosion, such as pitting and SCC and since the corrosion-rate values obtained with LPR are at best approximate, the method is best suited for use during periods when substantial corrosion-rate changes may occur (Dean, 2003; Standard Guide for On-Line Monitoring of Corrosion in Plant Equipment (Electrical and Electrochemical Methods), 2001).

ZERO RESISTANCE AMMETRY With this electrochemical technique galvanic currents between dissimilar electrode materials are measured with a zero resistance ammeter.³ The design of dissimilarities between sensor elements may be made to target a feature of interest in the system being monitored (e.g., different compositions, heat treatments, stress levels, or surface conditions). The technique may also be applied to nominally identical electrodes in order to reveal changes occurring in the corrosive environment and thus serve as an indicator of changing corrosion rates.

The main principle of the technique is that differences in the electrochemical behavior of two electrodes exposed to a process stream give rise to differences in the redox potential at these electrodes. Once the two electrodes are externally electrically connected, the

³ A zero resistance ammeter is a current to voltage converter that produces a voltage output proportional to the current flowing between its two input terminals while imposing a “zero” voltage drop to the external circuit.

noble electrode becomes predominantly cathodic, while the more active electrode becomes predominantly anodic and sacrificial. When the anodic reaction is relatively stable the galvanic current monitors the response of the cathodic reaction to the process stream conditions. When the cathodic reaction is stable, it monitors the response of the anodic reaction to process fluctuations (Techniques for Monitoring Corrosion and Related Parameters in Field Applications, 1999).

This technique has been found particularly useful to study depolarization effects of the cathode of a galvanic pair of electrodes to obtain feedback of low levels of dissolved gases, particularly oxygen, or the presence of bacteria, which depolarize the cathode of the galvanic pair and increase the coupling current. When used for detection of low levels of oxygen, other dissolved gases may interfere. Calibration against a dissolved oxygen meter is usually required if quantitative values are needed.

The ZRA method can provide a quantitative measure if the number of influencing factors is limited and preferably verifiable through other means. For monitoring the effect of dissolved gases, the conversion of the signal to a gas concentration level is not accurate. When other factors play a role as well, for example, during the formation of biofilm compounds or in the presence of inhibitors, the method cannot really provide a quantitative indication of any kind.

Additionally, results from galvanic probes do not always reflect the actual galvanic corrosion rates, because galvanic corrosion depends on the relative areas and specific geometries of the components, which can easily vary between a probe design and the actual plant components being monitored. The ZRA method cannot distinguish between either activation of the anodic, or the cathodic reaction. For example, an increase in the measured current can result from cathodic activation by increased dissolved oxygen, from anodic activation by increased bacterial activity, or by a combination of these. Separate analysis is sometimes performed if it is necessary to distinguish between these electrochemical components.

POTENTIODYNAMIC—GALVANODYNAMIC POLARIZATION In this technique, a three electrode corrosion probe is used to polarize the electrode that serves as the sensing element. The current or potential response is measured as the potential (potentiodynamic) or current (galvanodynamic) is shifted away from the free corrosion potential. The basic difference with the LPR technique is that the applied polarization can be of several hundred millivolts (Figure 33.36). While the technique is used quite commonly in a laboratory environment, it is used only occasionally in the field mostly to estimate the anodic and cathodic Tafel slopes for systems on which the basic corrosion rate theory is based (Figure 33.37). The formation of passive films and the onset of pitting corrosion can also be identified at characteristic potentials, which can assist in assessing the overall corrosion risk.

Potentiodynamic polarization is generally used in aqueous systems while galvanodynamic polarization has been used in systems containing oil, so that current density is controlled. One of the most common uses of the technique is to determine whether crevice corrosion or pitting is a problem and whether general corrosion can be estimated by another technique. It is also often used to estimate the relative susceptibility of various materials to localized corrosion in the process stream (Techniques for Monitoring Corrosion and Related Parameters in Field Applications, 1999).

Ideally, each point on the current–potential graph should be made by allowing a polarization time of tens of seconds to several minutes to permit the complete charging of the double-layer capacitance associated with the metal–fluid interface. Such long-time

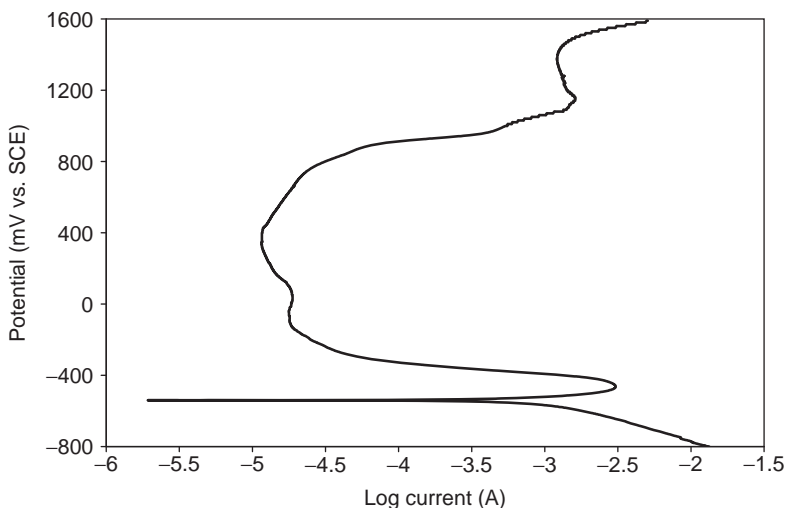


FIGURE 33.36 Typical anodic polarization plot for S43000 steel in a 0.05 M H_2SO_4 solution.

intervals also permit the electrode surface to oxidize or reduce all of the surface deposits and to polarize to the applied potential. In practice, the applied current or applied potential, whichever is the controlling variable, is changed continuously in an analog form at some preset rate (potentiodynamic or galvanodynamic) or in a digital form of small discrete steps at some preset rate (potential staircase or galvanic staircase).

In either case, the faster the rate of change of the applied signal, the greater is the lag of the measured signal behind the true steady-state values desired. This compromise

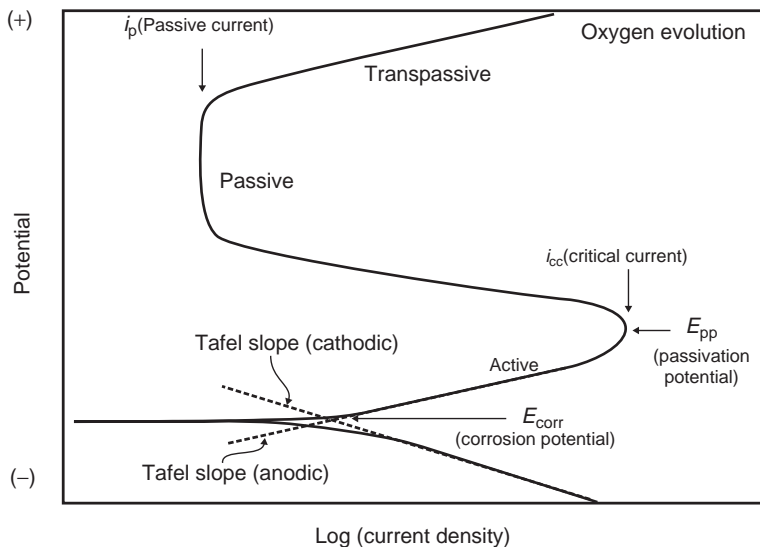


FIGURE 33.37 Generalized polarization diagram showing the various potential regions of a passivable metal and the Tafel extrapolation lines.

between a practical and reasonable time to complete a scan and the degree of lag from measured-to-ideal response is a key question with this technique, particularly so in the field where the tested environment is continually changing and cannot usually be controlled as it can be in a laboratory.

A potentiodynamic polarization variant is cyclic voltammetry which involves sweeping the potential in a positive direction until a predetermined value of current or potential is reached, then the scan is reversed toward more negative values until the original value of potential is reached. In some cases, this scan is done repeatedly to determine changes in the current–potential curve produced with scanning.

Another variation of potentiodynamic polarization is the potentiostaircase method. This refers to a technique for polarizing an electrode in a series of potential steps where the time spent at each potential is constant, while the current is often allowed to stabilize prior to changing the potential to the next step. The step increase may be small, in which case, the technique resembles a potentiodynamic curve (Van Orden, 1998).

Electrochemical potentiodynamic reactivation (EPR) is another polarization method that evaluates the degree of sensitization of stainless steels, such as S30400 and S30403 steels. This method uses a potentiodynamic sweep over a range of potentials from passive to active (called reactivation).

However, probably the most popular variant is the cyclic polarization test. This test is often used to evaluate the pitting susceptibility of a material. The potential is swept in a single cycle or slightly less than one cycle usually starting the scan at the corrosion potential. The voltage is first increased in the anodic or noble direction (forward scan). The voltage scan direction is reversed at some chosen current or voltage toward the cathodic or active direction (backward or reverse scan) and terminated at another chosen voltage. The presence of a hysteresis between the currents measured in the forward and backward scans is believed to indicate pitting, while the size of the hysteresis loop itself is often related to the amount of pitting.

This technique has been especially useful to assess localized corrosion for passivating alloys such as S31600 stainless steel, nickel-based alloys containing chromium, and other alloys, such as titanium and zirconium. Though the generation of the polarization scan is simple, its interpretation can be difficult (Silverman, 1998).

In the following example, the polarization scans were generated after 1 and 4 days of exposure to a chemical product maintained at 49°C. The goal of these tests was to examine if S31600 steel could be used for short-term storage of a 50% commercial organic acid solution (aminotrimethylene phosphonic acid) in water. A small amount of chloride ion (1%) was also potentially present in this acidic chemical.

In this example, the potential scan rate was 0.5 mV/s and the scan direction was reversed at 0.1 mA/cm². Coupon immersion tests were run in the same environment for 840 h. The S31600 steel specimens were exposed to the liquid, at the vapor–liquid interface, and in the vapor phase. The reason for the three exposures was that in most storage situations, the containment vessel would be exposed to a vapor–liquid interface and a vapor phase at least part of the time. Corrosion in these regions can be very different from liquid exposures. The specimens were also fitted with artificial crevice formers.

Figure 33.38 shows the polarization scan generated after 1 day and Figure 33.39 shows the polarization scan generated after 4 days of exposure. The important parameters considered were the position of the “anodic-to-cathodic” transition relative to the corrosion potential, the existence of the repassivation potential and its value relative to the corrosion potential, the existence of the pitting potential and its value relative to the

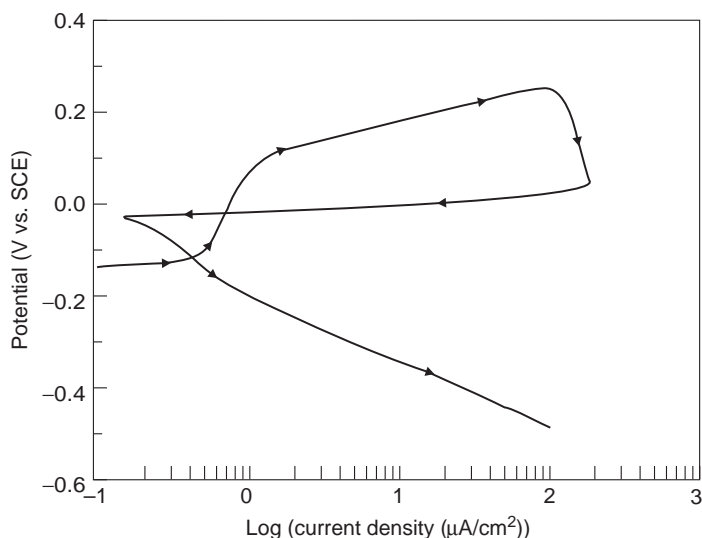


FIGURE 33.38 Polarization scan for S31600 steel in 50% aminotrimethylene phosphonic acid after one day of exposure (the arrow indicates scanning direction).

corrosion potential, and the hysteresis (positive or negative). The interpretation of the results is summarized in Table 33.6.

The presence of the negative hysteresis would typically suggest that localized corrosion may be a problem depending on the value of the corrosion potential relative to the

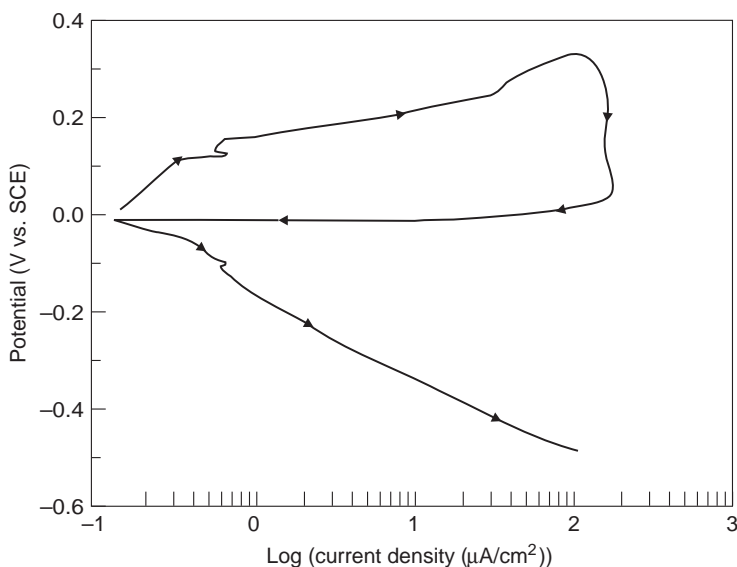


FIGURE 33.39 Polarization scan for S31600 steel in 50% aminotrimethylene phosphonic acid after four days of exposure (the arrow indicates scanning direction).

TABLE 33.6 Features and Values Used to Interpret Figures 33.38 and 33.39

Feature	Value in Figure 33.38	Value in Figure 33.39
Repassivation potential–corrosion potential	0.12 V	0.0 V
Pitting potential–corrosion potential	0.22 V	0.12 V
Potential of anodic-to-cathodic transition–corrosion potential	0.12 V	0.0 V
Hysteresis	Negative	Negative
Active-to-passive transition	No	No

characteristic potentials present in these polarization plots. After the first day of exposure, pitting was not expected to be a problem because the pitting potential was far away from the corrosion potential. The currents generated were much higher than those normally associated with S31600 steel in a passive state. These observations suggested that there was a risk of initiation of corrosion, particularly in localized areas where the pH can decrease drastically (Silverman, 1998).

After 4 days, the risk of localized corrosion increased. At this time, the repassivation potential and the potential of the change from anodic-to-cathodic current were equal to the corrosion potential. The pitting potential was only ~ 0.1 V more noble than the corrosion potential and the hysteresis still negative. The risk of pitting had increased to become a concern.

Coupon immersion tests confirmed the long-term predictions. Slight attack was found under the artificial crevice formers in the complete liquid exposure.

The practical conclusion of this in-service study was that, since localized corrosion often takes time to develop, a few days of exposure to this chemical product could be acceptable. However, it was recommended to avoid long-term exposure since both pitting and crevice corrosion would be expected for longer exposure periods.

While this example was an interesting study, the potentiodynamic techniques are generally not considered for on-line or real-time monitoring because the electrodes have to be replaced after only one or two test runs. The high anodic potentials used in the tests simply corrode the sensing elements and permanently changes their surface. Correcting the ohmic(IR) drop due to the solution resistance may also be particularly important with this technique because of the relatively high currents used compared with other electrochemical techniques. The IR induced error can be much smaller in highly conductive environments.

ELECTROCHEMICAL IMPEDANCE SPECTROSCOPY With electrochemical impedance spectroscopy (EIS), the sensing element is polarized by the application of an alternating potential that in turn produces an alternating current response. For corrosion monitoring, the frequency range of the applied ac polarization is typically between 0.1 and 100 kHz with a polarization level within 10 mV of the corrosion potential. Full frequency scans provide phase shift information that can be utilized in combination with equivalent circuit models to obtain useful information on the system complex interface.

Amongst the numerous equivalent circuits that have been proposed to model electrochemical interfaces only a few really apply to a freely corroding system. The circuit shown in Figure 33.40a is the simplest equivalent circuit that can describe a metal–

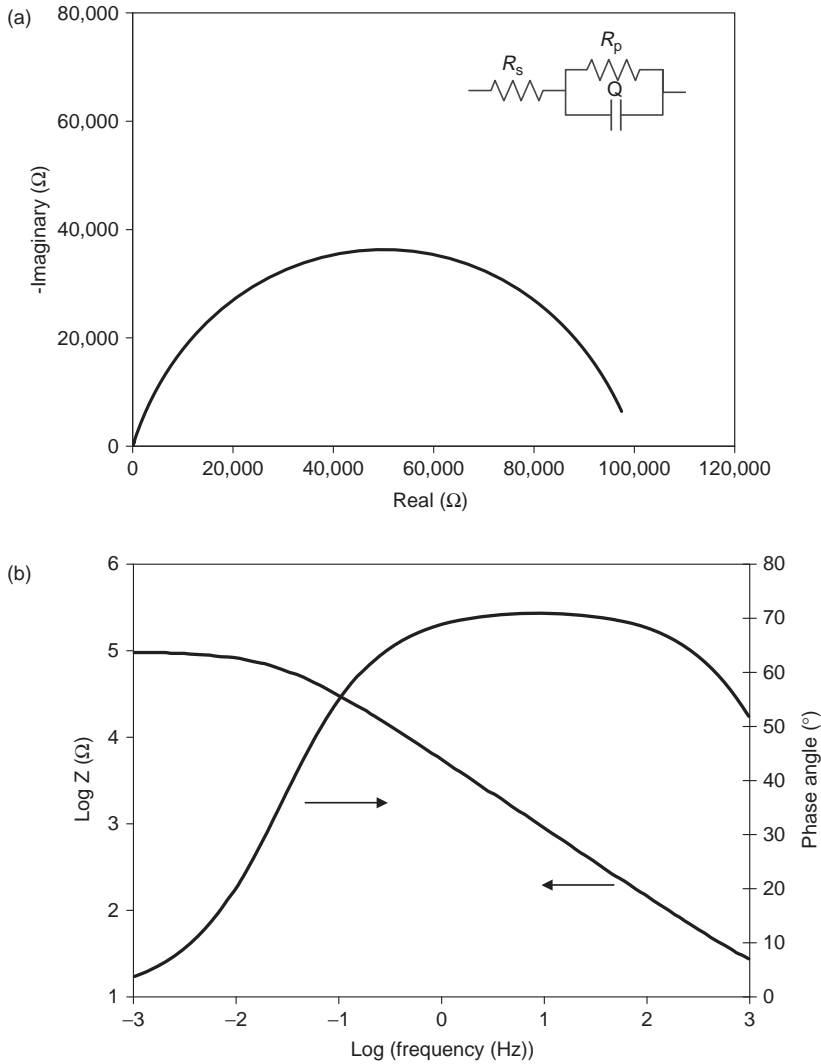


FIGURE 33.40 (a) Complex plane and RC model circuit with simulated data, where $R_s = 10 \Omega$, $R_p = 100 \text{ k}\Omega$, and Q decomposes into $C_{dl} = 40 \mu\text{F}$ and $n = 0.8$. (b) Bode plot corresponding to the same simulated data.

electrolyte interface. Its behavior is described by Equation (33.6).

$$Z(\omega) = R_s + \frac{R_p}{1 + (j\omega R_p C_{dl})^\beta} \quad (33.6)$$

where R_s is the solution resistance, R_p is the polarization resistance, ω is the ac polarization frequency, and C_{dl} is the double layer capacitance.

The term Q in Figure 33.40a describes the “leaky capacitor” behavior corresponding to the presence of a constant phase element (CPE) (Boukamp, 1989). Figure 33.40a also illustrates the complex plane presentation of this EIS model circuit in which $R_s = 10\ \Omega$, $R_p = 100\ \text{k}\Omega$, and Q is a function of $C_{dl} = 40\ \mu\text{F}$ and $n = 0.8$. Figure 33.40b shows how the same data would appear in a Bode plot format.

The high-frequency response is used to determine the component of solution resistance (R_s) included in the measurement. The polarization resistance (R_p) can then be determined by subtracting the R_s value from the low-frequency measurement. To convert the polarization resistance into a corrosion rate involves an empirical measurement of the Tafel slopes that have to be determined by other techniques, such as potentiodynamic polarization and harmonic distortion analysis, or again obtained from the literature (Grauer et al., 1982).

The measurement cycle time depends on the frequency range used, especially the low frequencies. A single frequency cycle at 1 mHz, for example, takes 15 min. A high-to-low frequency scan going to such a low frequency would take $>2\ \text{h}$. In order to make routine corrosion monitoring with EIS certain simplifications are needed to maximize the use of high frequency data and drastically shorten the measurement time. It is also important to simplify the data processing and analysis to make the technique user friendly for field corrosion monitoring. However, the need for a field instrument that can be easily deployed has always been an impediment for on-line corrosion monitoring with EIS.

In order to simplify the analysis of field EIS results, a method was developed that consists of finding the geometric center of an arc formed by three successive data points on a complex impedance diagram (Figure 33.41) (Roberge and Sastri, 1994; Roberge, 1992). This technique was designed as an improvement over the two point method based on the comparison of high- and low-frequency data points for which the impedance would be

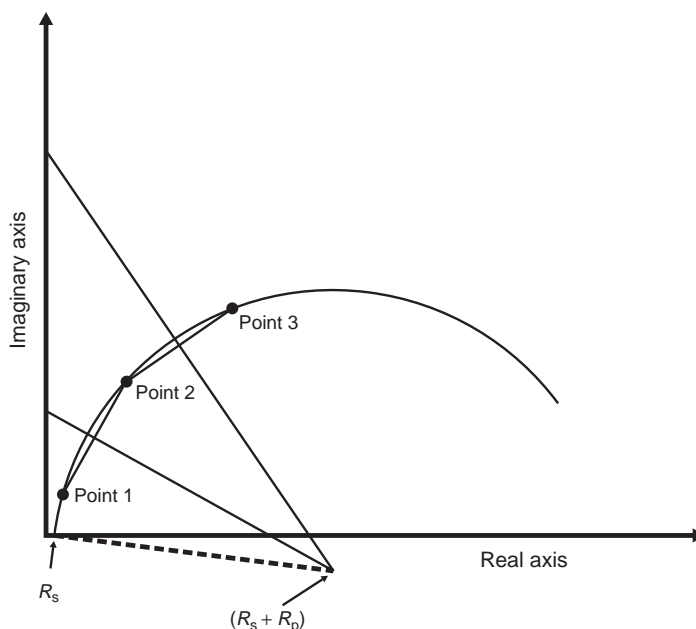


FIGURE 33.41 Schematic of the extrapolation method to obtain the polarization resistance from EIS data.

proportional to the R_s at the high frequency point and the summation of R_s and R_p at the low frequency point (Haruyama and Tsuru, 1981). In real-world situations, one difficult assumption to satisfy with the two point method is that data points should contain negligible imaginary components (i.e., 0 phase shift), a condition usually hard to achieve in a meaningful manner at low measuring frequencies.

The three point analysis technique was further developed by permuting the data points involved in the projection of centers in order to obtain a population of projected centers. This improvement has permitted to automate the data analysis while providing some information concerning the adherence of the results with the RC behavior described by Equation (33.6), what is assumed for the evaluation of the parameters associated with uniform corrosion. The technique was extensively tested in a laboratory environment before being applied in field tests with the same laboratory equipment (Roberge and Sastri, 1994).

Very recently, a full-spectrum relatively low cost EIS corrosion monitoring system has been developed, which is wireless, small (5-cm diameter, 1.2-cm height), requires nominally 10-mW power during its 200-s measurement period, and has an electronic identifier, which allows for a single data logger to monitor multiple devices in the same general vicinity (Figure 33.2a and b) (Davis et al., 2005).

The wireless EIS sensor determines the impedance at 15–20 independent frequencies, by measuring amplitude and phase at each frequency (Figure 33.42). It computes corrosion rate, conductivity and coating impedance, and transmits the result wirelessly to a data logger. The miniature and wireless features make it suitable for embedding in concrete or placing in hidden and inaccessible locations, for example, in HVAC systems. Its minimal power consuming aspect lends itself useful for long-term monitoring of coating integrity.

The miniature EIS system has been tested in various environments, namely, concrete, water, and under coatings (Figure 33.43a and b). In addition, it was also tested against a

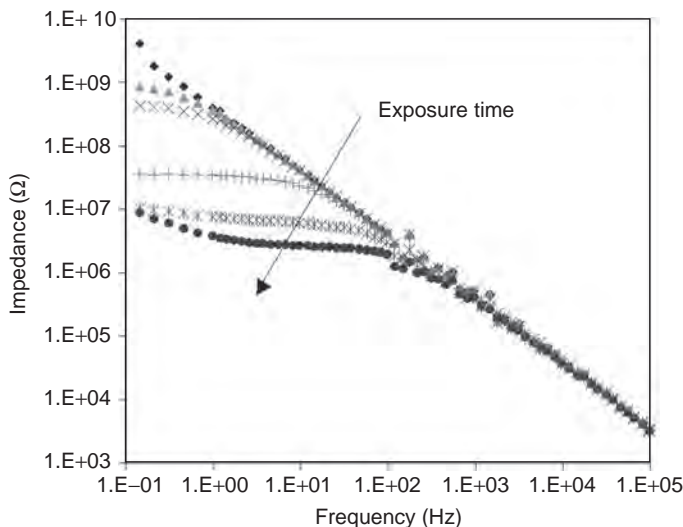


FIGURE 33.42 Magnitude of impedance of a coating versus frequency. Low-frequency impedance values show good correlation with long-term exposure behavior. (Courtesy of Guy D. Davis, DACCO SCI, Inc.)

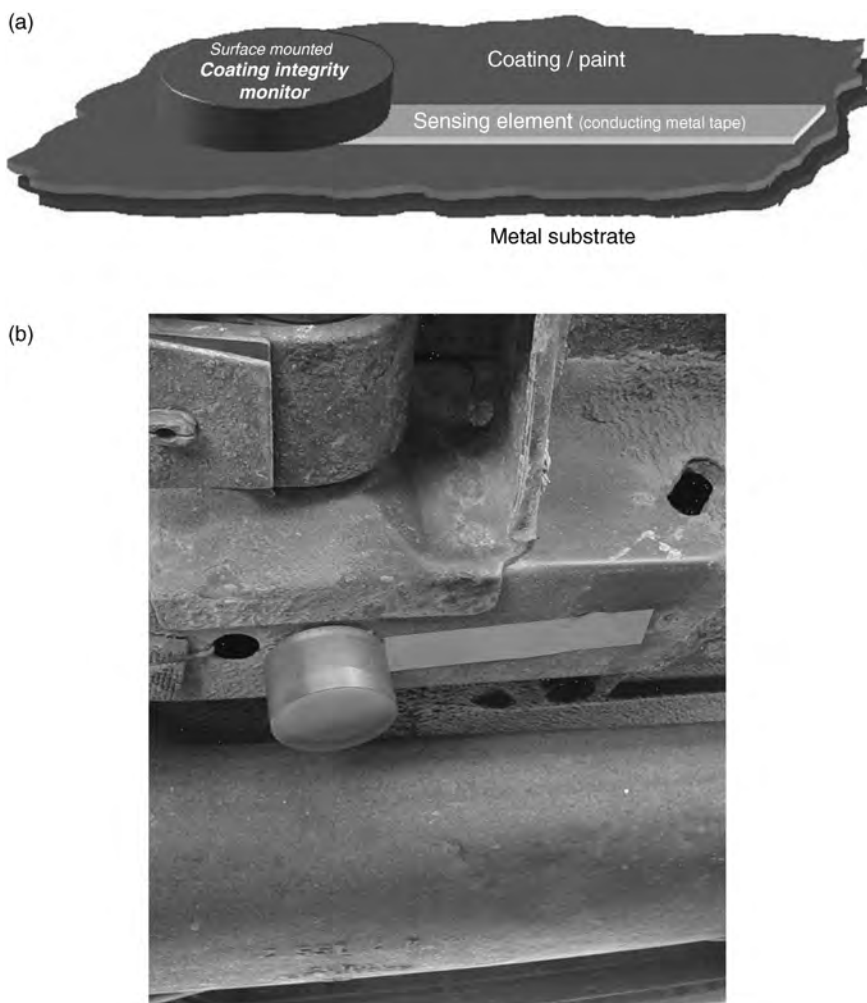


FIGURE 33.43 (a) Schematic of the Coating Health Monitor (CHM) with a tape sensing element mounted on a coated metal and (b) actual sensor tape electrode and electronics housing mounted on the frame of a commercial vehicle. (Courtesy of Guy D. Davis, DACC SCI, Inc.)

simple resistor capacitor circuit and validated with two different commercial instruments. The miniature EIS system successfully monitored the corrosivity of water contaminated with chloride, concrete in contact with water containing salt, and health and integrity of coatings on metals. These results correlated well with data obtained using conventional bench-top EIS and LPR instruments. However, unlike the conventional instruments, the wireless EIS sensor is small, requires very little power to operate and can be hermetically sealed. Therefore, it can be embedded in structures and immersed in liquids without much concern about drawing wires and cables.

HARMONIC DISTORTION ANALYSIS With this technique, a low-frequency sinusoidal potential is applied to a three-electrode measurement system, and the resulting current is measured.

As the corrosion process is nonlinear in nature, a potential perturbation by one or more sine waves should generate responses at more frequencies than the frequencies of the applied signal. Current responses can be measured at zero, harmonic, and intermodulation frequencies.

Measuring the dc at frequency “zero” is called the Faraday rectification (FR) technique. The FR technique can be used for corrosion rate measurements if at least one of the Tafel parameters is known. The corrosion rate and both Tafel parameters can be obtained with one measurement by analyzing the harmonic frequencies. The speed with which the Tafel slopes can be determined (typically <1 min) is a particular attraction of the technique (Bosch et al., 2001).

With harmonic distortion analysis (HDA), a single, low-frequency and low-distortion, sinusoidal voltage is applied to the corroding interface. As a quality check, three different frequencies are used to verify the repeatability of the technique. The amplitude is in the range of 10–30-mV peak to peak. The frequency used is typically 0.1–10 Hz. Other frequencies can be used depending on the corrosion process.

The theoretical analysis for the computation of Tafel slopes and corrosion current makes no assumptions about solution resistance effects and measurements cannot be performed unless the system is free of significant electrical noise in the frequency range of the applied measurement potential, or one of its harmonic frequencies (Techniques for Monitoring Corrosion and Related Parameters in Field Applications, 1999).

In principle, analyzing the primary frequency and the harmonics makes it possible to extract all the information to calculate Tafel slopes and corrosion rate. In practice, this has been used in a very limited number of applications to date. A serious limitation of the existing nonlinear techniques is that their application has been restricted mainly to activation controlled corrosion processes, for example, corrosion rate measurements in acid media with and without inhibitors.

Besides FR, variations of harmonic distortion analysis developed have been called nonlinear EIS (NLEIS), harmonic analysis (HA), harmonic impedance spectroscopy (HIS), and electrochemical frequency modulation (EFM). These techniques all analyze nonlinear response of a corrosion system after perturbation with a mono -or biharmonic signal, and on this basis allow extraction of the required kinetic parameters of the corrosion process. Except for the FR method, all other HAD techniques allow the simultaneous determination of corrosion current and both Tafel coefficients (Jankowski, 2003).

ELECTROCHEMICAL NOISE (EN) ANALYSIS Fluctuations of potential or current of a corroding metallic specimen are a well known and easily observable phenomenon. Electrochemical noise analysis (ENA) as a corrosion tool has increased steadily since Iverson's paper in 1968 (Iverson, 1968) and the number of industrial applications of electrochemical noise monitoring has grown significantly in recent years. The study of EN has been found to be uniquely appropriate for monitoring the onset of events leading to localized corrosion and understanding the chronology of the initial events typical of this type of corrosion.

The EN technique differs in many ways from the other electrochemical techniques described so far. One important difference is that ENA does not require that the sensing element be polarized in order to generate a signal. However, it is also possible to measure current noise under an applied potential, or measure potential noise under an applied current. The potential and current between freely corroding electrodes (in many cases <1 μ V and <1 nA) are measured with sensitive instrumentation. A measurement frequency of 1 Hz is usually appropriate to provide meaningful data. For simultaneous measurement of

electrochemical potential and current noise, a three-electrode sensor is required. In field corrosion monitoring, the three sensor elements are usually made of the same material.

While the measurement of electrochemical noise is relatively straightforward, the data analysis can be complex and inconclusive. Even if ENA was first applied in field corrosion monitoring in the late 1960s, an understanding of the method of analysis is still evolving, partly because the technique has been used to look at several types of corrosion. The relationships between potential and current noise are inherently complex to analyze quantitatively because the naturally occurring fluctuations do not have controlled frequencies as are applied, for example, in EIS. For these reasons, much of the investigations with EN have been centered on frequency analysis of the data. There are still varying conclusions about the accuracy and effectiveness of the technique in its own right (Techniques for Monitoring Corrosion and Related Parameters in Field Applications, 1999).

Electrochemical Noise Analysis The analysis of EN, begun a few decades ago, has only been recently introduced as a credible corrosion monitoring technique. In this context, the pioneering work of Eden *et al.* has been instrumental in introducing the idea of a corrosion cell with two working electrodes (WE), where both current and voltage fluctuations can be measured (Eden *et al.*, 1986). The remaining question was how to establish the data interpretation on a firm basis. Because EN measurements on a single corroding electrode are not sufficient to evaluate corrosion rates, most applications in the field are based on the use of cells with two identical electrodes (same material, same size, same surface preparation), connected through a zero-resistance ammeter (ZRA) so as to have both working electrodes set at a common corrosion potential (Huet *et al.*, 2001).

There are many cases, particularly in field applications where the use of low-noise reference electrodes commonly used in laboratories would be impractical. In these cases, one can use a third electrode made of material similar to the other two working electrodes. Obviously, such a reference electrode would contribute to the noise of the system. It turns out, however, that in such an arrangement the noise impedance Z_n is equal to a fraction of the total noise impedance (Z), that is, $\sqrt{3}|Z|$, so that a simple correction is sufficient to correct the problem. However, a more serious concern is that the noise signals depend on the three electrodes having the same impedance and contributing the same noise. As every corrosion worker knows, initially identical electrodes tend to diverge in behavior with time. Experience has shown that this is particularly troublesome in the case of localized corrosion for which it could introduce significant errors difficult to correct.

There are basically three categories of ENA: visual examination, sequence-independent methods that treat the collection of voltage or current values without regard to their position in the sequence of readings (moments, mean, variance, standard deviation, skewness, and kurtosis), and those that take the sequence into account (autocorrelation, power spectra, fractal analysis, stochastic process analysis) (Cottis *et al.*, 2001).

Visual examination of the time record trace can give indications as to the type of corrosion processes that are occurring. The following example illustrates how a simple examination of EN measurements could reveal the corrosivity of various points of an industrial gas scrubbing system where highly corrosive thin-film electrolytes may form (Roberge, 2000). These conditions arise when gas streams are cooled to a temperature below the dewpoint. The resulting thin electrolyte layer (moisture) is often highly concentrated in corrosive species.

The corrosion probe used in this example is illustrated in Figures 33.44 and 33.45. A retractable probe with flexible depth was selected, in order to mount the sensor surface

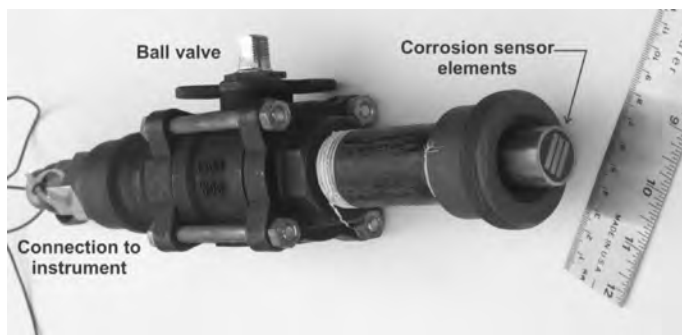


FIGURE 33.44 Corrosion sensor and access fitting used for thin-film corrosion monitoring. (Courtesy Kingston Technical Software.)

flush with the internal scrubber wall surface. The close spacing of the carbon steel sensor elements was designed to work with a discontinuous thin surface electrolyte film. This corrosion sensor was connected to a handheld multichannel data recorder by shielded multi-strand cabling Figure 33.1. As the ducting of the gas scrubbing tower was heavily insulated, no special precautions were taken to cool the corrosion sensor surface.

Potential noise and current signals recorded during the first hour of exposure at the conical base of the gas scrubbing tower are presented in Figure 33.46. According to the operational history of the plant, condensate had a tendency to accumulate at this location where highly corrosive conditions had been noted. The high levels of potential noise and current noise in Figure 33.46 are indicative of a massive pitting attack that is consistent with the operational experience. Note that the current noise is actually off-scale for most of the monitoring period, in excess of 10 mA. The high corrosivity indicated by the



FIGURE 33.45 Close-up of corrosion sensing elements used for thin-film corrosion monitoring. (Courtesy Kingston Technical Software.)

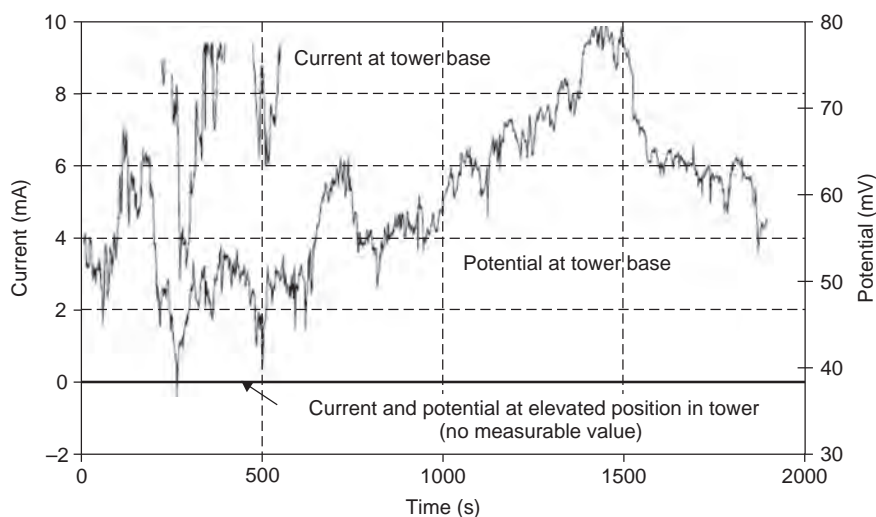


FIGURE 33.46 Potential and current noise records at two locations in a gas scrubbing tower.

electrochemical noise data from this sensor location was confirmed by direct evidence of severe pitting attack on the sensor elements, revealed by scanning electron microscopy (Figure 33.47). In contrast, both current and voltage signals remained relatively constantly small at a position higher up in the tower, where the sensor surface remained mostly dry.

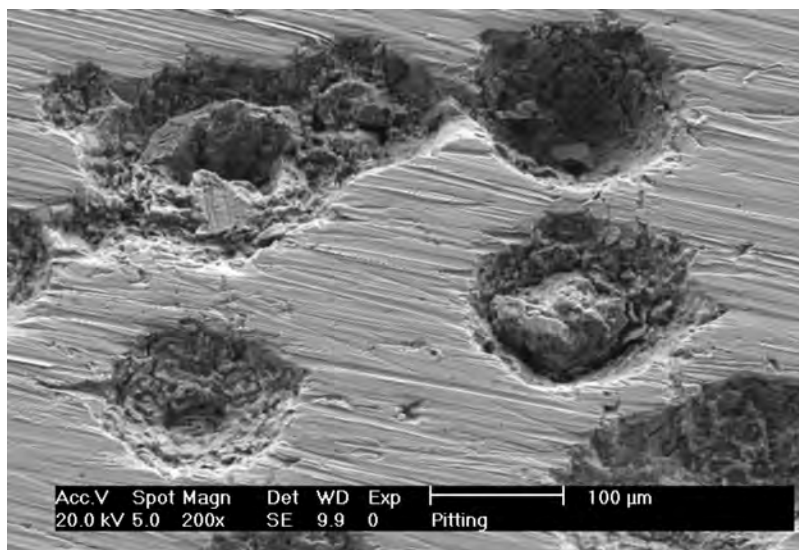


FIGURE 33.47 Scanning electron microscope image of a sensor element surface after exposure at the base of the scrubbing tower clearly showing corrosion pits.

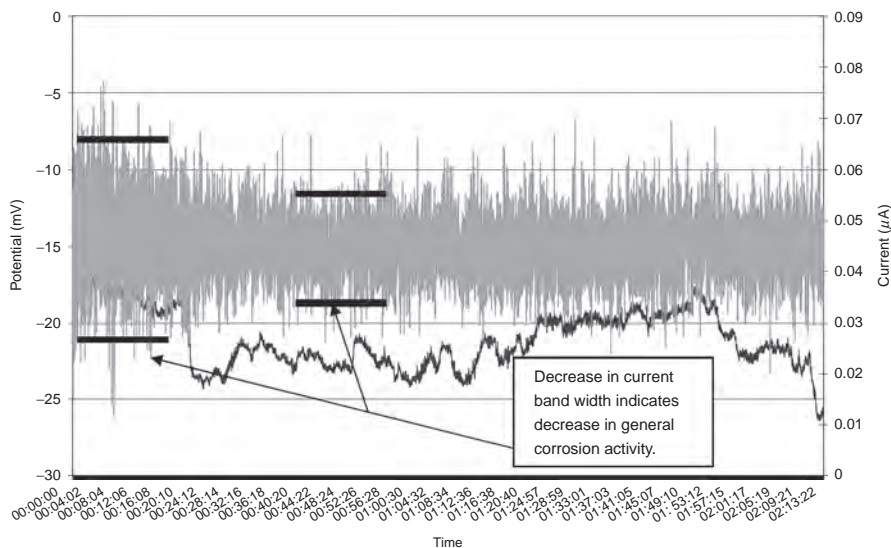


FIGURE 33.48 Electrochemical current noise (large band) and potential noise (lower signal) in a debutanizer overhead piping obtained with the Concerto VT noise system. (Courtesy of CAPCIS Ltd.)

An improvement over this simple analysis is commonly practiced in industry by tracking the width of the potential and current signals as an indication of corrosion activity in the system being monitored. Figure 33.48 illustrates how the decrease in the current band obtained with a monitoring system was interpreted as a reduction in general corrosion activity in a debutanizer overhead piping where the interaction between operational changes and the corrosion mechanism were being investigated.

Most ENA techniques have been developed to relate sometimes subtle features in the noise data records to changes in the corrosivity of a system and particularly on how these changes affect localized corrosion. Three techniques have also been proposed for obtaining the polarization resistance as a measure of uniform corrosion rate if the Tafel constants of the system are known or measured.

The noise resistance (R_n), a sequence-independent method, is the ratio of the standard deviation of potential to the standard deviation of current. The impedance resistance (R_{imp}), a sequence-dependent method, is the ratio of the square root of the potential power to the square root of the current power in the low frequency limit. The potential and current power densities are calculated from power spectrum distribution (PSD) techniques, such as fast Fourier transform (FFT) or the maximum entropy method (MEM) (Cottis and Turgoose, 1999). Finally, the EN transient resistance (R_{tran}) is the ratio of the amplitude of potential to current for single transients (Klassen and Roberge, 2002). The parameter R_{tran} provides the most exact corrosion resistance value, but the data analysis is also the most complicated since the technique hinges on finding transients amenable to a clear interpretation.

While these techniques have been extensively used in controlled laboratory experiments, the high variability of field environments combined with the availability of well established methods such as ER, LPR, or the use of coupons have created the impression that ENA is not an appropriate option for obtaining corrosion rates when uniform corrosion is the prevalent corrosion mechanism.

In contrast, localized corrosion processes, such as pitting and SCC, have characteristic transients in the time record traces that may help to distinguish between various possible types of corrosion. The modern success of ENA is in great part based on this unique property since more traditional on-line techniques are quite ineffective in this area. This success is also due to the fact that most forms of localized corrosion are much more troublesome in the management and operation of many engineering systems than uniform corrosion.

The effect of localized corrosion on EN signals tends to create a deviation from a normal distribution (Poisson distribution) as may be determined from the skewness and kurtosis of the EN signals. If techniques, such as linear polarization resistance (LPR) and harmonic distortion analysis (HDA), are also used, then independent derivation of the general corrosion current (I_{corr}) is possible. The ratio of the current noise (standard deviation of the current) to the general corrosion current may then be used as an improved indicator of localized corrosion activity (Eden, 2000).

Skewness and kurtosis are measures of the probability density of the noise signals in relation to the mean value. Skewness provides an indication of the symmetry of the distribution. A value of zero implies that the distribution is symmetrical whereas a positive skewness implies that there is a tail in the positive direction and a negative skewness a tail in the negative direction. Kurtosis is a measure of the shape of the distribution compared with the normal distribution. A kurtosis of zero implies that the distribution has a shape similar to that of the normal distribution. A positive kurtosis implies a narrower distribution, whereas a negative kurtosis implies a flatter distribution. It was found experimentally that the kurtosis of a signal generated by uniform corrosion was approximately 1.5–2 while a kurtosis value >5 would be indicative of localized corrosion (Eden, 2005).

The localization parameter is the ratio of the standard deviation of the current, σ_I , to the root-mean-square (rms) current (Cottis and Turgoose, 1999). However, the rms current is related to the mean coupling current, which, over an infinite period of time with two similar electrodes is zero. This gives the localization parameter the possibility of having the unfortunate value of infinity. The pitting factor is a modification of the localization parameter (Roberge et al., 1996). The idea is to use a measure of i_{corr} instead of rms current. If one assumes that R_p is representative of the polarization resistance then the pitting factor is equivalent to the ratio of the standard deviation of potential, σ_V , to the Stern–Geary constant (B).

Figures 33.49 and 33.50 illustrate how the corrosion pitting indication and associated system instability were tracked during the debutanizer overhead piping evaluation project mentioned earlier.

The shot-noise charge is a measure of the average charge of a transient based on shot-noise analysis (Cottis and Turgoose, 1999). This is the ratio of $\sigma_I \sigma_V / bB$, where b is the bandwidth of measurement (s^{-1}). The bandwidth is the frequency range from essentially zero to the Nyquist frequency, $1/(2\Delta t)$, where Δt is the sampling period.

Certain characteristics of localized corrosion have been interpreted by analyzing EN data in the frequency domain. The slope of the linear region of the PSD plot of current or potential versus frequency has been interpreted as being related to the shape, amplitude and frequency of elementary transients (Cottis and Turgoose, 1999). The expected slopes can be in the range of 0 to -4 with zero for white noise processes, -0.5 for diffusion controlled processes, -1 for processes having a Gaussian distribution and less than -2 for processes with fluctuations (Eden, 2005). However, the exact value of this slope corresponding to localized corrosion behavior may differ significantly between systems and environments.

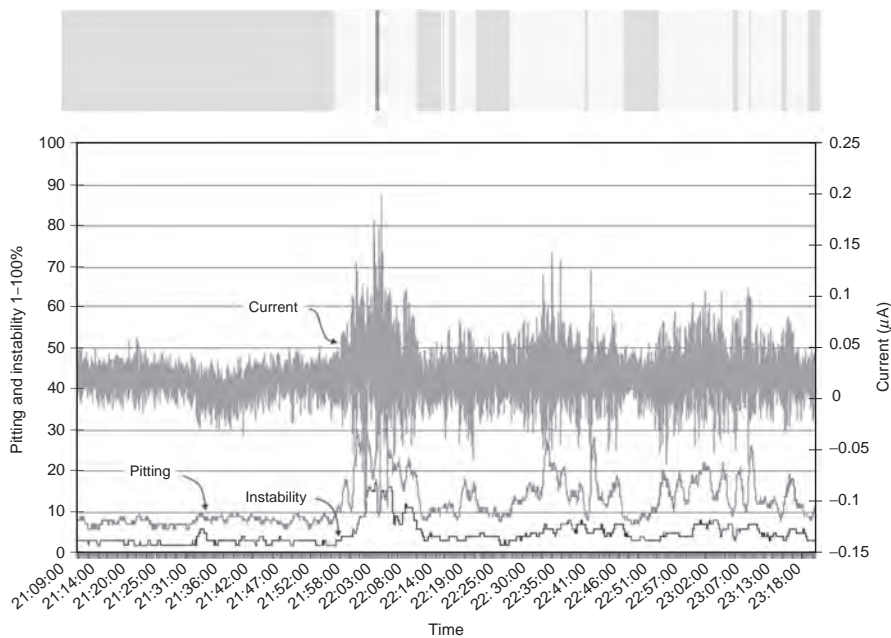


FIGURE 33.49 Increased pitting activity at end of period from large current transients observed in a debutanizer overhead piping obtained with the Concerto VT noise system. (Courtesy of CAPCIS Ltd.)

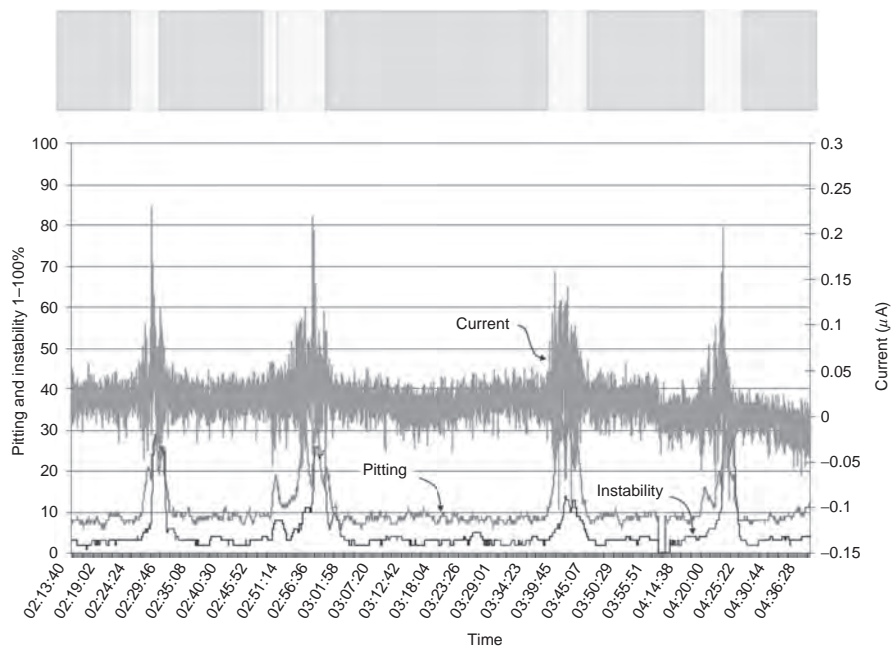


FIGURE 33.50 Moderate pitting events on fairly regular intervals observed in a debutanizer overhead piping with the Concerto VT noise system. (Courtesy of CAPCIS Ltd.)

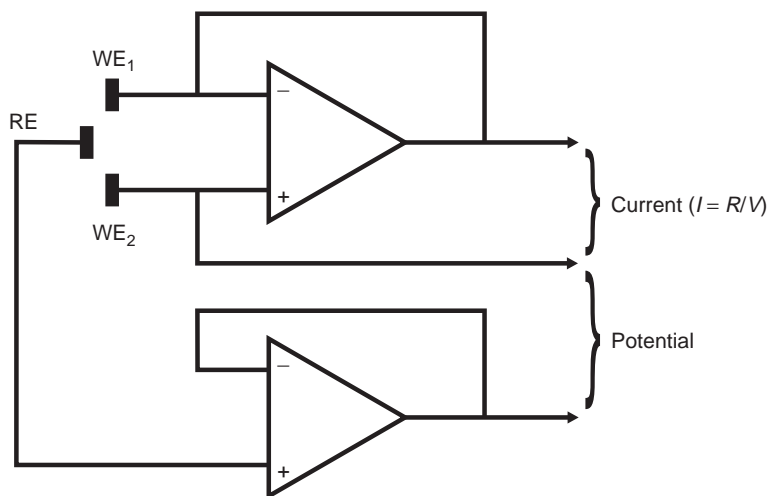


FIGURE 33.51 Simultaneous current and potential monitoring using three identical electrodes. (Adapted from Eden (2005).)

Probe Design and Signal Analysis An appreciable level of effort has been directed towards developing electrode arrangements for monitoring localized corrosion with EN, in particular pitting, such as the use of both very small ($\leq 1 \text{ mm}^2$) or very large ($\geq 100 \text{ cm}^2$) electrodes, as well as electrode arrays. A great variety of instrumentation approaches have also been used (Eden, 2005):

- Three identical electrodes, configured as two coupled via a ZRA and one pseudoreference, with simultaneous (Figure 33.51) or sequential (Figure 33.52) measurements of current and potential.
- Three identical electrodes configured in potentiostatic mode (working, reference and auxiliary) with the reference potential set at 0 V (Figure 33.53), with simultaneous measurement of current and potential. This particular arrangement is useful in that the potentiostatic control may be used to perform other controlled polarization measurements such as LPR or HDA.
- Multiple electrode arrays of the test material using scanning techniques to measure current and/or potential between individual or combinations of elements in order to identify locally anodic behavior on one or more of the electrodes (Figure 33.10) (Yang and Sridhar, 2003).
- Single electrode probe consisting of a working electrode with current (Figure 33.54) or potential (Figure 33.55) measured with respect to the structure of interest. As with the three electrode arrangement this may also be used to evaluate polarization response.

Higher levels of electrochemical corrosion activity are generally associated with higher EN levels. Certain electrochemical phenomena, such as the breakdown of passivity during pit initiation show up distinct noise “signatures”, which can be exploited for corrosion monitoring purposes. Pit initiation and growth, for example, can be detected with EN measurements long before it becomes evident by visual examination.

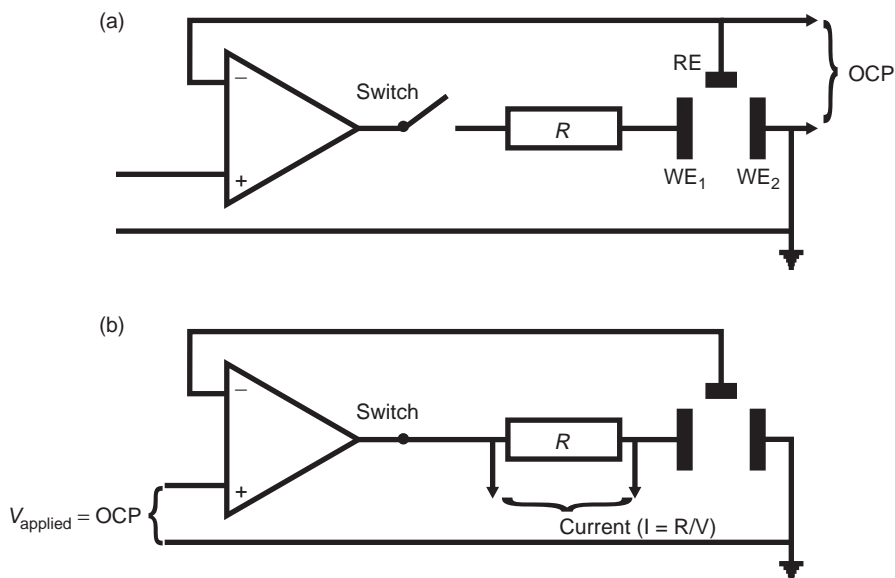


FIGURE 33.52 Sequential current and potential logging of three electrodes: (a) the open circuit potential of WE₂ and (b) switched-on mode of operation. (Adapted from Eden (2005).)

The analysis of EN results obtained between two working electrodes (WEs) with an imposed or naturally developing asymmetry can be carried out by considering Equation (33.7), which indicates that the noise impedance of a cell (Z_n) depends on the impedances of the two WEs, as well as their noise levels, represented by the power density spectra (Ψ_{i_1} and Ψ_{i_2}) obtained by performing the analysis of noise signals with either FFT or with the maximum entropy method (MEM).

$$Z_n(f) = \sqrt{\frac{\Psi_V(f)}{\Psi_I(f)}} = |Z_1(f)Z_2(f)| \sqrt{\frac{\Psi_{i_1}(f) + \Psi_{i_2}(f)}{|Z_1(f)|^2\Psi_{i_1}(f) + |Z_2(f)|^2\Psi_{i_2}(f)}} \quad (33.7)$$

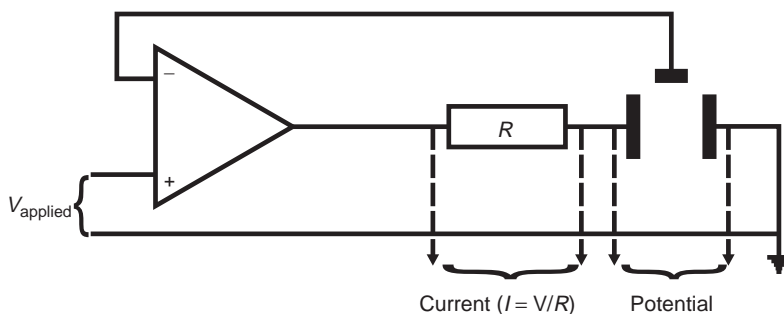


FIGURE 33.53 Three identical electrodes simultaneous current and potential measurement. (Adapted from Eden (2005).)

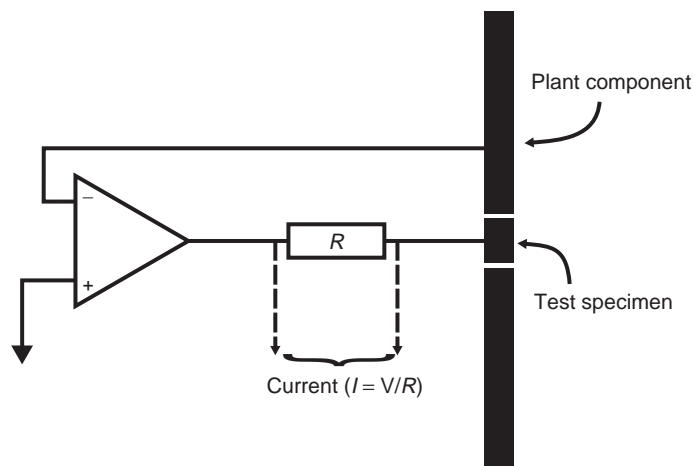


FIGURE 33.54 Single electrode measurement of current with respect to plant. (Adapted from Eden (2005).)

For the simplest case of two WEs with the same impedance ($Z_1 = Z_2$) the noise impedance is equal to the modulus of the electrode impedance $|Z(f)|$. This result is valid independently of the origin of the noise signals (localized or uniform corrosion, bubble evolution due to the cathodic reaction) and the shape of the impedance plot, even if the noise levels of the two electrodes are different. In such case, noise measurements are equivalent to impedance measurements for which the external signal perturbation has been replaced by the internal noise generated by the corrosion processes (Huet et al., 2001).

However, when the two WEs do not have the same impedance, the noise impedance analysis requires a more cautious interpretation. Depending on the source of the current

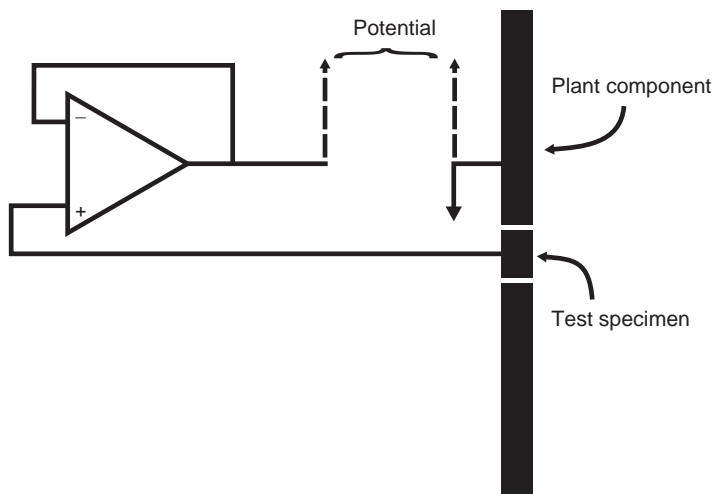


FIGURE 33.55 Single electrode measurement of potential with respect to plant. (Adapted from Eden (2005).)

noise the measured impedance may either be that of a quiet cathode or that of a quiet anode while in intermediate cases the results of the measurement of Z_n would show a mixed behavior and be more difficult to interpret. For example, if hydrogen bubbles are evolving on the cathode while the anode undergoes generalized corrosion, the noise of the cathode is orders of magnitude larger than that of the anode, so that Z_n becomes equal to the impedance modulus of the anode, $|Z_a|$. In these conditions, while the time records appear to show only the cathodic processes, the impedance measured is that of the anode, using the noise of the cathode as input signal.

An opposite case would be a cell where the anode is undergoing pitting, while the cathodic reaction is the reduction of dissolved oxygen or an imposed galvanic situation. Since the anodic noise is preponderant, Equation (33.7) shows that Z_n is equal to the impedance modulus of the cathode, $|Z_c|$. The anodic noise is the internal signal source utilized for the measurement of the impedance of the cathode.

However, this special arrangement can produce current noise signals valid on their own (Klassen et al., 2001; Schmitt et al., 2005). The design of asymmetric corrosion probes opens new possibilities for the evaluation of the susceptibility of a system to localized problems difficult to instrument otherwise: galvanic cells (inhibitors, presence of oxygen), crevice corrosion, stress corrosion cracking, etc. This is surely a very promising area of research since no other field technique has close to the sensitivity and flexibility that ENA could have for the detection of such problems.

33.3.4.2 Direct Nonintrusive Techniques By definition, direct nonintrusive measurement techniques do not require access to the inside of the system being monitored and so avoid the usual risks associated with on-line insertion and retrieval operations of intrusive devices. An important advantage of these techniques over intrusive corrosion monitoring techniques is that the actual plant material is monitored instead of a possibly different test coupon material. Depending on the specific technique and implementation, the area of the plant material monitored is generally larger than on an inserted specimen.

As mentioned earlier, most of the techniques belonging to this category in the NACE International report (Techniques for Monitoring Corrosion and Related Parameters in Field Applications, 1999) are described in Chapter 5 since they are usually described as NDE, NDT, or NDI techniques.

Thin-Layer Activation and Gamma Radiography With thin-layer activation (TLA) and gamma radiography, a technique developed from the field of nuclear science, a small section of material is exposed to a high energy beam of charged particles to produce a radioactive surface layer. For example, a proton beam may be used to produce the radioactive isotope Co-56 within a steel surface. This isotope decays to Fe-56 with the emission of gamma radiation. The concentration of radioactive species is sufficiently low that metallurgical properties of the monitored component are essentially unchanged. The radioactive effects utilized are much lower than conventional radiography and so the health risk concerns. The low levels of radiation present involve modest handling procedures and product quality, other than for human consumables, is not compromised.

Changes in gamma radiation emitted from the surface layer are measured with a separate detector to study the rate of material removed from the surface. The radioactive surfaces can be produced directly on components (nonintrusive) or on separate sensors (intrusive). The measured gamma radiation depends on the quantity of the original

radioactive tracer present and the natural decay of the isotope with time (related to its half-life). The technique has been applied to measure wear, erosion, and corrosion when the corrosion product is removed from the surface. Different components in a system can be irradiated with different isotopes to allow simultaneous measurements of these components. Activation area can be small ($<1 \text{ mm}^2$), which allows monitoring from specific metallurgical areas such as welds and weld heat-affected zones.

The gamma radiation can penetrate typically up to 5 cm of steel with acceptable signal attenuation, so that on-site external monitoring of internally generated activated layers is possible without any physical interconnection between internal and external surfaces. To compensate for the natural isotope decay and the natural background levels, three measurements are used for each reading, namely the activated component, the reference sample of the same isotope (for natural decay), and the background radiation (for naturally occurring radiation in the system and atmosphere) (Techniques for Monitoring Corrosion and Related Parameters in Field Applications, 1999). However, these techniques measure strictly material loss and do not distinguish between erosion and corrosion effects. Additionally, corrosion products still adherent to the material surface are not measured as material loss.

Field Signature Method Field signature method (FSM) is a nonintrusive technique that can monitor corrosion of a pipe wall directly. The original development of this technique was largely directed at oil and gas production. Typical applications involve pin attachment to the external surface of a pipeline, which is normally protected against corrosion, to monitor corrosion damage to the inside of the pipe wall. The technique measures corrosion damage over several meters of an actual structure and can be used for locations where access for traditional intrusive probes is difficult, and for high temperatures ($>150^\circ\text{C}$) where the application of probes and ultrasonic testing (UT) is limited. The technique is useful in difficult access areas because generally no access is required after the initial installation. There are no consumable parts in this system except for the pipe spool itself that may be consumed when used in high corrosion rate conditions.

In FSM, a current is induced in the monitored section of interest, typically 3–10 m apart for a large pipe or a few centimeters for small pipes, and the resulting voltage distribution is measured to detect corrosion damage (Figure 33.56). An array of pins is attached by stud welding, gluing, or spring-loading to make electrical connections externally over the structure.

From a corrosion perspective, the location and type of corrosion to be expected is critical in designing a FSM monitoring matrix. Up to 84 pin pairs can be typically utilized for one location. If monitoring for localized attack, the pin pairs are set up closer together, whereas for uniform corrosion the pin pairs can be spread out to cover a much larger area. A general rule for covering both types of corrosion is to position the pin pairs at three times the pipe wall thickness (Scanlan et al., 2003).

The voltage readings are monitored and compared looking for any nonuniformity that could be due to cracking or pitting in the monitored section (Figures 33.57 and 33.58). Comparison of the average voltage drop with that of a reference element is used to monitor general metal loss. The area monitored is dependent on the pin spacing with a resolution inversely proportional to pin spacing and wall thickness (Techniques for Monitoring Corrosion and Related Parameters in Field Applications, 1999). When a reference element is used, the sensitivity to general metal loss is approximately one part per thousand of wall thickness. The FSM monitoring unit has a lower sensitivity when not permanently

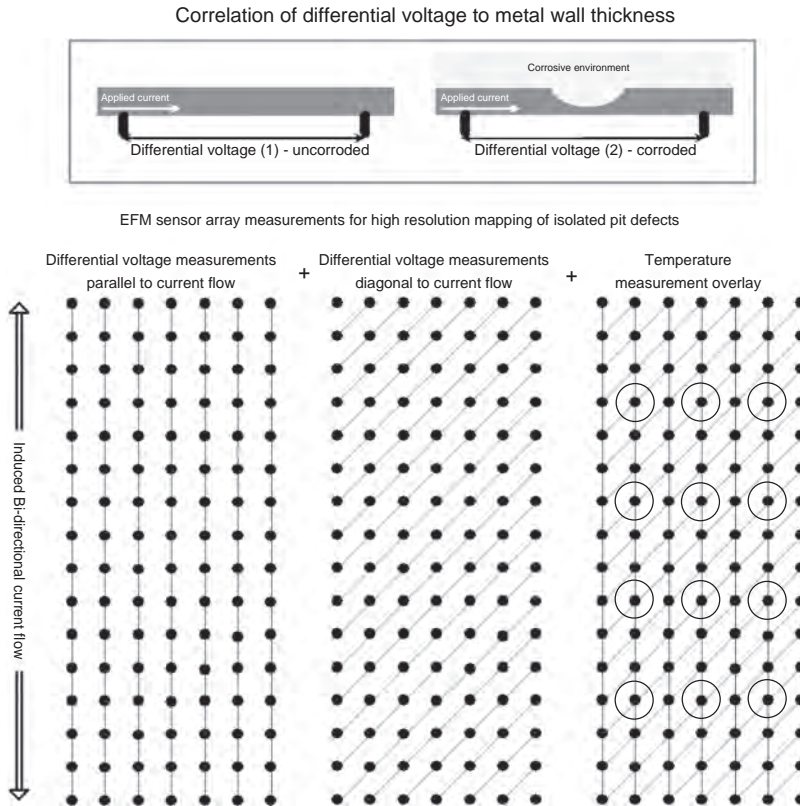


FIGURE 33.56 Schematic of the electric field monitoring sensor array for high-resolution mapping of isolated pit defects with FSM. (Courtesy of Eric Kubian, PinPoint Corrosion Monitoring Inc.)

attached because of the need to relocate the contact pins accurately for voltage measurements.

Once the locations vulnerable to the corrosion mechanism are determined, the process temperature needs to be considered. Two types of FSM systems are available. One is a high-temperature system and the other a low temperature system, with the cut off being $\sim 150^{\circ}\text{C}$ due to hardware requirements. The high-temperature system requires a further junction box located outside the insulation. However, in theory the FSM technique has no temperature limits (Scanlan et al., 2003). Following the installation of the FSM system, attention must be given to data interpretation, which can be quite complex. This technique does not distinguish between internal and external flaws or general material loss.

Acoustic Emission Acoustic emission (AE) is based on measuring acoustic sound waves that are emitted during the growth of microscopic defects, such as stress corrosion cracks. The sensor elements are thus essentially very sensitive microphones, which are strategically positioned on structures. The sound waves are generated from mechanical stresses generated during pressure or temperature changes. Background noise effects have to be

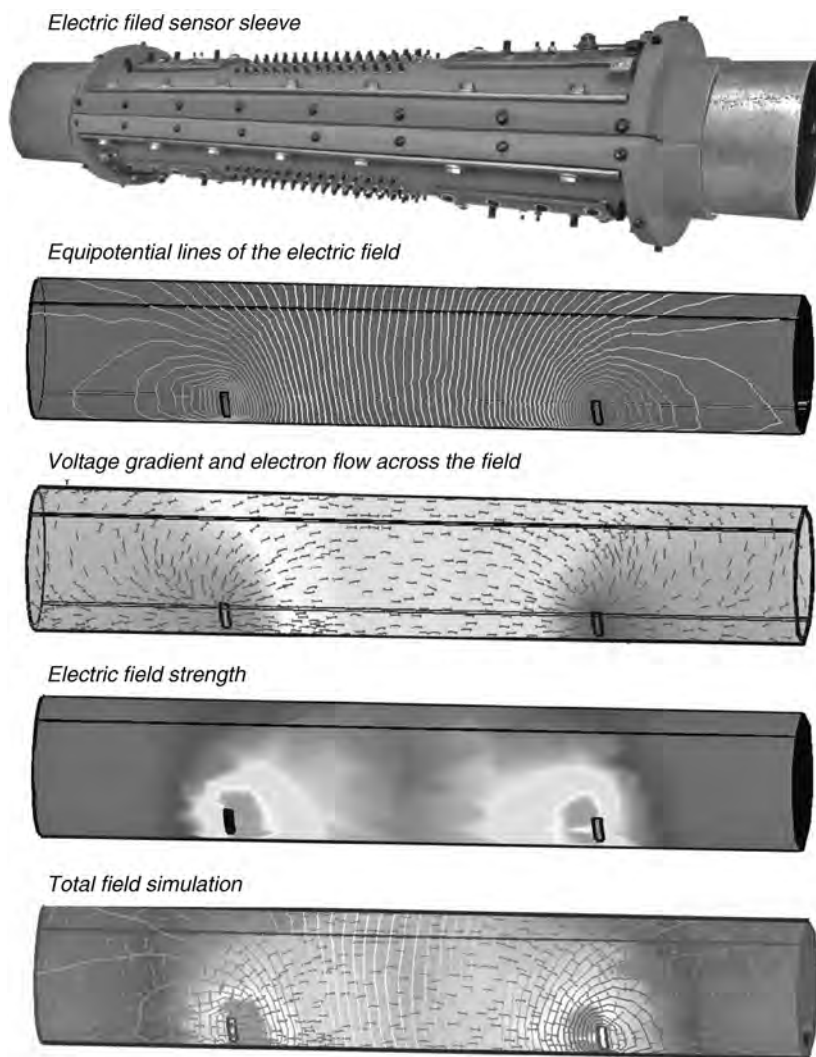


FIGURE 33.57 Electric field sensor sleeve and various representations of the electric behavior between two sensor points when monitoring with FSM. (Courtesy of Eric Kubian, PinPoint Corrosion Monitoring Inc.)

taken into consideration and can be particularly troublesome in on-line measurements. The technique produces large amounts of data and requires relatively sophisticated filtering and analysis.

It is the flaw growth or even plastic deformation that generates the sound waves that are detected by the acoustic sensors. Secondary emissions from crack fretting or breaking of corrosion pockets can also be employed for the detection of flaws. The technique is essentially qualitative in identifying areas with flaws that may be further investigated with other nondestructive techniques, (e.g., ultrasonics) (Techniques for Monitoring Corrosion and Related Parameters in Field Applications, 1999).

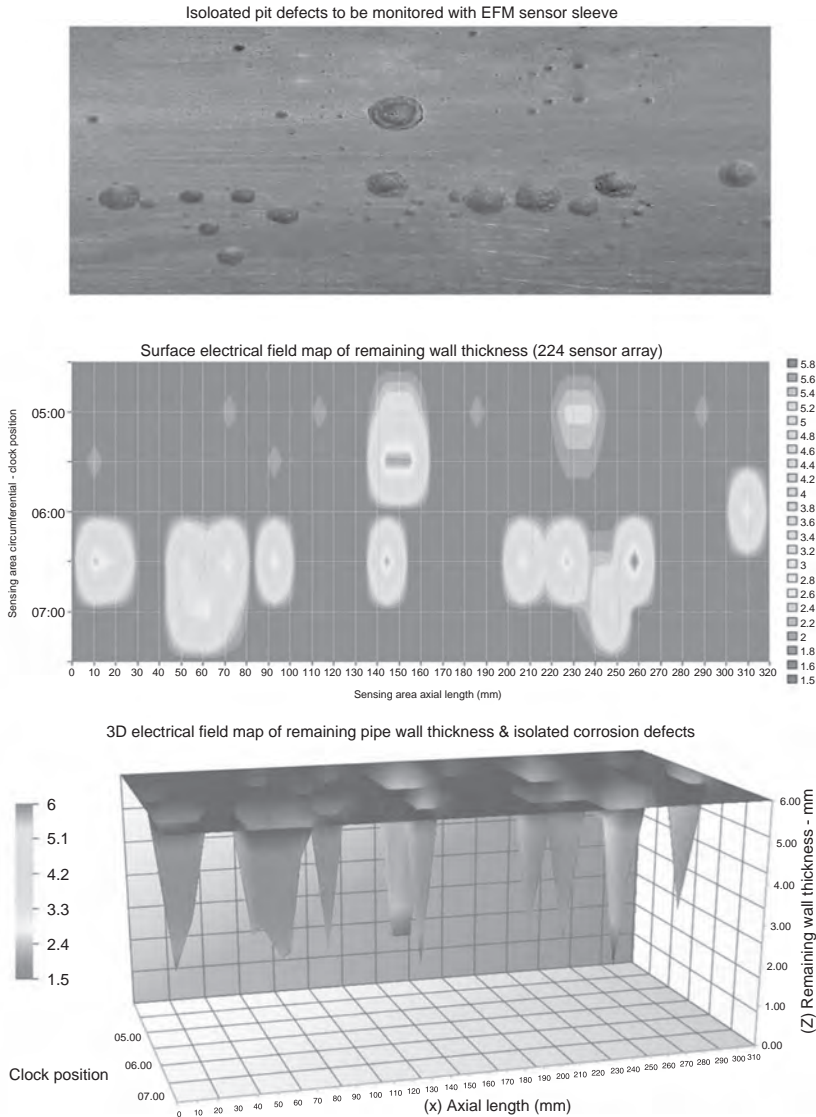


FIGURE 33.58 Example of FSM results plotted as a three-dimensional (3D) map obtained on a pitted pipe using a 224 sensor array. (Courtesy of Eric Kubian, PinPoint Corrosion Monitoring Inc.)

Testing is most commonly done off-line by applying pressure changes, or during shut-down as the temperature is changing, to minimize the background noise effects. On-line testing is also done, but the range of detection is reduced due to background noise. Frequency and gain of the sensors are modified according to the procedure to adjust filtering. Off-line tests are short-term tests to monitor for integrity whereas on-line tests are used to track the operational conditions that promote flaw propagation. A frequency range of 150–175 kHz is used for initial wide coverage in off-line applications. For on-line

applications in which flaw noise is the targeted problem, the frequency used is ~ 1 MHz, which reduces the range of coverage to ~ 0.50 m.

Triangulation can be used to estimate the location of growing flaws. Acoustic energy measured at the sensor depends on the magnitude and distance of the flaw from the sensor. In off-line tests, overpressure only detects flaws that have sufficient energy during the monitoring period to be on the verge of propagation. It does not detect flaws that are not actively growing under the applied stress excursions, even though these flaws may be potential weak points or actively growing under a more severe process excursion.

33.3.4.3 Indirect On-Line Techniques A great variety of techniques have been developed to measure the indirect changes, either in the environment or in the metallic component of interest, that can occur during corrosion processes or that can increase the rate at which these processes are occurring. This variety is reflected in the number of disciplines (e.g., metallurgy, physics, biology, chemistry, nuclear science) involved in the development of these tools designed to track and assess the effects and factors associated with corrosion damage.

Hydrogen Monitoring The principle of the hydrogen probe relies on the fact that one of the cathodic reaction products in nonoxidizing acidic systems is hydrogen which, in its atomic form, can diffuse through the thickness of vessel or pipe walls and recombine into hydrogen molecules at the exterior surface. This “uptake” of hydrogen occurs when the recombination of hydrogen atoms and subsequent release as molecular hydrogen in the environment is inhibited by the presence of some chemical species or poisons (e.g., cyanides, arsenic, antimony, selenium, or sulfur compounds). The generation of atomic hydrogen can be used for corrosion monitoring purposes in both intrusive and nonintrusive ways. In the latter, hydrogen monitoring sensors are attached to the outside walls of vessels and piping. It is the diffusion of atomic hydrogen through the metallic substrate that is of most concern, as it can lead to problems such as hydrogen induced cracking.

Hydrogen monitoring is highly applicable to the oil refining and petrochemical industries with hydrocarbon process streams. The presence of hydrogen sulfide in these industries promotes the uptake of hydrogen into plant components. Several types of hydrogen probe have been developed to monitor hydrogen flux (Figure 33.59) (Kane and Cayard, 1998). These probes are based on one of the following three principles:

1. *Hydrogen Pressure (or Vacuum) Probe:* One version of this probe technology is intrusive and consists in the insertion of a steel tube or cylinder that includes an inner cavity. The pressure in the inner cavity of the tube or cylinder is measured with a pressure gauge. Another version is nonintrusive and consists of a patch-type device, in which a patch or foil is welded or otherwise sealed to the outside of the pipe or vessel to create a cavity. In some implementations, the pressure range from zero absolute to atmospheric pressure, that is, the vacuum region. Other types use positive pressure above atmospheric pressure. Hydrogen passing through the wall of the tube or cylinder is detected as an increase in pressure in the cavity as a function of time. The higher the hydrogen flux, the greater the rate of pressure increase with time (Techniques for Monitoring Corrosion and Related Parameters in Field Applications, 1999).
2. *Electrochemical Hydrogen Patch Probe:* This nonintrusive device is attached to the outer surface of the pipe or vessel to be monitored. The patch probe itself consists of

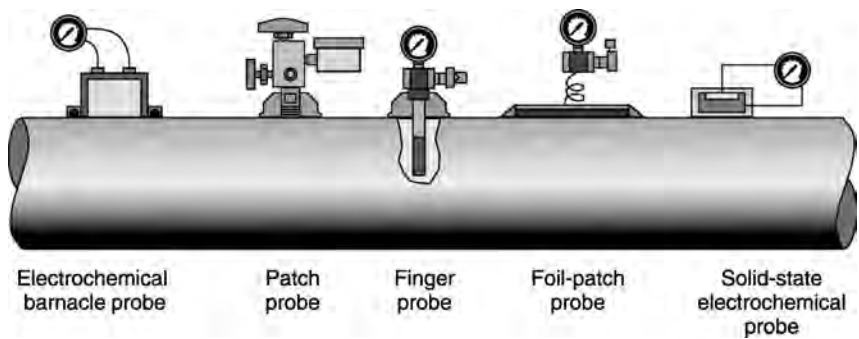


FIGURE 33.59 Various hydrogen probes, which are inserted into a vessel or pipe, or are external. (Adapted from Kane and Cayard (1998).)

a small electrochemical cell containing an appropriate electrolyte in contact with the component to be monitored (Figure 33.60). Typically, this cell uses the nickel electrode from a nickel cadmium battery with a caustic electrolyte put directly on the pipe. Often the electrolyte is isolated from the outer surface of the pipe or vessel by a thin palladium foil. This foil is also used as an electrode in the detection circuit. The cell is operated at a potential that oxidizes hydrogen as it enters the cell. The current used to maintain this potential is proportional to the hydrogen flux into the cell (Techniques for Monitoring Corrosion and Related Parameters in Field Applications, 1999).

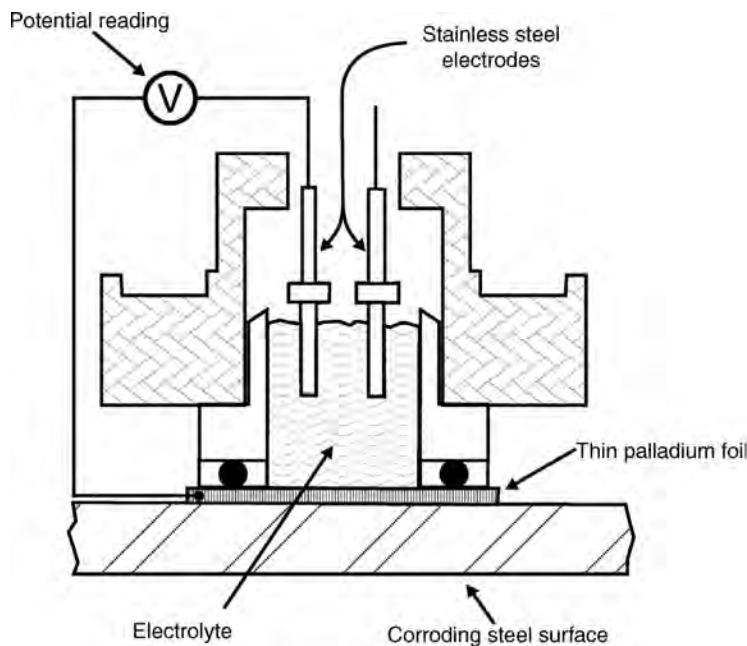


FIGURE 33.60 Schematic of an electrochemical patch probe.

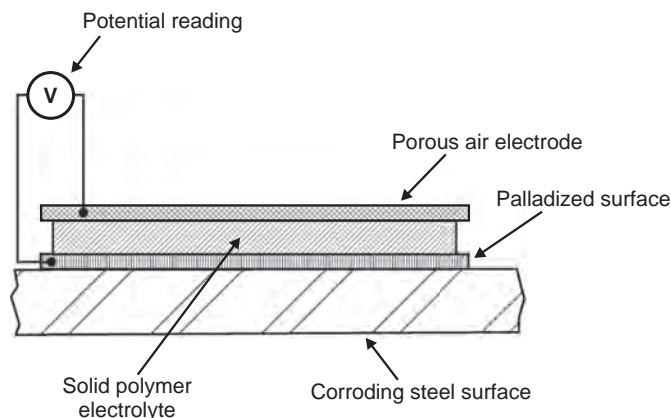


FIGURE 33.61 Schematic of the hydrogen detector using a solid electrolyte proton exchange membrane fuel cell (Vera et al., 2002).

3. *Hydrogen Fuel Cell Probe*: This other nonintrusive technique is also an electrochemical device that consists of a small fuel cell. As shown in Figure 33.61, the cell contains a solid electrolyte membrane, a cathode material, and an anode which consists of a surface catalyzed with palladium. Hydrogen entering the cell is reacted on the activated palladium surface while the cathode material reacts with ambient oxygen (air) causing a current flow in an external circuit between the cathode and anode. This current is directly proportional to the hydrogen flux through the palladium membrane (Vera et al., 2002).

When the relative distribution of atomic hydrogen in a material is constant, the permeating fraction of hydrogen can give relative corrosion rate information. However, the technique is used more commonly to detect high flow rates of the hydrogen passing through the steel that could lead to hydrogen blistering and hydrogen-induced cracking (HIC).

Hydrogen evolution is only one of a few possible cathodic reactions that can occur concurrently with the actual purely anodic corrosion reaction. This technique is not considered suitable to indicate corrosion rates in conditions or environments when any of the other cathodic reactions (e.g., reduction of dissolved oxygen) would be dominant. Currently there is no known absolute correlation between hydrogen diffusion rates and corrosion rates, cracking, or blistering without specific analysis of a piece of the actual steel affected.

Corrosion Potential Measurement of the corrosion potential (E_{corr}) is a relatively straightforward intrusive technique that is widely used in industry for monitoring reinforcing steel corrosion in concrete and structures, such as buried pipelines under cathodic or anodic protection. Changes in corrosion potential can also give an indication of active-passive behavior in stainless steel. However, this technique only indicates a corrosion risk and does not measure corrosion or pitting rates.

The parameter E_{corr} (also known as rest potential, open-circuit potential, or freely corroding potential) is measured relative to a reference electrode, which is characterized by a

stable half-cell potential. Either a reference electrode has to be introduced into the corrosive medium for these measurements, or an electrical connection has to be established to a structure in conjunction with an external reference electrode in contact with the structure with a wet electrolyte.

The success of corrosion-potential measurements depends on the long-term stability of the reference electrode. Such electrodes have been developed for continuous application for measuring corrosion potentials. However, conditions of temperature, pressure, electrolyte composition, pH, and other possible variables can limit the usage of these electrodes for corrosion monitoring service. Special reference electrodes are needed for operating temperatures above about the boiling point of water (100°C).

The use of E_{corr} measurements for in-service corrosion monitoring in process environments is not as widespread as the use of polarization resistance. However, this approach can be valuable in some cases, particularly where an alloy could show both active and passive corrosion behavior in a given process stream. For example, stainless steels can provide excellent service, as long as they remain passive. However, if an upset occurs that might introduce either chlorides or reducing agents into the process stream, these alloys could become active and exhibit excessive corrosion rates. Corrosion-potential measurements would indicate the development of active corrosion that could be confirmed by using LPR, for example (Dean, 2003).

On-Line Water Chemistry Analyses Various types of chemical analyses can provide valuable information to a corrosion monitoring programs. The measurement of pH, conductivity, dissolved oxygen, metallic and other ion concentrations, water alkalinity, concentration of suspended solids, inhibitor concentrations, and scaling indexes all fall within this domain. Several of these measurements can be made on-line using appropriate sensors.

All of the following measurements are used particularly in water treatment, and in general apply only to aqueous environments. Many of these parameters are used in combination to generate various indices, such as the Langelier index,⁴ in order to rate the scaling tendency and corrosivity of waters.

pH The pH^5 of an aqueous environment may be measured with a meter or calculated if certain parameters are established. Water itself dissociates to a small extent to produce equal quantities of H^+ and OH^- ions displayed in the equilibrium described in Equation (33.8).



The term pH was derived from the manner in which the hydrogen ion concentration is calculated. It is the negative logarithm of the hydrogen ion (H^+) concentration as shown in Equation (33.9).

$$\text{pH} = -\log_{10}(a_{\text{H}^+}) \quad (33.9)$$

⁴ See Appendix B for additional information on Langelier, Ryznar, and other scaling indices.

⁵ The pH, originally defined by Danish biochemist Søren Peter Lauritz Sørensen in 1909, is a measure of the concentration of hydrogen ions.

where \log is a base 10 logarithm and a_{H^+} is the activity (related to concentration) of hydrogen ions.⁶

A higher pH means that there are fewer free hydrogen ions. A change of one pH unit reflects a 10-fold change in the concentrations of the hydrogen ion. For example, there are 10 times as many hydrogen ions available at pH 7 than at pH 8. Substances with a $\text{pH} < 7$ are considered to be acidic and substances with $\text{pH} > 7$ are considered to be basic or alkaline. Thus, a pH of 2 is very acidic and a pH of 12 very alkaline. In general, for steel, low pH (or high acidity) produces a particularly corrosive environment. For other alloys this can vary.

The pH is an important corrosion factor for two reasons. First, the H^+ ion can be reduced into hydrogen and thus participate to the overall corrosion processes as the cathodic reaction. Second, the pH influences the solubility of the products of the chemical corrosion reactions, particularly the passivation reactions involving oxides, sulfides, or carbonates. As a measure of alkalinity, pH is a very important component in scaling indices.⁷

Generally, pH measurements have been limited to a maximum temperature of boiling water and to a maximum pressure of ~ 2 MPa, although some special-purpose probes are available for up to 70 MPa.

The pH measurements can be affected by interfering ions, such as lithium, sodium, and potassium ions that can interact with the sensing glass membrane of a pH electrode. However, because lithium ions are normally not found in sample solutions and potassium ions cause little interference, the most significant interference is from sodium ions.

Additionally, fouling of the probe measurement element by hydrocarbons, for example, can reduce or even completely block the probe response to pH changes. Frequent maintenance might thus be required to ensure cleanliness and maintain calibration. Low-conductivity solutions can also be a problem. Even low-ionic-strength pH probes can have problems at conductivities less than $20 \mu\text{S}/\text{cm}$. Some low-conductivity probes inject an electrolyte to correct for high solution resistance.

CONDUCTIVITY In the present context, conductivity is the electrolytic current-carrying capacity of water, which is largely determined by the concentration of ions from dissolved salts and solids (Standard Test Methods for Electrical Conductivity and Resistivity of Water, 2005). When a salt dissociates, the resulting ions interact with surrounding solvent molecules or ions to form charged clusters known as solvated ions. These solvated ions can move through the solution under the influence of an externally applied electric field. Such motion of charge is known as ionic conduction. Ionic conductance, which for the bulk solution is the only conductance present, is the reciprocal of ionic resistance. The dependence upon the size and shape of the conductor can be corrected by using conductivity κ rather than conductance G , as expressed in Equation (33.10) for the simple geometry shown in Figure 33.62.

$$\kappa = \left(\frac{\ell}{A}\right) \frac{1}{R} = G \left(\frac{\ell}{A}\right) \quad (33.10)$$

⁶ The p in Equation (33.9) stands for the German word for “power,” *potenz*, so pH is an abbreviation for power of hydrogen.

⁷ See Appendix B for additional information on Langelier, Ryznar, and other scaling indices.

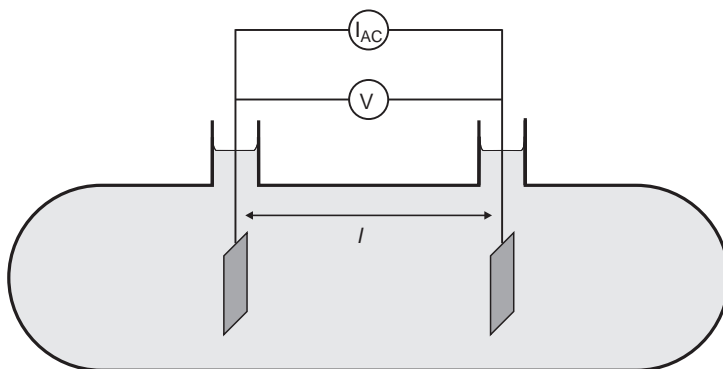


FIGURE 33.62 Schematic of a conductivity cell containing an electrolyte and two inert electrodes of surface A parallel to each other and separated by distance ℓ .

where ℓ is the length of the conductor, that is, the gap separating the electrodes in Figure 33.62; A is the cross-sectional area of each electrode, assuming that both electrodes have the same dimensions.

Because corrosion is an electrochemical process, an increase in conductivity is generally associated with an increase in corrosivity. However, this does not mean that zero electrolyte conductivity signifies that corrosion is stifled. Extra pure water such as used in nuclear reactors, for example, has proven to be quite corrosive unless small amounts of oxygen or other passivating agents are added to counter this aggressiveness.

Electrochemical reactions at the electrode interfaces can affect the readings and can be dependent on the measurement frequency used. Therefore a frequency of 1 kHz is commonly used as the measurement frequency to avoid this problem. Routine cleaning of the measurement cell has often been necessary to prevent false readings due to bridging of the electrodes, or fouling on the membrane.

DISSOLVED OXYGEN Dissolved oxygen refers to the amount of oxygen dissolved in a liquid, usually expressed as parts per billion (ppb) or parts per million (ppm) in the engineering world or in milligrams per liter (mg/L) in the chemical world. The solubility of oxygen depends on temperature, pressure, and molarity of the solution. Increased pressure increases oxygen solubility while an increase in temperature decreases its solubility (see Chapter 1 for additional details).

Dissolved oxygen can be measured by ion-selective electrodes, or for approximate low levels of oxygen a galvanic probe can be used (Figure 33.63). The oxygen ion-selective electrode has a thin organic membrane covering a layer of electrolyte and two metallic electrodes. Oxygen diffuses through the membrane and is electrochemically reduced at the cathode. There is a fixed voltage between the cathode and anode so that only oxygen is reduced. The greater the oxygen partial pressure, the more oxygen diffuses through the membrane in a given time. The result is a current that is proportional to the oxygen in the sample (Standard Test Methods for Dissolved Oxygen in Water, 2003).

Temperature sensors can be built into the probe to make corrections for the sample and membrane temperatures. The cathode current, sample and membrane temperatures, barometric pressure, and salinity information are used to calculate dissolved oxygen content of the sample in either concentration (ppb or ppm or mg/L) or percent saturation (% sat).

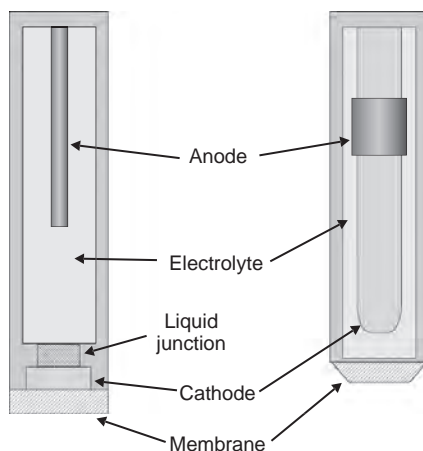


FIGURE 33.63 Two different electrode chemical cell designs to measure dissolved oxygen.

The affinity of oxygen for most structural metals is the cause of many corrosion phenomena. Oxygen is responsible for both corrosive attack and passivation. The approach to controlling corrosion of carbon steel in oil fields and in boilers in North America has generally been to remove all or most of the molecular oxygen by various combinations of physical and chemical means. Bringing oxygen levels <20 ppb or $20\text{ }\mu\text{g/L}$ has a significant effect on corrosion. With an oxygen scavenger, typical levels are <1 ppb or $1\text{ }\mu\text{g/L}$. In some boilers that use high-purity deionized water, no oxygen scavenger is used but instead small amounts of oxygen or hydrogen peroxide are deliberately added to the feed-water to act as a passivating agent under carefully controlled boiler water chemistry conditions.

Generally, dissolved oxygen measurements have been limited to a maximum temperature of 65°C and a maximum pressure of $200\text{--}350\text{ kPa}$. Consumption of oxygen by the probe may also lower the oxygen concentration near the probe, so that a flowing or stirred sample is generally recommended. The electrolyte is also depleted of its active ingredients as it reduces the oxygen and needs to be replaced at regular intervals of a few weeks or months depending on the design and use of the probe. Periodic calibration of the probes is therefore a recommended practice. Nonetheless, poisoning of the electrode can occur. This has precluded the use of the method in many industrial environments such as processes involving heavy hydrocarbon and water.

OXIDATION–REDUCTION (REDOX) POTENTIAL Oxidation reduction (redox) potential is the potential of a reversible oxidation–reduction electrode measured with respect to a reference electrode in a given electrolyte. The measurement system comprises a potentiometer (high-sensitivity and high-impedance voltmeter) connected to a reference electrode and a noble metal sensing electrode. The change in potential of the sensing electrode produced by the oxidizing or reducing species, is measured relative to the reference electrode. This voltage difference is then converted electronically to display on a meter or recorder for real-time on-line measurement.

This technique has also been used to detect the end point of oxidation or reduction reactions during a water treatment to control more closely the addition of oxidizing

biocides, such as chlorine and bromine, which may have a significant effect on increasing the corrosion rates.

Redox potential is also a key measure to evaluate soil corrosivity where the redox potential is essentially related to the degree of aeration. A high redox potential indicates a high oxygen level. Low redox values may provide an indication that conditions are conducive to anaerobic microbiological activity. Sampling of soil will obviously lead to oxygen exposure and unstable redox potentials are thus likely to be measured in disturbed soil.

Process Variables

FLOW REGIME Corrosion usually occurs on the metal surface wetted by the aqueous phase. This wetting action and transport of corrosive agent to the surface can be controlled by the flow regime. In some flow regimes, such as slug flow, high surface shear stresses and extreme turbulence can make corrosion inhibition difficult and increase corrosion rates to several centimeters per year. At higher flow regimes, impingement, cavitation, and erosive conditions created by two-phase flows can be extremely corrosive.

In single-phase flows, the flow regime is typically described as laminar, transition, or turbulent flow as defined by the Reynolds number, which depends on the flow velocity, the pipe diameter, the fluid viscosity, and the fluid density. In multiphase flows the relative pattern of multiple phases is important as defined by mist flow, annular flow, or slug flow. In single-phase environments, the flow regime affects the mass transfer to the metal surface and the shear stresses on the metal surface. These have a direct impact on the corrosion rates at the metal surface (see Chapter 2 for a description of the effects of flow on corrosion).

In multiphase flows the flow regime varies with varying flow rates of each phase, elevation changes, or specific geometry of a line or conduit. Acoustic monitoring and on-line γ -radiography have been used to detect slug flow. Flow characteristics can also be used to provide predictions of corrosion rates based on modeling and other empirical data.

PRESSURE Pressure can affect the proportion of phases present in a vessel or pipe, or the composition of the process fluids. Different phases and constituents can produce quite different corrosive environments. For example, the partial pressure of CO_2 affects the amount of CO_2 dissolved in water, which in turn affects the corrosivity of the fluid due to the presence of carbonic acid. Similarly, the partial pressure of H_2S is a major determining factor in the susceptibility of various alloys to sulfide stress cracking (SSC).

Total pressure measurements can be made on-line, and are a real-time measurement. Determination of partial pressures utilizes knowledge of the composition of the process fluid, the temperature, and the total pressure. This involves sampling of the process fluid. In some gas-liquid systems, pressure affects gas solubility, but the relationships are complex. Consequently, the pressure is usually used only to analyze and predict which phases are present, rather than as an on-line measurement.

TEMPERATURE The temperature of a process fluid can have a direct effect on its corrosion dynamics. These temperature effects can be nonlinear. Low temperature can produce condensation of water or other corrosive liquids. High temperature increases chemical reaction rates and can change the composition of the process. In general, increased temperature increases the chemical reaction rates. The temperature can lead either to vaporization (a dry condition) or condensation (from a dry to a wet condition). Both these changes in state can affect corrosion.

DEWPOINT Dewpoint is the temperature at which a liquid starts condensing from a vapor or gaseous phase. Dewpoint monitoring is important because the region of condensation of water and corrosive fluids in otherwise dry environments has a major impact on corrosion rates. This can be complex in some environments where multiple gases and condensable combinations are present. In atmospheric environments in which the condensing liquid is water, the dewpoint temperature is measured with a wet-bulb thermometer. In a process system, this may not be practical, and a system in which cooling of the flow to the point of condensation on an optical mirror has been used.

The direct effects of dewpoint conditions can be measured with mass-loss coupons or electrical resistance techniques. Electrochemical methods, such as electrochemical noise and multielectrode array systems (CMAS), have also been used in a similar way, either as a time-of-wetness indicator or for corrosion rate measurements, since these electrochemical techniques will function even when surfaces are only partially wetted. Additionally, forced cooling of the corrosion measurement surfaces can be used to generate the appropriate dewpoint on the measurement surfaces.

Fouling Fouling is an accumulation of both organic and inorganic substances from a fluid stream onto the surfaces of the equipment through which the fluid is circulated as well as *in situ* corrosion products and inorganic deposits, such as hardness salts that occur on the metal surface. Fouling is a prime cause of underdeposit corrosion since it can produce highly anodic areas and give rise to hot spots in boilers, for example. Fouling may also restrict flow by causing an increase in pressure drop through the equipment or retard heat transfer by formation of an insulating deposit. Side-stream and in-line intrusive measurement techniques, together with visual inspection, have been used for determination of fouling. One of two techniques has been employed, that is, heat-transfer or pressure-drop monitoring.

33.3.4.4 Indirect Off-Line Measurement Techniques A great variety of chemical analysis techniques have been used over the years to provide operators and providers of corrosion control services some estimates of the effects of corrosion on systems, on the corrosivity of given environments, or on the effective usage of chemical additives. The following sections describe the type of information commonly sought with these techniques (Techniques for Monitoring Corrosion and Related Parameters in Field Applications, 1999).

Off-Line Water Chemistry Parameters

ALKALINITY Alkalinity is considered to be an important measurement when handling waters because carbonate and hydroxide scales are common problems. The alkalinity of a water as determined by titration with standard acid solution to the methyl orange endpoint (pH \sim 4.5) is sometimes abbreviated as “M alkalinity”. Total alkalinity includes many alkalinity components, such as hydroxides (OH^-), carbonates (CO_3^{2-}), and bicarbonates (HCO_3^-).

METAL ION ANALYSIS Some corrosion products are soluble and can be detected by metal ion analysis of the process stream for determining the amount of metal lost that has dissolved in the process stream or the amount that has been carried along in the process stream as corrosion product (e.g., iron, copper, nickel, zinc, and manganese). The analysis

can be done easily, inexpensively, and quickly in the field. The assumption that metal loss occurred over the total surface area or that the concentration of ions in solution is proportional to the corrosion rate is often incorrect, in which case the technique may only provide a trend indication. An increase in the metal ion concentration can indicate an increase in corrosion. However, a low metal ion concentration is not a guarantee of low corrosion, due to the possibility of localized corrosion, deposition of the metal ion due to temperature or pH changes, or a significant time delay before analysis.

Metal ion analysis is most useful when applied in closed systems and if the corrosion products are soluble or relate to particular concentrations of soluble species. In open systems, the relative changes in ion concentrations between locations can provide some useful information. The technique is generally not reliable in fluids containing hydrogen sulfide due to the precipitation of insoluble sulfides from soluble metallic ions, or in alkaline solutions because of precipitation of hydroxides.

Much care should be used in obtaining a representative sample of the aqueous phase, including the sampling point design, because the sampling point can accumulate corrosion products. Obtaining a representative sample from a process stream can also be difficult, because fluid velocity, temperature, and pressure can vary greatly with time unless specific precautions are taken to prevent this.

CONCENTRATION OF DISSOLVED SOLIDS Total dissolved solids (TDS) is the sum of minerals dissolved in water. It is therefore an important parameter for the determination of a scaling index.⁸ Total dissolved solids is determined by evaporating the water from a pre-weighed sample. A calculated TDS value can be determined by adding the various cations and anions from an analysis.

GAS ANALYSIS Gas analysis commonly includes hydrogen, H₂S, or other dissolved gases. This technique is generally done in a laboratory, but can also be done for certain gases in the field to determine potentially corrosive constituent gases, particularly acid gases, that become corrosive when hydrated, for example, when the temperature falls below the dewpoint. However, the technique is limited to very specific gaseous process situations and the equipment for analysis can be relatively expensive to acquire and maintain.

RESIDUAL OXIDANT Ozone and halogens, specifically chlorine, chlorine dioxide, and bromine, are powerful oxidizing agents and are widely used to control microbiological fouling in aqueous systems. Residual halogen can directly oxidize the inhibitors used to protect against corrosion or fouling. Dissolved halogens in aqueous environments can be measured using redox potential or one of a variety of colorimetric techniques.

Halides are salts that come from halogen acids, typically HF, HCl, HBr, and HI. Halides have been directly implicated as a causative agent in SCC and have been indirectly linked to galvanic corrosion by increasing the conductivity of the environment. Halides can be measured using specific ion electrodes or with a colorimetric method.

MICROBIOLOGICAL ANALYSIS Microbiological influenced corrosion is a very pervasive factor in many corrosion situations. Due to the importance of the subject, a special section of

⁸ See Appendix B for additional information on Langelier, Ryznar, and other scaling indices.

this chapter is devoted entirely to the techniques and methods that have been developed to monitor microbes and bacteria that are the most susceptible to cause corrosion damage.

Residual Inhibitor Measurement of corrosion inhibitor residuals in a system provides an indication of the concentrations of inhibitor at various locations of a system. When the reliability of an analytical test method is established and the minimum acceptable inhibitor concentration has been determined, measurement of inhibitor concentrations throughout a system can indicate whether adequate corrosion protection is likely to be achieved at each sample location. Additional inhibitor loss due to reaction with the system hardware or reactions with the environment (e.g., absorption, neutralization, precipitation, adsorption on solids or corrosion products) can be detected and compensated for by dosage selection.

Filming Corrosion Inhibitors: Filming corrosion inhibitors are chemical products used in low concentrations to adsorb on system surfaces and shield them from corrosive agents. Filming inhibitor residual measurements are usually made to ensure that an adequate supply of inhibitor has been introduced in the system to compensate for a reduction of the inhibitor concentration by adsorption and maintenance of inhibitor film;

Reactant Corrosion Inhibitors: Reactant corrosion inhibitors are designed to react with potential corrosive agents in a system to neutralize their harmful effect and combine with constituents in a system to produce *in situ* corrosion inhibitors. Reactant corrosion inhibitor residual measurements are used to ensure that sufficient reactant inhibitors have been injected into the system to provide an excess of reactant. Measurement of residual reactant inhibitor can be taken after the point of introduction of oxidant (oxygen or air entry, e.g.) to ensure that sufficient inhibitor has been injected to perform the reaction while leaving some residual inhibitor.

Reactant corrosion inhibitors can be alkaline neutralization chemicals used to maintain a safe operating pH and, for example, adjust the pH at the point of condensation of acid gases in a distillation system. A typical parameter in determining the effectiveness of such an inhibitor is the ability to measure the pH at appropriate points in a system. Reactant inhibitors can also be used to react with or reduce oxidants in a system and thus remove them from the environment. This type of inhibitor or oxygen scavenger can include sulfites, bisulfites, and hydrazine for example.

Chemical Analysis of Process Samples Chemical analysis of process samples taken at times of high and low corrosion rates can be useful in identifying the constituents that cause the high corrosion rates. From this information, the source of aggressive constituents can be found and corrected or modified. In petroleum production, crude oil and gas condensate samples are typically analyzed for organic nitrogen and acid. Sulfur, organic acid, nitrogen, and salt content are typically measured for refining purposes. These parameters are used in combination to predict the corrosivity of the oil.

In gas handling and gas processing operations, samples of produced and processed natural gas and gas-liquids are typically analyzed to determine hydrogen sulfide (H_2S), carbon dioxide (CO_2), water (H_2O), carbonyl sulfide (COS), carbon disulfide (CS_2), mercaptans (RSH), and/or oxygen (O_2) to predict and assess corrosion potential in producing wells, gas gathering systems, and gas processing operations.

In this monitoring technique, a representative sample of a process stream is taken and kept in a vessel that maintains the original condition of the sample for subsequent analysis. The sampling methods can be quite complex since the samples need to be maintained under the pressure conditions of the process stream. This can be particularly difficult for high-pressure process systems.

SULFUR CONTENT Sulfur is the most abundant element in petroleum other than carbon and hydrogen. Corrosion of carbon steel may become extremely high (~ 100 mm/year!!) when the sulfur $>0.2\%$. It can be present as elemental sulfur, dissolved hydrogen sulfide, mercaptans, sulfides, and polysulfides. The total sulfur content is generally analyzed according to ASTM D 4294. Halides and heavy metals interfere with this method and the capability of a sulfur compound to form H_2S during heating in the refinery process, rather than the total amount of the compound, is believed to correlate with corrosion in the plants.

TOTAL ACID NUMBER Acid content is generally expressed in terms of total acid number (TAN) or neutralization (neut) number. For the purpose of predicting corrosion in crude distillation units in refineries, the TAN threshold is believed to be ~ 0.5 for whole crude and 2.0 for the cuts. In petroleum production, it was found that the corrosion rate is usually inversely proportional to the algebraic product of the nitrogen concentration and the TAN.

With this monitoring technique, an oil sample is dissolved in a mixture of toluene and isopropyl alcohol containing a small amount of water. The solution is then titrated with an alcoholic potassium hydroxide solution (KOH). Both ASTM D 664 and ASTM D 974 can be used to determine the TAN. However, ASTM D 664 yields numbers 30–80% higher than does ASTM D 974.

Inorganic acids, esters, phenolic compounds, sulfur compounds, lactones, resins, salts, and additives, such as inhibitors and detergents, sometimes interfere with the measurement. Universal Oil Products (UOP) Methods 565 and 587 give procedures to remove most of the interfering materials before performing the analysis for organic acids.

NITROGEN CONTENT The total nitrogen is generally analyzed according to ASTM D 3228 for the assessment of the corrosivity of process feedstocks. High nitrogen content indicates corrosion-inhibitive properties of the crude oil or condensate in petroleum production. As already stated, the corrosion rate is inversely proportional to the algebraic product of the organic nitrogen concentration in weight percent and the TAN. In refining, however, when the organic nitrogen concentration exceeds 0.05%, cyanide and ammonia can form, collect in the aqueous phase, and corrode certain materials.

SALT CONTENT OF CRUDE OIL Salt (primarily sodium chloride with lesser amounts of calcium and magnesium chloride) is present in produced water, and this produced water can be dispersed, entrained, and/or emulsified in crude oil. Salt can cause corrosion of refinery equipment and piping by the formation of hydrochloric acid through hydrolysis and heating. Salt precipitates can form scale in heaters and heat exchangers and this can result in accelerated corrosion of equipment. Refineries generally limit salt content of crude oil for processing to 2.5–12 mg/L.

The salt content of crude oil can be determined using ASTM D 3230. The analytical procedure assumes that calcium and magnesium exist as chlorides and all the chloride is calculated as sodium chloride.

33.3.5 Monitoring Microbiologically Influenced Corrosion

An effective biocorrosion mitigation program needs to include corrosion monitoring as a periodic or continual means of assessing whether program goals are being achieved. This is particularly true in industrial water-handling systems with known susceptibility to biocorrosion (e.g., cooling water and injection water systems, heat exchangers, wastewater treatment facilities, storage tanks, piping systems, and all manner of power plants, including those based on fossil fuels, hydroelectric, and nuclear) (Dexter, 2003). Table 33.7 lists potential problem areas by industry (Scott, 2004).

As mentioned in Chapter 1, the interaction of microbial metabolism and corrosion processes can produce localized attack at very high rates. Monitoring techniques that detect the presence of microbes, especially on metallic surfaces, can provide an early indication of incipient MIC or the potential for MIC. A number of methods for the detection of microorganisms, including specific types of organisms and estimates of their numbers and activity, have been developed (Zintel et al., 2001).

The first biocorrosion monitoring systems were focused on assessing the number of microbes per unit volume of water sampled from the system. These data were combined with electrochemical corrosion measurements, using ER or LPR probes in addition to coupon weight loss measurements. The problem with this approach is that the number of free-floating planktonic organisms in the water does not correlate well with the organisms present in biofilms on the metal surface where the corrosion actually takes place. An effective monitoring scheme for controlling both biofouling and biocorrosion should include data gathering of as many of the following types of data as possible (Dexter, 2003):

- Sessile bacterial counts of the organisms in the biofilm on the metal surface done by either conventional biological techniques or optical microscopy.
- Direct observation of the community structure of the biofilm. This can be done on metal coupons made from the same alloy used for the system. Several types of probe systems are commercially available for holding and inserting such coupons into the system. Examination of the biofilm has been done by SEM, epifluorescence optical microscopy, or confocal laser scanning microscopy.
- Identification of the microorganisms found in both the process water and on the metal surface.
- Surface analysis to obtain chemical information on corrosion products and biofilms.
- Evaluation of the morphology of the corrosive attack on the metal surface after removal of biological and corrosion product deposits. Conventional macrophotography, as well as low-power stereomicroscopy, optical microscopy, metallography, and SEM may all be helpful in this regard.
- Electrochemical corrosion measurements.
- Water quality and redox potential measurements.
- Other types of information specific to each operational system, including duty cycle and downtime information, concentrations and timing for addition of biocides and other chemical inputs, local sources and nature of pollutants, and so on.

TABLE 33.7 Where MIC Problems Are Most Likely to Occur

Industry/Application	Potential Problem Sites for MIC	Organisms Responsible
Pipelines—oil, gas water, and wastewater	Internal corrosion primarily at the bottom (6:00) position Dead ends and stagnant areas Low points in long-distance pipes Waste pipes-internal corrosion at the liquid/air interface Buried pipelines-on the exterior of the pipe, especially in wet clay environments under disbonded coating	Aerobic and anaerobic acid producers, SRB, manganese and iron-oxidizing bacteria, sulfur-oxidizing bacteria
Chemical process industry	Heat exchangers, condensers, and storage tanks-especially at the bottom where there is sludge build-up Water distribution systems (See also “Cooling water systems,” “Fire protection systems,” and “Pipelines” in this table)	Aerobic and anaerobic acid producers, SRB, manganese, and iron-oxidizing bacteria In oil storage tanks also methanogens, oil-hydrolyzing bacteria
Cooling water systems	Cooling towers Heat exchangers-in tubes and welded areas-on shell where water is on shell side Storage tanks-especially at the bottom where there is sludge build-up	Algae, fungi, and other microorganisms in cooling towers Slime-forming bacteria, aerobic and anaerobic bacteria, metal-oxidizing bacteria, and other microorganisms and invertebrates
Fire protection systems	Dead ends and stagnant areas	Anaerobic bacteria, including SRB
Docks, piers, oil platforms, and other aquatic structures	Just below the low-tide line Splash zone	SRB beneath barnacles, mussels, and other areas sequestered from oxygen
Pulp and paper	Rotating cylinder machines Whitewater clarifiers	Slime-forming bacteria and fungi on paper making machines Iron-oxidizing bacteria SRB in waste water
Power generation plants	Heat exchangers and condensers Firewater distribution systems Intakes	As above for heat exchangers and fire protection systems Under mussels and other fouling organisms on intakes
Desalination	Biofilm development on reverse osmosis membranes	Slime-forming bacteria

33.3.5.1 Planktonic Organisms Depending on the type of industrial system, planktonic organisms may include, besides bacteria, unattached algae, diatoms, fungi, and other microorganisms present in a system bulk fluids. In most cases, it is planktonic bacteria that are the focus of monitoring for MIC using microbiological detection techniques, since system fluids are generally easier to sample than metallic surfaces. Unfortunately, the levels of planktonic bacteria present in the liquids are not necessarily indicative of MIC problems or their severity (Zintel et al., 2001).

At best, the detection of viable planktonic bacteria may serve as an indicator that living microorganisms are present in a particular system, some of these organisms are capable of participating in the microbial attack. It is generally very important that additional monitoring methods be performed to confirm that actual corrosion due to microbial processes has occurred.

33.3.5.2 Sessile Organisms Microorganisms that are attached to a surface are termed “sessile” organisms. These organisms are most often present as a consortium or community, collectively referred to as a biofilm. Complex assemblages of various species may occur within both planktonic and sessile microbial populations. The environmental conditions largely dictate whether the microorganisms will exist in a planktonic or sessile state.

Sessile microorganisms do not attach directly to the actual surface, but rather to a thin layer of organic matter adsorbed on the surface (Figure 33.64, Stages 1 and 2). As microbes attach to and multiply, a biofilm composed of immobilized cells and their extracellular polymeric substances builds up on the surface.

The growing biofilm increasingly prevents the diffusion of dissolved gases and other nutrients coming from the bulk liquid. These changing conditions become inhospitable to some microorganisms at the base of the biofilm, and eventually many of these cells die. As the foundation of the biofilm weakens, shear stress due to adjacent fluid flow, for example, may cause sloughing of cell aggregations exposing the bare surface to the bulk fluid in localized areas (Figure 33.64, Stage 5). The exposed areas are subsequently recolonized and new microorganisms and their exopolymers are woven into the fabric of the existing biofilm (Figure 33.64, Stage 6). This phenomenon of biofilm instability occurs even when the physical conditions in the bulk liquid remain constant. Thus, biofilms are constantly in a state of flux (Geesey, 1993).

Since MIC occurs directly on metal surfaces, the presence of sessile organisms is an important component to monitor. Monitoring sessile organisms either requires that the system be regularly opened for sampling or that accommodations be made in the system design to allow for regular collection or on-line tracking of attached organisms while the system continues to operate. It should be pointed out, however, that the presence of viable sessile organisms does not always translate into actual attack on the metal surfaces. Again, it is a good idea to use additional methods that directly determine the presence of active MIC.

33.3.5.3 Sampling Samples for analysis can be obtained by scraping accessible surfaces. In open systems or on the outside of pipelines or other underground facilities, this can be done directly. Bull plugs, coupons or inspection ports can provide surface samples in low pressure water systems (Sanders, 1988). More sophisticated devices

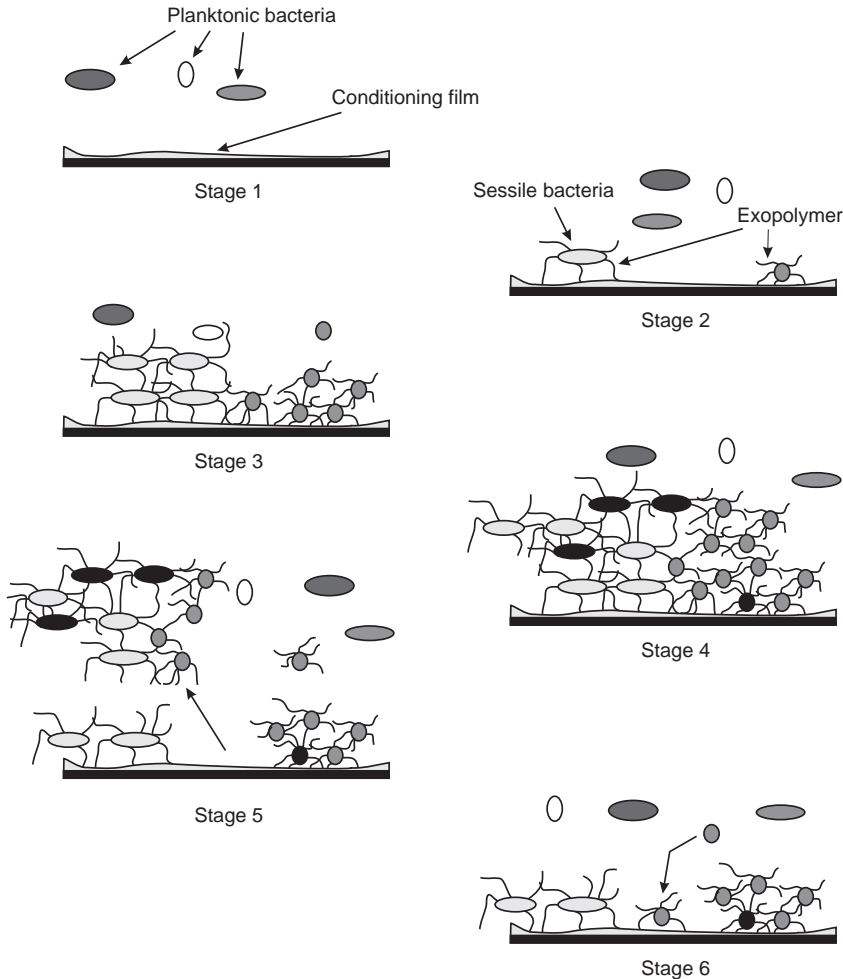


FIGURE 33.64 Different stages of biofilm formation and growth. Stage 1: Conditioning film accumulates on submerged surface. Stage 2: Planktonic bacteria from the bulk water colonize the surface and begin a sessile existence by excreting exopolymer that anchors the cells to the surface. Stage 3: Different species of sessile bacteria replicate on the metal surface. Stage 4: Microcolonies of different species continue to grow and eventually establish close relationships with each other on the surface. The biofilm increases in thickness. Conditions at the base of the biofilm change. Stage 5: Portions of the biofilm slough away from the surface. Stage 6: The exposed areas of surface are recolonized by planktonic bacteria or sessile bacteria adjacent to the exposed areas (Geesey, 1993).

are commercially available for use in pressurized systems (Gilbert and Herbert, 1987). In these devices, coupons are held in an assembly that mounts on a standard pressure fitting.

If the biofilm developing on the coupon is to be representative of the behavior in a system it is important for the sampling coupons to be made of a material similar to the system material and that they be flush mounted in the wall of the system to have the same

flow effects as those of the surrounding surface. While pressure fittings allow inserting coupons directly in process units, these fittings can be expensive. Pressure vessel codes and accessibility can also restrict possible locations. For these reasons, sidestream installations are often preferred.

Handling of field samples should be done carefully to avoid contamination with foreign matter including biological materials. Many different types of sterile sampling tools and containers are commercially available. Because many systems are anaerobic, proper sample handling and transport is essential to avoid exposure to oxygen from the air. One option is to analyze samples on the spot with special kits. Where transportation to a laboratory is required, Torbal jars or similar anaerobic containers can be used (Gerhardt, 1981). In many cases, simply placing samples directly in a large volume of the process water in a completely filled screw cap container is adequate.

Processing in the lab should also be done anaerobically using special techniques or in anaerobic chambers designed for this purpose. Because viable organisms are involved, processing should be done quickly to avoid growth or death of cells stimulated or inhibited by changes in temperature, oxygen exposure or other factors (Jack, 1999).

33.3.5.4 Biological Assessment Biological assays can be performed on liquid samples or on suspensions of solid deposits to identify and enumerate viable microorganisms, quantify metabolic or specific enzyme activity, or determine the concentration of key metabolites (Jack, 2002). Table 33.8 summarizes some of the methods that have been used to detect and describe microorganisms in terms of total cells present, viable cell numbers, and metabolic activity. Table 33.8 also identifies assays that can be used to establish the presence of biomass in a field sample, identify the organisms present, and assess the activity of enzymes, such as hydrogenase, that are thought to accelerate corrosion through cathodic depolarization. A more detailed description of these assays and other techniques is provided in the following sections.

Direct Inspection Direct inspection is best suited for the enumeration of planktonic organisms suspended in relatively clean water. In liquid suspensions, cell densities $>10^7$ cells cm^{-3} may cause the sample to appear turbid. Quantitative enumerations using a phase contrast microscopy can be done quickly using a counting chamber that holds a known volume of fluid in a thin layer. Visualization of microorganisms can be enhanced by fluorescent dyes that cause cells to light up under ultraviolet radiation. The technique is usually performed in the laboratory (Techniques for Monitoring Corrosion and Related Parameters in Field Applications, 1999).

Using a stain, such as acridine orange, cells separated by filtration from large aliquots of water can be visualized and counted on a 0.25- μm filter using the epifluorescent technique. Newer stains, such as fluorescein diacetate, 5-cyano-2,3-ditolyltetrazolium chloride, or *p*-iodonitrotetrazolium violet, indicate active metabolism by the formation of fluorescent products (Jack, 1999). Antibody fluorescence microscopy is similar to general fluorescent microscopy, except that the fluorescent dye used is bound to antibodies specific to SRBs. Only bacteria recognized by the antibodies fluoresce. Results can be analyzed within 2 h. The technique detects viable and nonviable bacteria, but it is limited to the type of SRB used in the manufacture of these antibodies (Techniques for Monitoring Corrosion and Related Parameters in Field Applications, 1999).

TABLE 33.8 Inspection, Growth, and Activity Assays for Microbial Populations

Assay	Method	Comments
Microorganisms	<i>Microscopic examination:</i> Cell numbers are obtained with a Petroff-Hausser counting chamber. Fluorescent dyes can light up specific microbes.	Requires a microscope with high magnification, phase contrast, and ultraviolet fluorescence, as appropriate. Cell counts are straightforward, but particulates and fluorescent materials may interfere.
Viable cell counts	<i>Most probable number:</i> Sample is diluted into a series of tubes of specific growth medium. Cell numbers based on growth at various dilutions.	Commercially available kits can be inoculated in the field, but growth takes days to weeks.
SRB APB Other organisms	<i>Dip slides:</i> A tab coated with growth medium is immersed, then incubated 2–5 days. Microbial colonies are counted visually.	
Identification of microorganisms	<i>Fatty acid methyl ester analysis:</i> Methyl esters of fatty acids from field sample are analyzed by gas chromatograph.	Crude numbers of bacteria, yeast, or fungi can be estimated in contaminated crankcase oil or fuels, for example. Commercial kits require minimal cost and expertise.
	<i>Probes based on nucleic acid base sequences:</i> Probes bind to DNA or RNA for specific proteins or organisms.	Fatty acid composition fingerprints specific organisms. The technique is available commercially.
	<i>Reverse sample genome probing:</i> DNA from MIC microbes is spotted on a master filter. Labeled DNA from field samples will bind to that DNA, if the reference organism is present.	Probes require special expertise and lab facilities.
	<i>Protein, lipopolysaccharide, nucleic acid analysis:</i> Established methods, widely available.	The assay answers the question, “Is this organism present?” It can also be used to track changes in a population after chemical treatment but it requires special expertise and equipment.
Biomass		Concentration of cell constituents correlates to level of organisms present in field sample.
Metabolic activity	<i>Adenosine triphosphate (ATP):</i> Fluorometer and supplies commercially available for field use.	ATP level correlates well with microbial metabolism level. Its concentration reflects the level of activity in a sample.
Sulfate reduction	<i>³⁵S sulfate reduction:</i> Radioactively labeled sulfate is incubated in field sample. Hydrogen sulfide formed is liberated by acid addition, trapped on a zinc acetate wick, and measured by scintillation counter.	A specialized lab technique useful in discovering nutrient sources for MIC and in quickly screening biocide activity in samples taken from the target system.
Enzyme activity	<i>Hydrogenase assay or sulfate reductase assay:</i> Measures enzyme activity in a field sample as a rate of reaction.	Commercial kits are available. Hydrogenase activity may be related to MIC, while sulfate reductase assesses the presence of SRB.

Identification of organisms can be accomplished by use of antibodies generated as an immune response to the injection of microbial cells into an animal, typically a rabbit. These antibodies can be harvested and will selectively bind to the target organism in a field sample. A second antibody tagged with a fluorescent dye is then used to light up the rabbit antibody bound to the target cells. In effect, the staining procedure can selectively light up target organisms in a mixed population or in difficult soil, coating or oily emulsion samples (Hunik et al., 1993).

Such techniques can provide insight into the location, growth rate, and activity of specific kinds of organisms in mixed biofilm populations. Antibodies that bind to specific cells can also be linked to enzymes that produce a color reaction in an enzyme-linked immunosorbent assay. The extent of color produced in solution can then be correlated with the number of target organisms present (Pope, 1992). While antibody based stains are excellent research tools, their high specificity means that they can only identify the target organisms. Other organisms potentially capable of causing problems are therefore missed.

Growth Assays The most common way to assess microbial populations in water samples is through growth tests using commercially available growth media for groups of organisms most commonly associated with industrial problems. These are packaged in a convenient form suitable for use in the field. Serial dilutions of suspended samples are grown on solid agar or liquid media. Based on the growth observed for each dilution estimates of the most probable number (MPN) of viable cells present in a sample can be obtained (Costerton and Colwell, 1977). The test can show some results in a few days, but the usual incubation period for the test is 14–28 days (Techniques for Monitoring Corrosion and Related Parameters in Field Applications, 1999).

However, despite the common use of these assays, only a small fraction of wild organisms actually grow in commonly available artificial media. Estimates for SRB in marine sediments, for example, indicate that only 1 in 1000 of the organisms present actually show up in standard growth tests (Jorgenson, 1978).

Activity Assays

WHOLE CELL Approaches based on the conversion of a radioisotopically labeled substrate can be used to assess the potential activity of microbial populations in field samples. This technique is specific to SRBs. It depends on bacterial growth for detection, but it generates results in ~2 days.

The sample is incubated with a known trace amount of radioactive-labeled sulfate. (SRBs reduce sulfate to sulfide.) After incubation, the reaction is terminated with acid to kill the cells and the radioactive sulfide is fixed in zinc acetate, which is sent to a laboratory for evaluation. This is a highly specialized technique, involving expensive laboratory equipment and the handling of radioactive substances (Techniques for Monitoring Corrosion and Related Parameters in Field Applications, 1999).

The radiorespirometric method that can use field samples directly without the need to separate organisms is very sensitive. Selection of the radioactively labeled substrate is key to interpretation of the results, but the method can provide insights into factors limiting growth by comparing activity in native samples with supplemented test samples under various conditions. Oil degrading organisms, for example, can be assessed through the mineralization of ^{14}C labeled hydrocarbon to carbon dioxide.

Radioactive methods are not routinely used by field personnel. However, they have been particularly useful in a number of applications including biocide screening programs, identification of nutrient sources, and assessment of key metabolic processes in various corrosion situations (Jack, 1999).

ENZYME-BASED ASSAYS An increasingly popular approach is the use of commercial kits to assay the presence of enzymes associated with microorganisms suspected to cause problems. For example, kits are available for the sulfate reductase enzyme common to SRB associated with corrosion problems (Odom et al., 1991). This technique takes advantage of the fact that SRBs reduce sulfate to sulfide through the presence of an enzyme, APS-reductase, common to all SRBs. Measurement of the amount of APS-reductase in a sample gives an estimate of total numbers of SRBs present. The test does not require bacterial growth and the entire test takes 15–20 min (Techniques for Monitoring Corrosion and Related Parameters in Field Applications, 1999).

Another example is the hydrogenase enzyme implicated in the acceleration of corrosion through rapid removal of cathodic hydrogen formed on the metal surface (Bryant et al., 1991). The test analyzes for the hydrogenase enzyme that is produced by bacteria able to use hydrogen as an energy source. The test is usually performed on sessile samples. The sample is exposed to an enzyme-extracting solution, and the degree of hydrogen oxidation in an oxygen-free atmosphere is detected by the addition of a dye (Techniques for Monitoring Corrosion and Related Parameters in Field Applications, 1999).

Performance of several of these kits has been assessed by field personnel in round robin tests. Correlation of activity assays and population estimates is variable. In general, these kits have a narrower range of application than growth-based assays making it important to select a kit with a range of response appropriate to the problem under consideration (Scott and Davies, 1992).

METABOLITES Adenosine triphosphate (ATP) is present in all living cells, but it disappears rapidly on death. The measure of ATP may thus provide a measure of living material. The ATP can be measured using an enzymic reaction, which generates flashes of light that are detected by a photomultiplier (Techniques for Monitoring Corrosion and Related Parameters in Field Applications, 1999). Commercial instruments are available that measure the release of light by the firefly luciferin–luciferase with ATP. The method is best suited to clean aerobic aqueous samples since suspended solids and chemical quenching can affect the results. Detection of metabolites, such as organic acids in deposits or gas compositions including methane or hydrogen sulfide by routine gas chromatography can also indicate biological involvement in industrial problems (Jack, 1999).

CELL COMPONENTS Biomass can be generally quantified by assays for protein, lipopolysaccharide, or other common cell constituents, but the information gained is of limited value. An alternate approach is to use cell components to define the composition of microbial populations with the hope that the insight gained may allow future damaging situations to be recognized and managed. Fatty acid analysis and nucleic acid sequencing provide the basis for the most promising methods in this category.

Fatty Acid Profiles Analyzing fatty acid methyl esters derived from cellular lipids can fingerprint organisms rapidly and, provided pertinent profiles are known, organisms in industrial and environmental samples can be identified with confidence. The immediate

impact of events, such as changes in operating conditions or the application of biocides, can be monitored by such analysis. Problem populations of certain organisms can also be identified in order to implement appropriate management responses in a timely fashion.

Nucleic Acid-Based Methods Specific DNA probes have been developed to detect segments of genetic material coding for known enzymes. A gene probe developed to detect the hydrogenase enzyme that occurs broadly in SRB from the genus *Desulfovibrio* has been tested on samples from an oilfield waterflood plagued with iron sulfide related corrosion problems. The enzyme was detected with this probe in 12 of 20 samples, suggesting that sulfate reducers that did not have this enzyme were also present in the operation (Voordouw et al., 1995). In principle, probes could be developed to detect potentially all sulfate-reducers. However, the operation of a battery of probes could be a daunting task where large numbers of field samples have to be analyzed.

To overcome this obstacle, the reverse sample genome probe (RSGP) was developed. With RSGP, the DNA from organisms previously isolated from field problems is spotted on a master filter following which the DNA isolated from field samples is labeled with either a radioactive or fluorescent indicator and exposed to the filter. Labeled DNA from the field sample sticks to the corresponding spot on the master filter when complementary strands of DNA are present. Organisms represented by the labeled spots are then known to be in the field sample (Voordouw et al., 1995).

Detailed Coupon Examinations A great deal of information can be learned by careful, in-depth examination of corrosion coupon surfaces using commonly available analytical techniques. A wide variety of samplers for introducing metallic surfaces of interest into a system are available. A popular sampling device is shown in Figure 33.65. Special handling of coupons after removal from the system being monitored is crucial to ensure that subsequent laboratory tests provide representative information. Biofilms in particular are highly sensitive to dehydration, exposure to air, temperature, mechanical damage, and other gross environmental changes that can occur during removal and transport of the coupon (Zintel et al., 2001).

Examination of coupons for microbial populations can be performed either directly or indirectly, using histological embedding techniques to preserve and remove the biofilm. Although fairly involved, the embedding technique offers several advantages over direct observation in that the biofilm and corrosion products are preserved for future analysis. Environmental scanning electron microscopy (ESEM) may also be utilized to examine biofilms on test coupons; however, exopolymers and corrosion products often obscure the cells, making quantification and identification difficult with this method.

33.3.5.5 Monitoring MIC Effects The presence of a biofilm on a metallic surface can greatly alter the local corrosion processes. In addition to the electrochemical changes that affect corrosion, biofilms can also modify other readily measured characteristics, such as pressure drop or heat-transfer resistance. Monitoring such microbiological influences can provide a useful indicator that a biofilm is present and that action should be taken to mitigate potential MIC.

Deposition Accumulation Monitors Methods for monitoring deposits can provide an indication of the accumulation of biofilm and other solids on surfaces or in orifices. For example, monitoring pressure drop across an orifice provides a simple method for

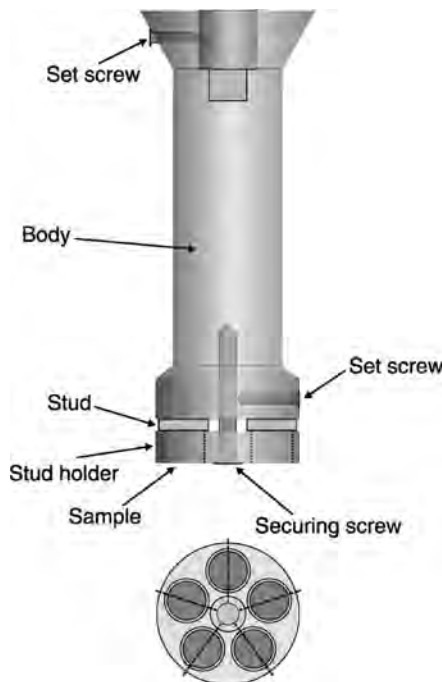


FIGURE 33.65 Biofilm sampling device with removable “buttons.”

continuous monitoring of deposit accumulation and biofilm accumulation. The main disadvantage of the pressure drop technique is that it is not specific to the biofilm build-up since it detects the total scaling and deposition effects in a line (Zintel et al., 2001).

These measurements can be made on actual operating units on-line, but they may also be done using model heat exchanger units or instrumented pipe loops run in parallel to system flow. Figure 33.17 shows such an instrumented pipe loop test unit with five parallel, instrumented pipe runs. Water flow from the target system is diverted through this unit, so that conditions are representative of the actual operating system (Jack, 2002).

Measurements of friction factors and heat exchange efficiencies can indicate fouling. While deposits of any sort can affect flow and heat transfer in an operating industrial system, biofilms are especially effective. A 165- μm thick biofilm shows 100 times the relative roughness of a calcite scale and a thermal conductivity close to that of water, that is, almost 100 times less than carbon steel (Jack, 1999). The assumption is generally made that a susceptible system showing extensive fouling is prone to MIC. Systems showing the effects of extensive fouling are operating inefficiently and may warrant remedial action, in any case.

Electrochemical Methods An electrochemical method for on-line monitoring of biofilm activity has been developed for continuous monitoring of biofilm formation without the need for excessive involvement of plant personnel (Figure 33.66a and b). A series of stainless steel or titanium disks are exposed to the plant environment. One set of disks is

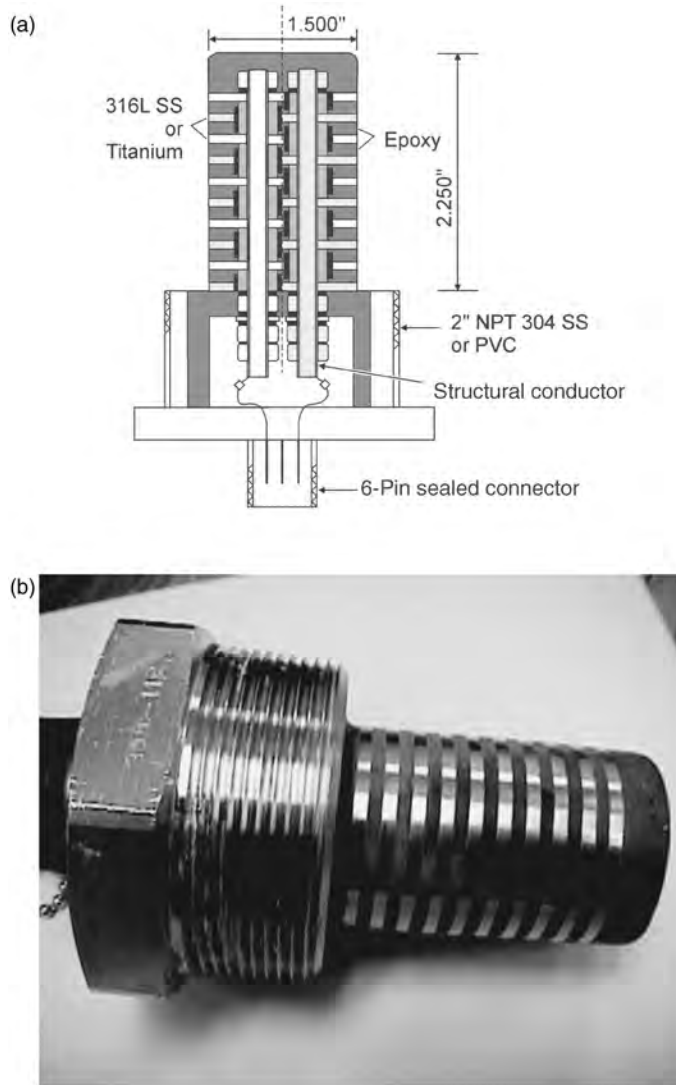


FIGURE 33.66 (a) Schematic and (b) picture of the BioGEORGE MIC detection probe. (Courtesy of George Licina, Structural Integrity Associates, Inc.)

polarized (relative to the other set) for a short period of time each day. The electrodes are connected through a shunt the remainder of the time. Biofilm activity, which is also an electrochemical process, is monitored by tracking changes in the applied current when the external potential is on and the generated current when the potential is off (Zintel et al., 2001).

The onset of biofilm formation on the probe is indicated when either of these independent indicators deviates from the baseline level (Figure 33.67). Such a departure would then trigger the alarm located in a control box (Figure 33.68). The level of biofilm activity is also measured by the amount of variation from the baseline. The applied and generated

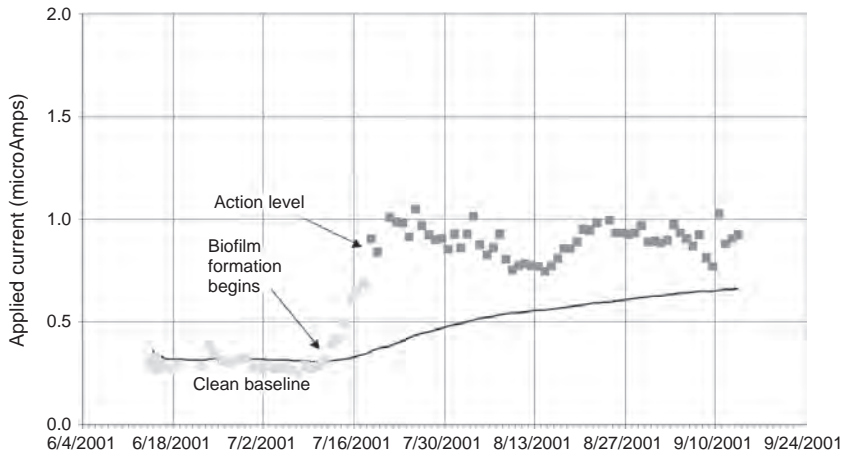


FIGURE 33.67 A plot of signals generated by a BioGEORGE MIC detection probe. (Courtesy of George Licina, Structural Integrity Associates, Inc.)

currents from a well-controlled system will be a flat line, devoid of any significant deviations.

Another experimental approach to detect MIC with an electrochemical signals is based on the use of small silver sulfide and silver chloride electrodes capable of detecting sulfides or chlorides by changes in the potential between the Ag/AgCl or



FIGURE 33.68 Control box of a BioGEORGE MIC detection probe. (Courtesy of George Licina, Structural Integrity Associates, Inc.)

Ag/Ag₂S electrode pairs (Zintel et al., 2001). The production of sulfides or concentration of chlorides by microbial action was shown to produce a nearly instantaneous change in the potential of the electrode pair, thus providing an indication of specific microbial activity.

33.3.6 Monitoring Pipeline CP Systems

Monitoring pipeline CP systems is a technically complex field. In many cases, condition monitoring requirements are specified by regulatory authorities. Since CP systems are expected to operate in demanding environmental conditions over long time periods, the reliability requirements of the associated hardware are high. Regular monitoring of the equipment is therefore an important aspect of any CP program.

An increasing trend toward selective remote rectifier monitoring, using modern communication systems and computer networks, is emerging to accomplish these tasks with reduced resources. Wireless cell phone and satellite communication systems are available for interrogating rectifiers in remote locations. The other aspect of monitoring a pipeline CP system is the monitoring of protection afforded to the pipeline. The following sections briefly describe some of the techniques used to carry out this work.

33.3.6.1 Close Interval Potential Surveys The principle of a close interval potentials surveys (CIPS) is to record the potential profile of a pipeline over its entire length by taking potential readings at ~ 1 m intervals. A reference electrode is connected to the pipeline at a test post and positioned in the ground over the pipeline at regular intervals for the measurement of the potential difference between the reference electrode and the pipeline (Figure 33.69).

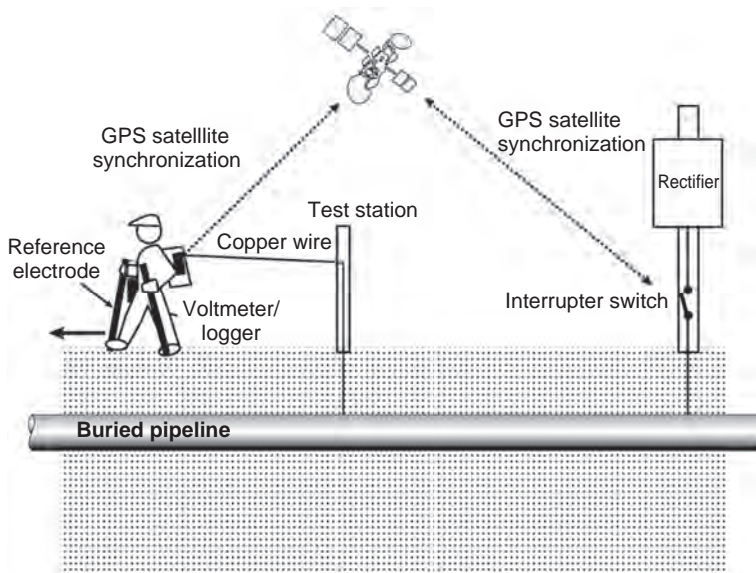


FIGURE 33.69 Schematic description of the CIPS methodology.

A three-person crew is typically required to perform these measurements. One person walking ahead locates the pipeline with a pipe locator to ensure that the potential measurements are performed directly above the pipeline. This person also carries a tape measure and inserts a distance marker (a small flag) at regular intervals over the pipeline. A second person carries a pair of electrodes connected to the test post by means of a trailing thin copper wire and the potential measuring instrumentation. The second person is also responsible for entering specific features as a function of the measuring distance. The third person collects the trailing wire, after individual survey sections have been completed.

The CIPS measurements are quite demanding on field crews and require extensive logistical support from both the pipeline operator and the CIPS contractor. Field crews are typically required to move over/around fences, roads, highways and other obstacles and difficult terrain. Breakage of the trailing copper wire is not uncommon and special strengthened wire has to be taped down onto road surfaces that are crossed.

An important consideration in the potential readings is the IR or ohmic drop error that is included in the potential measurements when a CP system is operational. A commonly used method to correct for the IR drop, often called making instant-off measurements, can be a very accurate way to make potential measurements. In reality, however, making measurements completely free of voltage drop error is not always possible because some currents cannot be easily interrupted.

Uninterruptible current sources may include sacrificial anodes directly bonded to the structure, foreign rectifiers, stray currents, telluric currents, and long-line cells (Ansuini and Dimond, 2005). Modern interrupters are based on solid-state switches and are programmable to perform switching only when the survey is performed during the day time. This feature minimizes the depolarization of the pipeline that may occur gradually due to the cumulative effects of the “off” periods.

When several rectifiers protect a structure, it is necessary that all rectifiers be interrupted at the exact same instant in order to obtain meaningful measurements. Pipeline operators usually specify that at least two rectifiers ahead of the survey team and two rectifiers behind the survey team have to be interrupted in a fully synchronized manner. The amount of time between current interruption and depolarization can vary from a fraction of a second to several seconds, depending on the details of the structure. In addition, capacitive spikes that occur shortly after current is interrupted may mask the instant-off potential. Measurements made with a recording voltmeter are preferred as they can be subsequently analyzed to determine the real instant-off potential (Ansuini and Dimond, 2005).

An example of graphical CIPS data is presented in Figure 33.70 (Bianchetti, 2001). In the simplest format, the “on” and “off” potentials are plotted as a function of distance. The usual sign convention is for potentials to be plotted as positive values. The difference between the “on” and “off” potential values should be noted. As is usually the case, the “off” potentials are less negative than the “on” values. When the relative position of these two lines is reversed, it indicates that some unusual conditions such as stray current interference may be at play.

The CIPS technique provides a complete pipe-to-soil potential profile and the interpretation of results, including the identification of defects, is relatively straightforward.

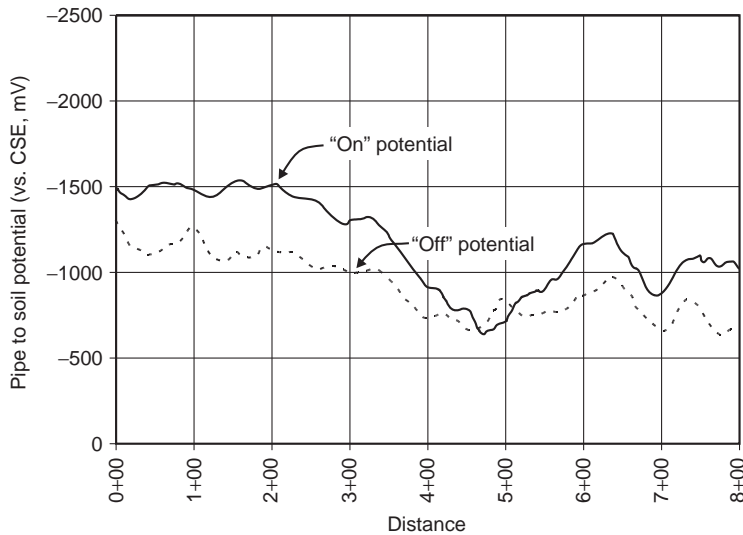


FIGURE 33.70 Over-the-line survey with cathodic protection.

33.3.6.2 Pearson Survey The Pearson Survey, named after its inventor, is used to locate coating defects in buried pipelines. Once these defects have been identified, the protection levels provided by the CP system can be investigated at these critical locations in more detail.

During a Pearson survey, an ac signal of ~ 1000 Hz is imposed onto the pipeline by means of a transmitter (Figure 33.71), which is connected to the pipeline and an earth spike as illustrated in Figure 33.72. Two survey operators make earth contact either



FIGURE 33.71 Instrumentation to carry out a Pearson survey. (Photo courtesy of Tinker & Rasor.)

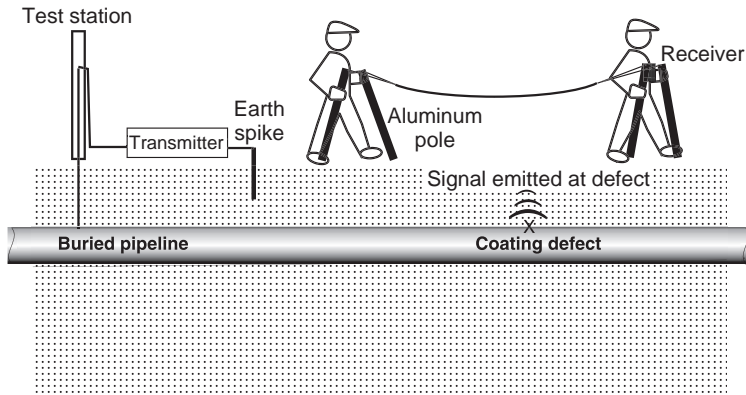


FIGURE 33.72 Schematic description of the Pearson survey technique.

through aluminum poles or metal studded boots (Figure 33.73). A distance of several meters typically separates the operators. Essentially, the signal measured by the receiver provides a measure of the potential gradient over the distance between the two operators. Defects are located by a change in the potential gradient, which translates into a change in signal intensity.

As in the CIPS technique, the measurements are usually recorded by walking directly over the pipeline. As the front operator approaches a defect, increasing signal intensity is recorded. As the front person moves away from the defect, the signal intensity drops and later picks up again as the rear operator approaches the defect. The interpretation of signals may obviously be more difficult when several defects are located between the two operators.

In principle, a Pearson survey can be performed with an impressed CP system still energized. However, sacrificial anodes should be disconnected since the signal from these may otherwise mask actual coating defects. A three-person team is usually required to locate the pipeline, perform the survey measurements, place defect markers into the ground, and move the transmitters periodically.



FIGURE 33.73 Pipeline CP fault testing with Pearson-type detector. (Photo courtesy of Tinker & Rasor.)

By walking the entire length of the pipeline, an overall inspection of the right of way can be made together with the measurements. In principle, all significant defects and metallic conductors causing a potential gradient will be detected. There are no trailing wires and the impressed CP current does not have to be turned on and off.

The disadvantages associated with Pearson surveys are similar to those of CIPS, as the entire pipeline has to be walked and contact established with the ground. The technique is therefore impractical in many areas, such as roads, paved areas, or rivers.

33.3.6.3 Direct Current Voltage Gradient Surveys The direct current voltage gradient (DCVG) surveys are a more recent method to locate defects on coated buried pipelines and to assess their severity. The technique again relies on the fundamental effect of a potential gradient being established in the soil at coating defects under the application of CP current. In general, the greater the size of the defect, the greater the potential gradient.

The potential gradient is measured by an operator between two reference electrodes, for example, copper sulfate electrodes, separated by a distance of approximately one-half of a meter. A pulsed dc signal is imposed on the pipeline for DCVG measurements. The pulsed input signal minimizes interference from other current sources (e.g., CP systems, electrified rail transit lines, telluric effects). This signal can be obtained with an interrupter on an existing rectifier or through a secondary current pulse superimposed on the existing “steady” CP current.

The operator walking the pipeline observes voltage deflections on a precision voltmeter to identify defect locations. The presence of a defect is indicated by an increased deflection as the defect is approached, no deflection when the operator is immediately above the defect and a decreasing deflection as the operator walks away from the defect (Figure 33.74). The high precision in locating defects (~ 0.1 to 0.2 m) represents a major advantage in minimizing the work of subsequent digs if corrective action needs to be taken.

The DCVG technique is particularly suited to complex CP systems, for example, areas with a relatively high density of buried structures. These are generally the most difficult survey conditions. The DCVG equipment is relatively simple and involves no trailing wires.

33.3.6.4 Corrosion Coupons Cathodic protection (CP) coupons are now being used as an alternative method to make potential measurements that may be substantially free of voltage drop error. A CP coupon is a small piece of metal that is electrically connected to the structure at a test station. The potential of a coupon will closely approximate the potential of any exposed portion of the structure (holiday) located in the vicinity of the coupon. By disconnecting the coupon from the structure at the test station, an instant-off potential measurement can be made on the coupon without having to interrupt any other current sources. However, these measurements are still not completely free of voltage drop error. Any voltage drop occurring in the electrolyte in the distance between the reference electrode and the coupon surface will still be incorporated into measurements. Placing a reference electrode as close as possible to a coupon can minimize this error. However, the reference electrode must not be placed so close that it shields the coupon (Ansuini and Dimond, 2005).

Perhaps the most important consideration in the installation of corrosion coupons is that a coupon must be representative of the actual pipeline surface/defect. The exact

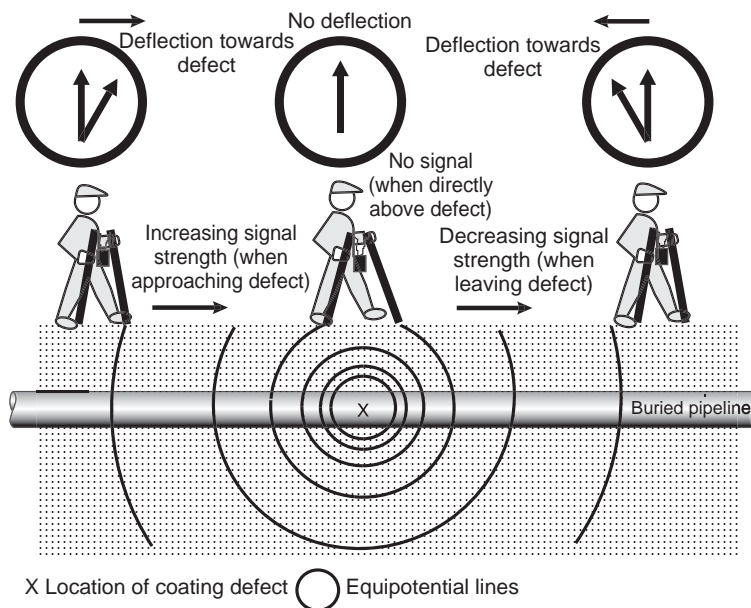


FIGURE 33.74 Schematic description of the DCVG methodology.

metallurgical detail and surface finish as found on the actual pipeline are therefore required on the coupon. The influence of corrosion product build-up may also be an important factor. The environmental conditions of the coupon have to be matched with those of the pipe being monitored, for example, temperature, soil conditions, soil compaction, and oxygen concentration.

Measurement of current flow to/from the coupon and its direction can also be determined, for example, by using a shunt resistor in the bond wire. Importantly, it is also possible to determine corrosion rates from the coupon. Electrical resistance sensors provide an option for *in situ* corrosion rate measurements, as an alternative to weight loss coupons.

33.3.7 Atmospheric Corrosion Monitoring

The economic losses caused by atmospheric corrosion are tremendous and account for the disappearance of a significant portion of metal produced. Atmospheric corrosion has been reported to account for more failures in terms of cost and tonnage than any other type of material degradation processes. Atmospheres have been classified into four basic types:

Industrial: An industrial atmosphere is characterized by its pollution level, two constituents of which are particularly corrosive (i.e., sulfur compounds and nitrogen oxides). When present, these contaminants may combine with dew or fog to produce a highly corrosive acid film on exposed surfaces.

Marine: Marine atmospheres are often laden with a sea salt mist or aerosol particles that can be carried by the wind to settle on exposed surfaces. The quantity of salt

contamination typically decreases with distance from the saline source, and is greatly affected by wind currents.

Rural: A rural environment does not usually contain strong chemical contaminants, but it does contain organic and inorganic dusts. Its principal corrosive constituent is moisture. In tropical environments, in addition to the high average temperature, the daily cycle includes a high relative humidity, intense sunlight, and long periods of condensation during the night.

Indoor: Indoor atmospheres can be generally considered to be quite mild. However, the presence of local contaminants can transform ambient air into a very corrosive environment if it is not properly ventilated or controlled through dehumidification, for example. Additionally, the consequences of indoor corrosion have been drastically amplified by the incredibly small volume of the material that needs to be damaged before causing a fault in modern computer and communication equipment. The microchip in an automobile, for example, is not directly subjected to the same environmental hazards as the car body. However, the tolerance for corrosion loss in electronic devices is many orders of magnitude less, that is, on the order of picograms (10^{-12} g). The minimum line width in the state-of-the-art printed circuit boards (PCBs) in 1997 was already $<100\text{ }\mu\text{m}$. On hybrid integrated circuits (HICs), line spacing may be $<5\text{ }\mu\text{m}$ (Frankenthal, 2000).

Various methods have been developed for measuring the factors that influence atmospheric corrosion. Temperature, RH, wind direction and velocity, solar radiation, and amount of rainfall are easily recorded. Not so easily determined are dwelling time of wetness (TOW), and the quantity of sulfur dioxide and chloride contamination. However, methods for these determinations have been developed and are in use at various test stations.

By monitoring these factors and relating them to corrosion rates, a better understanding of atmospheric corrosion can be obtained for planning maintenance procedures and inspection schedules, for the prediction or outdoor performance of materials and systems, and to circumvent the effects of corrosion products on the environment.

33.3.7.1 Relative Humidity and Time of Wetness Relative humidity is defined as the ratio of the quantity of water vapor present in the atmosphere to the saturation quantity at a given temperature, and it is expressed as percent (%). A fundamental requirement for atmospheric corrosion processes is the presence of a thin-film electrolyte that can form on metallic surfaces when exposed to a critical level of humidity (Figure 33.75). While this film is almost invisible, the corrosive contaminants it contains are known to reach relatively high concentrations, especially under conditions of alternate wetting and drying.

The critical humidity level is a variable that depends on the nature of the corroding material, the tendency of corrosion products and surface deposits to absorb moisture, and the presence of atmospheric pollutants (Roberge, 2000). It has been shown that, for example, this critical humidity level is 60% for iron if the environment is free of pollutants.

In the atmospheric classification scheme described in ISO 9223 (Corrosion of metals and alloys—Corrosivity of atmospheres—Classification, 1992). The TOW is an estimated parameter based on the length of time when the relative humidity is $>80\%$ at a temperature $>0^{\circ}\text{C}$. It can be expressed as the hours or days per year or the annual percentage of time. Experience from applying the ISO classification system has shown, however, that

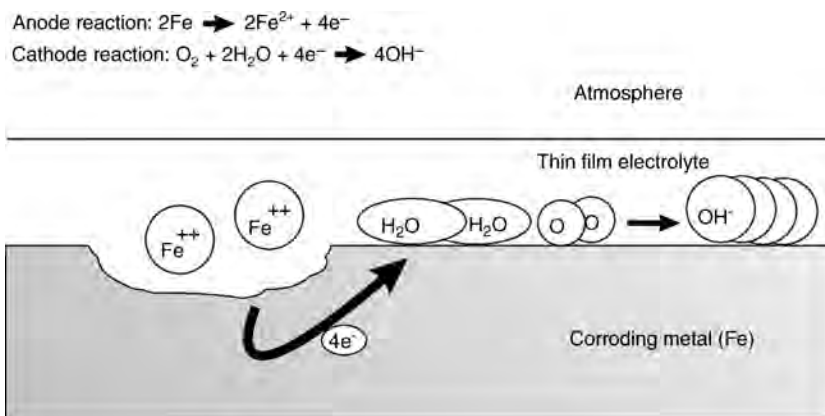


FIGURE 33.75 Schematic description of the atmospheric corrosion of iron.

certain observations need further clarification. Substantial corrosion has also been measured on specimens exposed at temperatures well below 0°C , which is in direct contradiction to the ISO criterion (King et al., 2001). It has been proposed on the basis of these results, that the TOW in cold environments be estimated as the length of time that relative humidity exceeds 50% and the ambient temperature exceeds -10°C .

A method of measuring the TOW has been developed by Sereda and correlated with the corrosion rates encountered in the atmosphere (Sereda et al., 1982). The moisture sensing elements in this sensor are manufactured by plating and selective etching of thin films of appropriate anode (copper) and cathode (gold) materials in an interlaced pattern on a thin nonconductive substrate (Figure 33.76). When moisture condenses on the sensor it activates the sensor, producing a small voltage (0–100 mV) across a $10^7 \Omega$ resistor.

Thin sensing elements are preferred in order to preclude influencing the surface temperature to any extent. Although a sensor constructed using a 1.5-mm thick glass

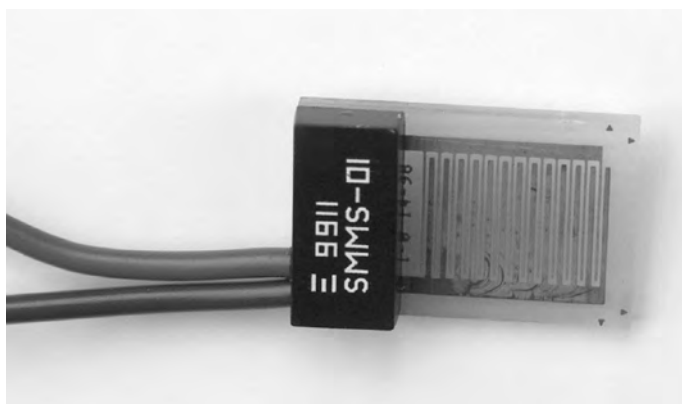


FIGURE 33.76 Interlocking combs of gold and copper electrodes in a ‘Sereda’ humidity sensor. (Courtesy Kingston Technical Software.)

reinforced polyester base has been found to be satisfactory on plastic surfaces, this will not be the case with the same sensing element on a metal surface with a high-thermal conductivity (Standard Practice for Measurement of Time-of-Wetness on Surfaces Exposed to Wetting Conditions as in Atmospheric Corrosion Testing, 1999). For metal surfaces, the sensing element should be appreciably thinner. Commercial epoxy sensor backing products of thickness of 1.5 mm, or less, are suitable for this purpose.

33.3.7.2 Pollutants Sulfur dioxide (SO_2), which is the gaseous product of the combustion of fuels that contain sulfur such as coal, diesel fuel, gasoline, and natural gas, has been identified as one of the most important air pollutants that contribute to the corrosion of metals.

Less recognized as corrosion promoters, are the nitrogen oxides (NO_x), which are also products of combustion. A major source of NO_x in urban areas is the exhaust fumes from vehicles. Sulfur dioxide, NO_x , and airborne aerosol particles can react with moisture and ultraviolet (UV) light to form new chemicals that can be transported as aerosols. A good example of this is the summertime haze or smog over many large cities. Up to 50% of this haze is a combination of sulfuric and nitric acids.

Sulfur dioxide is usually measured in terms of its concentration in air in units of $\mu\text{g}/\text{m}^3$. Precise methods are available to monitor continuously the amount of sulfur dioxide in a given volume of air. However, this is only indirectly related to the effect of sulfur dioxide on corrosion since only the actual amount of hydrated sulfur dioxide or sulfur trioxide deposited on metal surfaces is important.

Since it is the SO_2 deposited on the metal surface that affects the corrosion, deposited SO_2 has been measured as its reaction product in terms of sulfate deposition rate on the surface in units of mg/m^2 day. The pollution levels can also be measured in terms of the concentration of the dissolved sulfate (SO_4^{2-}) in rain water.

33.3.7.3 Airborne Particles (Chlorides) The behavior of aerosol particles in outdoor atmospheres can be explained by invoking the laws that govern their formation, movement, and capture. These particles are present throughout the planetary boundary layer and their concentrations depend on a multitude of factors including location, time of day or year, atmospheric conditions, presence of local sources, altitude, and wind velocity. Size is normally used to classify aerosol because it is the most readily measured property and other properties can be inferred from size information (Hidy, 1984). The highest mass fraction of particles in an aerosol is characterized by particles having a diameter in the range of 8–80 μm (Feliu et al., 1999). Some studies have also indicated that there is a strong correlation between wind speed and the deposition and capture of aerosols. In such a study of saline winds in Spain, a very good correlation was found between chloride deposition rates and wind speeds above a threshold of 3 m/s or 11 km/h (Morcillo et al., 2000).

The lifetime of any particular particle depends on its size and location. Studies of the migration of aerosols inland of a sea coast have shown that typically the majority of the aerosol particles are deposited close to the shoreline (typically 400–600 m) and consist of large particles ($>10\text{-}\mu\text{m}$ diameter), which have a short residence time and are controlled primarily by gravitational forces (Feliu et al., 1999; Morcillo et al., 2000). The aerosols that form also have mass and are subject to the influence of gravity, wind resistance, drop-let dry-out, and possibilities of impingement on a solid surface, as they progress inland.

Airborne salinity refers to the content of gaseous and suspended salt in the atmosphere. It is measured by its concentration in air in units of $\mu\text{g m}^{-3}$. Since it is the salt that is deposited on the metal surface that affects the corrosion, it is also often measured in terms of deposition rate in units of $\text{mg/m}^2 \text{ day}$. The chloride levels can also be measured in terms of the concentration of the dissolved salt in rain water.

A number of methods have been employed for determining the contamination of the atmosphere by aerosol transported chlorides (e.g., sea salt and road deicing salts). The “wet candle method”, for example, is relatively simple, but has the disadvantage that it also collects particles of dry salt that might not be deposited otherwise (Standard Test Method for Determining Atmospheric Chloride Deposition Rate by Wet Candle Method, 2002). This technique uses a wet wick of a known diameter and surface area to measure aerosol deposition (Figure 33.77). The wick is maintained wet using a reservoir of water or 40% glycol–water solution. Particles of salt or spray are trapped by the wet wick and retained. At intervals, a quantitative determination of the chloride collected by the wick is made and a new wick is exposed.

In reality, the wet candle method gives an indication of the salinity of the atmosphere rather than the contamination of exposed metal surfaces. The technique is considered to measure the total amount of chloride arriving to a vertical surface and its results may not be truly significant for corrosivity estimates.

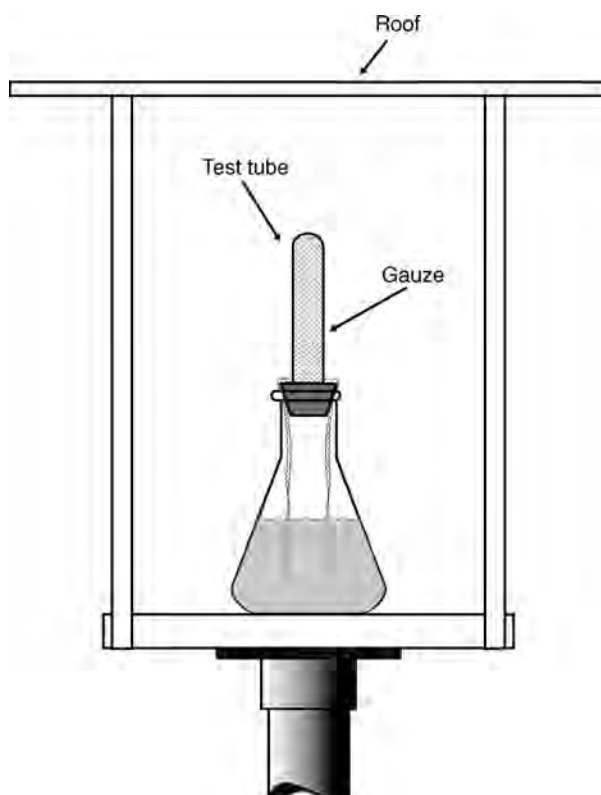


FIGURE 33.77 Schematic of a wet candle chloride apparatus.



FIGURE 33.78 A typical subfloor simulation box surmounted by a wet candle chloride measuring device that has been installed at several sites in Australia and New Zealand. (Courtesy of Branz, New Zealand.)

In order to evaluate the chloride deposition rates in a confined space, such as a ventilated subfloor where corrosion of fasteners can be quite severe, a special collecting box shown in Figure 33.78 was used in which the airborne chlorides were collected on horizontal and vertical filter papers positioned at different locations from the box openings (Figure 33.79).

For obvious reasons, the test chambers used in this study were thoroughly cleaned of chlorides before setting them outdoor. Rigid plastic material panels stuck at predetermined positions with self-adhesive pads were used to hold the filter papers with chloride-free plastic paper clips. The roof protected the surface from direct rain, but the filter papers were still exposed to deposition of the airborne chlorides and possibly some rain. The paper was removed every month for chemical analysis. A wet candle with its own roof cover was installed on each box (Figure 33.78) to provide a measure of the level of chlorides outside the boxes.

33.3.7.4 Atmospheric Corrosivity The simplest form of direct atmospheric corrosion measurement is by coupon exposures. Subsequent to their exposure, the coupons can be subjected to weight loss measurements, pit density and depth measurements, and to other types of examination. Flat panels exposed on exposure racks are a common coupon-type device for atmospheric corrosivity measurements. Various other specimen configurations have been used, including stressed U-bend or C-ring specimens for SCC studies. The main drawback associated with conventional coupon measurements is that extremely long exposure times are usually required to obtain meaningful data, even on a relative scale. It is not uncommon for such programs to run for 20 years or longer (Roberge, 2006).



FIGURE 33.79 Horizontal and vertical filter papers installed inside a subfloor simulation box. (Courtesy of Branz, New Zealand.)

Some variations of the basic coupon specimens can provide rapid material–corrosivity evaluations. The helical coil adopted in the ISO 9226 methodology is a high surface area–weight ratio coupon that gives a higher sensitivity than panel coupons of the same material (Corrosion of metals and alloys–Corrosivity of atmospheres–Determination of corrosion rate of standard specimens for the evaluation of corrosivity, 1992). The use of bimetallic specimens in which a helical wire is wrapped around a coarsely threaded bolt can provide additional sensitivity and forms the basis of the CLIMAT coupon that was developed to classify industrial and marine atmospheres (Doyle and Wright, 1982).

An ASTM standard describes the construction of CLIMAT coupons. According to this standard, CLIMAT coupons can be made from 1100 aluminum (UNS A91100) wire wrapped around threaded rods of nylon, 1010 mild steel (UNS G10100 or G10080), and CA110 copper (UNS C11000) (Standard Practice for Conducting Wire-on-Bolt Test for Atmospheric Galvanic Corrosion, 1999). The mass loss of aluminum wire after 90 days of exposure is considered to be an indication of atmospheric corrosivity. However, the relative corrosivity of atmospheres could be quite different for the various combinations of materials.

The aluminum wire on copper bolts has been found by many to be the most sensitive of the three proposed arrangements in the ASTM standard. This was corroborated in a recent paper that clearly demonstrated the superior sensitivity of the Al–Cu arrangement for all types of simulated environments, as is shown in Table 33.9 (Calderon and Arroyave, 2005). While this arrangement is the most sensitive, the use of triplicate coupons on a single holder additionally provides an indication of the reproducibility of the measurements.

TABLE 33.9 Mass Loss of Aluminum Wire Exposed in Different Environments

Environment	Helical Coil	Mass Loss (%)		
		Al–Nylon	Al–Cu	Al–Fe
Rural	0.02		0.88	0.19
Urban	0.03	0.060	0.77	0.25
Marine	0.03	0.066	5.48	3.9
Marine–urban	0.03		4.29	2.83

Such a CLIMAT coupon with three copper rods installed at the NASA Kennedy Space Center (KSC) beach corrosion test site (Figure 33.80) is shown immediately after it had been installed (Figure 33.81a), after 30 days (Figure 33.81b), and after 60 days (Figure 33.81c). The KSC has the highest corrosivity of any test site in the continental United States (Coburn, 1978). The mass loss recorded after a shorter exposure than usual can be very high. In the present example, it was already 16% of the original aluminum wire after only 60 days.

The CLIMAT coupons sensitivity to atmospheric corrosivity can be used to study fluctuations on a microenvironmental scale (Klassen et al., 2002). In the following example, a supporting panel (Figure 33.82) was placed ~2 m above the ground on a pedestrian bridge concrete support ~4 m from a moderately trafficked highway during the winter months when deicing salts are used. Six sets of CLIMAT units were deployed on the panel with each set in a different microenvironment produced by various baffle



FIGURE 33.80 Aerial view of the NASA Kennedy Space Center beach corrosion test site where atmospheric corrosivity is the highest corrosivity of any test site in the continental United States. (Courtesy of NASA.)

geometries. The overall atmospheric conditions, for example, temperature and relative humidity, were therefore the same for each set except for the differences in the rate of aerosol deposition.

The average mass loss experienced by the CLIMATs exposed in different baffling geometries is also shown in Figure 33.82. One obvious conclusion from these measurements is that shielding, whether from wind or direct precipitation, can dramatically reduce the corrosion rate of components exposed to the same time-of-wetness factor. In fact, there was a 24-fold difference between the average mass loss in the boxed-in and boldly exposed coupons.

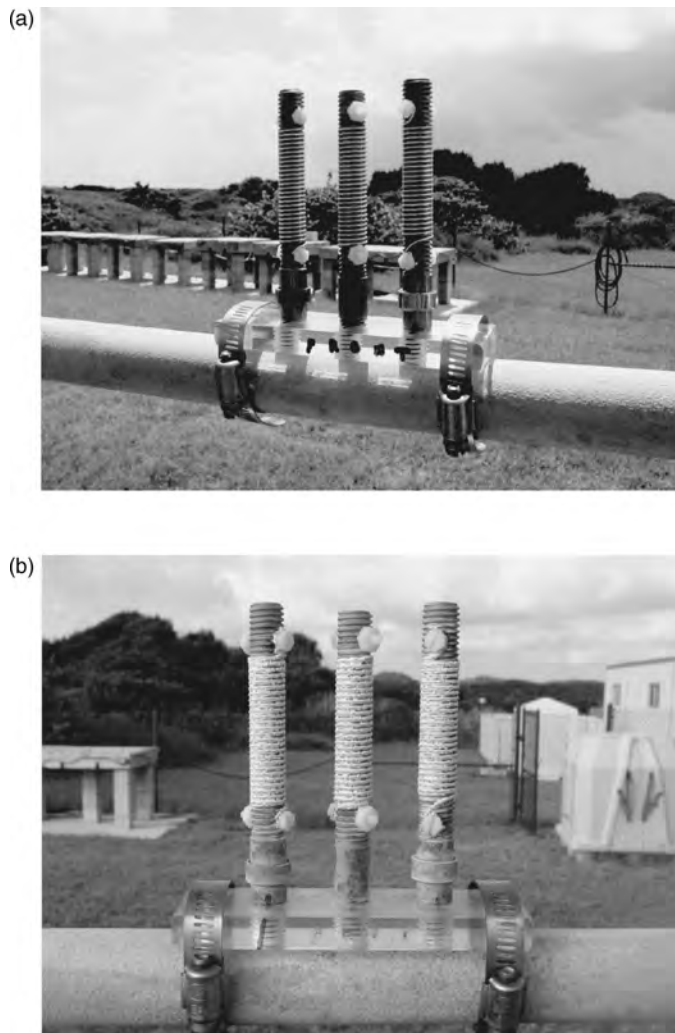


FIGURE 33.81 A CLIMAT coupon with three copper rods immediately after it was installed at the Kennedy Space Center beach corrosion test site (a), after 30 days (b), and after 60 days (c). (Courtesy of NASA.)



FIGURE 33.81 (Continued)

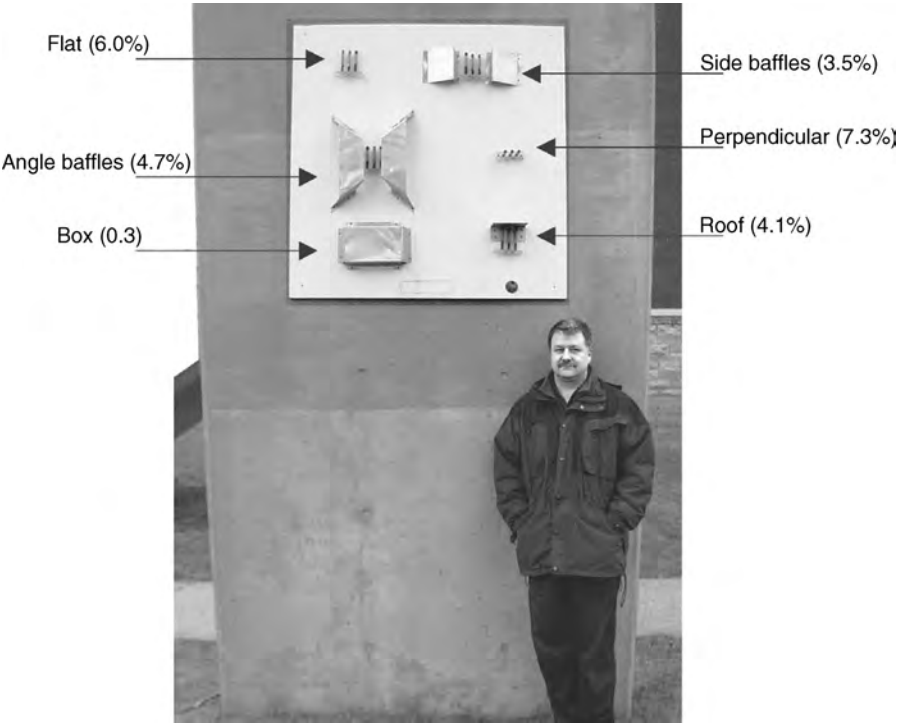


FIGURE 33.82 Test panel and CLIMAT coupons exposed in different microenvironments adjacent to a moderately trafficked highway during the winter months when deicing salts are used.

REFERENCES

- Ansuini FJ, Dimond JR. *Field Tests on an Advanced Cathodic Protection Coupon*. CORROSION 2005, Paper No. 39. Houston, TX: NACE International; 2005.
- Davis GD, Raghu S, Carkhuff BG, Garra F, Srinivasan R, Phillips TE. Corrosion Health Monitor for Ground Vehicles. Paper No. 103. 2005. Tri-Service Corrosion Conference, Orlando, FL, November 14–18 p. 2005.
- Bianchetti RL. Survey Methods and Evaluation Techniques. In: *Peabody's Control of Pipeline Corrosion*. Houston, TX: NACE International; 2001. p. 65–100.
- Bosch RW, Hubrecht J, Bogaerts WF, Syrett BC. Electrochemical frequency modulation: a new electrochemical technique for online monitoring. *Corrosion* 2001;57:60–70.
- Boukamp BA. *Equivalent Circuit (Equivrt.PAS) Users Manual*. Report CT89/214/128. The Netherlands: University of Twente; 1989.
- Bryant RD, Jansen W, Boivin J, Laishley EJ, Costerton JW. Effect of hydrogenase and mixed sulfate-reducing bacterial populations on the corrosion of steel. *Applied and Environmental Microbiology* 1991;57:2804–2809.
- Calderon JA, Arroyave CE. A laboratory approach to the mechanism of attack on the wire-on-bolt device used for atmospheric corrosion studies. *Corrosion* 2005;61:99–110.
- Coburn S. Atmospheric Corrosion. In: *Metals Handbook, 9th ed., Vol. 1, Properties and Selection, Carbon Steels*. Metals Park, Ohio: American Society for Metals; 1978. p. 720.
- Corrosion of metals and alloys—Corrosivity of atmospheres—Classification. ISO 9223. 1992. Geneva, Switzerland, International Standard Organization (ISO).
- Corrosion of metals and alloys—corrosivity of atmospheres—determination of corrosion rate of standard specimens for the evaluation of corrosivity. ISO 9226. 1992. Geneva, Switzerland, International Standard Organization (ISO).
- Costerton JW, Colwell RR. *Native Aquatic Bacteria: Enumeration, Activity and Ecology*. [STP 695] Philadelphia, PA: American Society for Testing and Materials; 1977.
- Cottis RA, Al-Awadhi MAA, Al-Mazeedi H, Turgoose S. Measures for the detection of localized corrosion with electrochemical noise. *Electrochimica Acta* 2001;46:3665–3674.
- Cottis RA, Turgoose S. *Corrosion Testing Made Easy: Electrochemical Impedance and Noise*. Houston, TX: NACE International; 1999.
- Dean SW. Overview of corrosion monitoring in modern industrial plants. In: Moran GC, Labine P, editors. *Corrosion Monitoring in Industrial Plants Using nondestructive testing and Electrochemical methods*. ASTM STP 908. Philadelphia: American Society for testing and Materials; 1986. p. 197–220.
- Dean SW. Corrosion monitoring for industrial processes. In: Cramer DS, Covino BS, editors. *Volume 13A: Corrosion: Fundamentals, Testing, and Protection*. Metals Park, OH: ASM International; 2003. p. 533–541.
- Denzine AF, Reading MS. An improved, rapid corrosion rate measurement technique for all process environments. *Materials Performance* 1998;37:35–41.
- Dexter SC. Microbiologically Influenced Corrosion. In: Cramer DS, Covino BS, editors. *Volume 13A: Corrosion: Fundamentals, Testing, and Protection*. Metals Park, OH: ASM International; 2003. p. 398–416.
- Doyle DP, Wright TE. Rapid Method for Determining Atmospheric Corrosivity and Corrosion Resistance. In: Ailor WH, editor *Atmospheric Corrosion*. New York: John Wiley and Sons; 1982. p. 227–243.
- Dravnieks A, Cataldi HA. Industrial applications of a method for measuring small amounts of corrosion without removal of corrosion products. *Corrosion* 1954;10:224–230.

- Eden DC, Cayard MS, Kintz JD, Schrecengost RA, Breen BP, Kramer E. *Making Credible Corrosion Measurements—Real Corrosion, Real Time*. CORROSION 2003, Paper No. 376. Houston, TX: NACE International; 2003.
- Eden DA, Hladky K, John DG, Dawson JL. *Simultaneous Monitoring of Potential and Current Noise Signals from Corroding Electrodes*. CORROSION 1986, Paper No. 274. Houston, TX: NACE International; 1986.
- Eden DA. Electrochemical Noise. In: Revie RW, editor *Uhlig Corrosion Handbook*. New York: John Wiley & Sons; 2000.
- Eden DA. *Electrochemical Noise—The Third Octave*. CORROSION 2005, Paper No. 351. Houston, TX: NACE International; 2005.
- Edgemon GL. *Electrochemical Noise Based Corrosion Monitoring: Hanford Site Program Status*. CORROSION 2005, Paper No. 584. Houston, TX: NACE International; 2005.
- Feliu S, Morcillo M, Chico B. Effect of distance from sea on atmospheric corrosion rate. *Corrosion* 1999;55:883–891.
- Frankenthal RP. Electronic Materials, Components, and Devices. In: Revie RW, editor *Uhlig's Corrosion Handbook*. New York: Wiley-Interscience; 2000. p. 941–947.
- Freedman AJ, Troscinski ES, Dravnieks A. An electrical resistance method of corrosion monitoring in refinery equipment. *Corrosion* 1958;14:175t–178t.
- Freeman RA, Silverman DC. Error propagation in coupon immersion tests. *Corrosion* 1992;48:463–466.
- Geesey GG. Introduction Part II—Biofilm Formation. In: Kobrin G, editor *Microbiologically Influenced Corrosion*. Houston, TX: NACE International; 1993.
- Gerhardt P, et al. *Manual of Methods for General Bacteriology*. Washington, DC: American Society of Microbiology; 1981.
- Gilbert PD, Herbert BN. Monitoring Microbial Fouling in Flowing Systems Using Coupons. In: Hopton JW, Hill EC, editors. *Industrial Microbiological Testing*. London, UK: Blackwell Scientific Publications; 1987. p. 79–98.
- Grauer R, Moreland PJ, Pini G. *A Literature Review of Polarisation Resistance Constant (B) Values for the Measurement of Corrosion Rate*. Houston, TX: NACE International; 1982.
- Haruyama S, Tsuru TA. *Corrosion Monitor Based on Impedance Method Electrochemical Corrosion Testing*. Mansfeld, F. and Bertocci, U. [STP 727], Philadelphia, PA: American Society for Testing and Materials; 1981. 167–186. Computer Modeling in Corrosion.
- Hausler RH. Corrosion Inhibitors. In: Baboian R, editor *Corrosion Tests and Standards*. 2nd Edition. West Conshohocken, PA: American Society for Testing of Materials; 2005. p. 480–499.
- Hidy GM. *Aerosols: An Industrial and Environmental Science*. Orlando, FL: Academic Press; 1984.
- Huet F, Bautista A, Bertocci U. Listening to corrosion. *The Electrochemical Society Interface* 2001;10:40–43.
- Hunik JH, van den Hoogen MP, de Boer W, Smit M, Tramper J. Quantitative Determination of the Spatial Distribution of *Nitrosomonas europaea* and *Nitrobacter agilis* Cells Immobilized in κ -Carrageenan Gel Beads by a Specific Fluorescent-Antibody Labelling Technique. *Applied and Environmental Microbiology* 1993;59:1951–1954.
- Iverson WP. Transient voltage changes produced in corroding metals and alloys. *Journal of the Electrochemical Society* 1968;115:617–618.
- Jack TR. Monitoring microbial fouling and corrosion problems in industrial systems. *Corrosion Reviews* 1999;17:1–31.
- Jack TR. Biological Corrosion Failures. In: Shipley RJ, Becker WT, editors. *ASM Handbook Volume 11: Failure Analysis and Prevention*. Materials Park, OH: ASM International; 2002.

- Jankowski J. Harmonic synthesis: a novel electrochemical method for corrosion rate monitoring. *Journal of the Electrochemical Society* 2003;150:B181–B191.
- Jorgenson BB. A comparison of methods for the quantification of bacterial sulfate reduction in coastal marine sediments. *Geomicrobiology Journal* 1978;1:49–64.
- Kane RD, Cayard MS. Use of corrosion monitoring to minimize downtime and equipment failures. *Chemical Engineering Progress* 1998;94(10):49–57.
- Kelly RG, et al. *Embeddable Microinstruments for Corrosion Monitoring*. Houston, TX: NACE International; 1997. 1–12. Corrosion 97.
- King G, Ganther W, Hughes J, Grigioni P, Pellegrini A. Studies in Antarctica Help to Better Define the Temperature Criterion for Atmospheric Corrosion; 2001. NACE International Northern Area Region Conference, February 26–28, 2001 Anchorage, Alaska.
- Klassen RD, Roberge PR, Hyatt CV. A novel approach to characterizing localized corrosion within a crevice. *Electrochimica Acta* 2001;46:3705–3713.
- Klassen RD, Roberge PR. *Self Linear Polarization Resistance*. CORROSION 2002, Paper No. 330. Houston, TX: NACE International; 2002.
- Klassen RD, Roberge PR, Lenard DR, Blenkinsop GN. *Corrosivity Patterns Near Sources of Salt Aerosols*. Townsend HE, editor [ASTM STP 1421], West Conshohocken, PA: American Society for Testing and Materials; 2002. p. 19–33. Outdoor and Indoor Atmospheric Corrosion.
- Morcillo M, Chico B, Mariaca L, Otero E. Salinity in marine atmospheric corrosion: its dependence on the wind regime existing in the site. *Corrosion Science* 2000;42:91–104.
- Odom JM, Jessie K, Knodel E, Emptage M. Immunological cross-reactivities of adenosine-5'-phosphosulfate reductases from sulfate-reducing and sulfide oxidizing bacteria. *Applied and Environmental Microbiology* 1991;57:727–733.
- Pope DH. State of the Art Report on Monitoring, *Prevention and Mitigation of Microbiologically Influenced Corrosion in the Natural Gas Industry*. GRI-92/0382. Chicago, IL: Gas Research Institute; 1992.
- Powell DE, Ma'ruf DI, Rahman IY. Practical considerations in establishing corrosion monitoring for upstream oil and gas gathering systems. *Materials Performance* 2001;4:50–54.
- Roberge PR, Tullmin MAA, Grenier L, Ringas C. Corrosion surveillance for aircraft. *Materials Performance* 1996;35:50–54.
- Roberge PR, Sastri VS. On-line corrosion monitoring with electrochemical impedance spectroscopy. *Corrosion* 1994;50:744–754.
- Roberge PR. *Analyzing Electrochemical Impedance Corrosion Measurements by the Systematic Permutation of Data Points*. Munn, R. S. [STP 1154] Philadelphia, PA: American Society for Testing and Materials; 197–211. 1992. Computer Modeling in Corrosion.
- Roberge PR. *Handbook of Corrosion Engineering*. New York: McGraw-Hill; 2000.
- Roberge PR. *Corrosion Basics—An Introduction*. 2nd edn. Houston, TX: NACE International; 2006.
- Sanders PF. Monitoring and Control of Sessile Microbes: Cost Effective Ways to Reduce Microbial Corrosion. In: Sequeira CAC, Tiller AK, editors. *Microbial Corrosion-1*. New York: Elsevier Applied Science; 1988. p. 191–223.
- Scanlan RJ, Boothman RM, Clarida DR. *Corrosion Monitoring Experience in the Refinery Industry Using the FSM Technique*. CORROSION 2003, Paper No. 655. Houston, TX: NACE International; 2003.
- Schmitt G, Buschmann R, Foehn K-H, Theunissen H. *Simultaneous Monitoring of Potential and Current Noise Signals from Corroding Electrodes*. CORROSION 2005, Paper No. 382. Houston, TX: NACE International; 2005.

- Scott PJB. Expert consensus on MIC: prevention and monitoring. *Materials Performance* 2004;43:50–54.
- Scott PJB, Davies M. Survey of field kits for sulfate reducing bacteria. *Materials Performance* 1992;31:64–68.
- Sereda PJ, Croll SG, Slade HF. Measurement of the Time-of-Wetness by Moisture Sensors and their Calibration. In: Dean SW, Rhea EC, editors. *Atmospheric Corrosion of Metals*. Philadelphia, PA: ASTM; 1982. p. 48.
- Silverman, *DC Tutorial on Cyclic Potentiodynamic Polarization Technique*. CORROSION 98, Paper No. 299. Houston, TX: NACE International; 1998.
- Standard Guide for Conducting and Evaluating Galvanic Corrosion Tests in Electrolytes. *Annual Book of ASTM Standards*. G71-81[Vol 03.02] Philadelphia, PA: American Society for Testing of Materials; 2003.
- Standard Guide for Conducting Corrosion Tests in Field Applications. G4-01. 2001. West Conshohocken, PA, American Society for Testing of Materials.
- Standard Guide for Crevice Corrosion Testing of Iron-Base and Nickel-Base Stainless Alloys in Seawater and Other Chloride-Containing Aqueous Environments. G78-01[Vol 03.02]. 2001. West Conshohocken, PA, American Society for Testing of Materials.
- Standard Guide for On-Line Monitoring of Corrosion in Plant Equipment (Electrical and Electrochemical Methods). *Annual Book of ASTM Standards*. G 96-90[Vol 03. 02] Philadelphia, PA: American Society for Testing of Materials; 2001.
- Standard Practice for Conducting Wire-on-Bolt Test for Atmospheric Galvanic Corrosion*. ASTM G116-99 edn. West Conshohocken, PA: American Society for Testing of Materials, 1999.
- Standard Practice for Laboratory Immersion Corrosion Testing of Metals. ASTM G31-72. 2004. West Conshohocken, PA, American Society for Testing of Materials.
- Standard Practice for Making and Using C-Ring Stress-Corrosion Test Specimens. G38-01[Vol 03. 02]. 2001. West Conshohocken, PA, American Society for Testing of Materials.
- Standard Practice for Making and Using U-Bend Stress-Corrosion Test Specimens. G30-97[Vol 03.02]. 2003. West Conshohocken, PA, American Society for Testing of Materials.
- Standard Practice for Measurement of Time-of-Wetness on Surfaces Exposed to Wetting Conditions as in Atmospheric Corrosion Testing. ASTM G84-89. [Annual Book of ASTM Standards, Vol 03.02] 1999. West Conshohocken, PA, American Society for Testing of Materials.
- Standard Practice for Preparation of Stress-Corrosion Test Specimens for Weldments. G58-85[Vol 03.02]. 1999. West Conshohocken, PA, American Society for Testing of Materials.
- Standard Practice for Preparing, Cleaning, and Evaluating Corrosion Test Specimens. G1-03[Vol 03.02]. 2003. West Conshohocken, PA, American Society for Testing of Materials.
- Standard Test Method for Determining Atmospheric Chloride Deposition Rate by Wet Candle Method. ASTM G140-02. [Annual Book of ASTM Standards, Vol 03.02]. 2002. Philadelphia, PA, American Society for Testing of Materials.
- Standard Test Method for Determining the Susceptibility to Intergranular Corrosion of 5XXX Series Aluminum Alloys by Mass Loss After Exposure to Nitric Acid (NAML Test). *Annual Book of ASTM Standards*. G67-04[Vol 03.02]. Philadelphia, PA: American Society for Testing of Materials; 2004.
- Standard Test Methods for Dissolved Oxygen in Water. *Annual Book of ASTM Standards*. D 888-03. West Conshohocken, PA: American Society for Testing and Materials; 2003.
- Standard Test Methods for Electrical Conductivity and Resistivity of Water. *Annual Book of ASTM Standards*. D West Conshohocken, PA, American Society for Testing and Materials; 2005. 1125–95.

- Standard Test Methods for Pitting and Crevice Corrosion Resistance of Stainless Steels and Related Alloys by Use of Ferric Chloride Solution. *Annual Book of ASTM Standards*. G 48-03[Vol 03.02]. Philadelphia, PA: American Society for Testing of Materials; 2003.
- Stern M, Geary AL. *Journal of the Electrochemical Society* 1957;104:56.
- Stern M. *Corrosion* 1958;14:440.
- Techniques for Monitoring Corrosion and Related Parameters in Field Applications. NACE 3T199. Houston, TX, NACE International; 1999.
- Thomas MJJS, Terpsta S. *Corrosion Monitoring in Oil and Gas Production*. CORROSION 2003, Paper No. 431. Houston, TX: NACE International; 2003.
- Van Orden, AC. *Applications and Problem Solving Using the Polarization Technique*. CORROSION 98, Paper No. 301. Houston, TX: NACE International; 1998.
- Vera JR, Méndez C, Hernández S, Cerpa S. *Field Results of the Hydrogen Permeation Sensor Based on Fuel Cell Technology*. CORROSION 2002, Paper No. 346. Houston, TX: NACE International; 2002.
- Voordouw G, Telang AJ, Jack TR, Foght J, Fedorak PM, Westlake DWS. Identification of sulfate-reducing bacteria by hydrogenase gene probes and reverse sample genome probing. In: Minear RA, Ford AM, Needham LL, Karch MJ, editors. *Applications of Molecular Biology in Environmental Chemistry*. Boca Raton: Lewis Publishers; 1995.
- Yang L, Sridhar N, Pensado O, Dunn DS. An in situ galvanically coupled multielectrode array sensor for localized corrosion. *Corrosion* 2002;58:1004–1014.
- Yang B. Method for on-line determination of underdeposit corrosion rates in cooling water systems. *Corrosion* 1995;51:153–165.
- Yang B. *Real time localized corrosion monitoring in refinery cooling water systems*. CORROSION 1998, Paper No. 595. Houston, TX: NACE International; 1998.
- Yang L, Sridhar N. Coupled multielectrode online corrosion sensor. *Materials Performance* 2003;42:48–52.
- Zintel TP, Licina GJ, Jack TR. Techniques for MIC monitoring. In: Stoecker II JG, ed. *A Practical Manual on Microbiologically Influenced Corrosion*. Houston, TX: NACE international; 2001.
- Zollars B, Salazar N, Gilbert J, Sanders M. *Remote Datalogger for Thin Film Sensors*. Houston, TX: NACE International; 1997.

34

SURFACE PROPERTIES MEASUREMENT

MRINALINI MULUKUTLA AND SANDIP P. HARIMKAR

- 34.1 Introduction
- 34.2 Surface properties
 - 34.2.1 Surface finish or surface texture
- 34.3 Microstructural analysis
 - 34.3.1 Optical microscopy
 - 34.3.2 Scanning electron microscopy
 - 34.3.3 Transmission electron microscopy
- 34.4 Compositional analysis
 - 34.4.1 Energy dispersive x-ray analysis
 - 34.4.2 X-ray photoelectron spectroscopy
 - 34.4.3 Auger electron spectroscopy
- 34.5 Phase analysis
 - 34.5.1 X-ray diffraction
- 34.6 Mechanical testing
 - 34.6.1 Hardness testing
 - 34.6.2 Fracture toughness
 - 34.6.3 Tribological properties
- 34.7 Corrosion properties
- 34.8 Standards for surface engineering measurement
- References

34.1 INTRODUCTION

A surface of a material is its interface between the bulk and the external phase, which can be solid, liquid, or gas, in contact with the bulk material. Surface engineering of materials is important for improving the properties of the surfaces for applications in diverse fields,

including mechanical, chemical, petrochemical, biomedical, electronic, power, automotive, construction, aerospace, missile, and many others covering almost all the fields of human activity (Stanley, 2005). As the demand for these engineered materials with improved surface properties is increasing, controlling and characterizing the surfaces is becoming increasingly important to ensure the designed performance of the components (Rok, 2009; Callister, 2006; Polak and Pande, 1999; Totten and Liang, 2004). There are several reasons for evaluation of the materials available, and some of them are listed below:

- to ensure that the material conforms to the specifications of the application it is used in;
- to investigate the development of material properties with processing and fabrication methods used;
- to monitor the performance of the material in different kinds of working environments in the intended application; thus, predicting the extent of damage (such as wear and corrosion).

Various standards are available for the measurement of surface properties of materials. The availability of these standards of measurement removes the ambiguities on the reliability of the obtained data. These standards specify measurement details, such as design specifications, apparatus used and calibration procedures for the same, test procedures, and range of data obtained and its reliability. Few standards organizations with interest in surface engineering include ASTM (American Society for Testing and Materials), SAE (Society of Automotive Engineers), ASME International (American Society of Mechanical Engineers), and ISO (International Organization for Standardization). In this chapter, an overview of the surface properties of materials and also the standard methods available for their measurement and characterization is presented. Various surface properties including, mechanical, physicochemical, topographic, and structural properties are discussed in detail. The combination of these characteristic properties can give an idea about the surface state of any particular material.

34.2 SURFACE PROPERTIES

34.2.1 Surface Finish or Surface Texture

The surfaces of most of the materials are not very smooth and have regular and irregular features, which tend to form a pattern or texture on the surface. These surface features are often formed during initial processing or in-service degradation. Surface finish is related with the geometric irregularities of the surfaces of the solid materials, and is defined in terms of roughness, waviness, lay, and flaws (Reidenbach et al., 1994). These basic components of surface texture are shown in Figure 34.1. Surface roughness is a measure of finer irregularities in the surface finish. Numerically, it is the average deviation of the surface valleys and peaks, expressed in micrometers (Blau, 1992). Surface waviness refers to the repeating irregularities on which the surface roughness is superimposed. These features tend to affect the performance of the engineered component in different ways, and hence should be measured and controlled for improved performance of the component. For example, the surface features significantly influence friction, lubrication, and wear characteristics of materials.

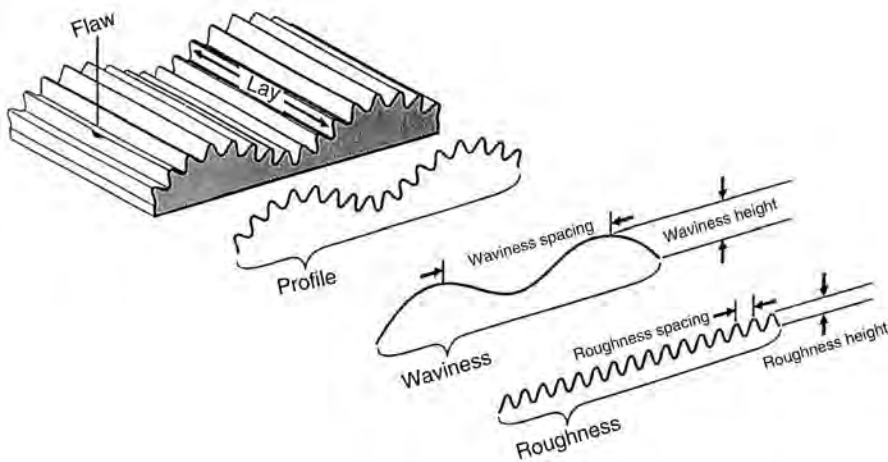


FIGURE 34.1 Basic components of surface texture.

In order to analyze the surface topography of a specimen effectively, the profile generated by the measuring instrument has to be evaluated according to the internationally accepted surface parameters. Each of these parameters characterizes a specific aspect of the surface. These parameters are categorized as follows:

- *Amplitude Parameters:* They are a measure of the vertical characteristics of the deviations on the surface.
- *Spacing Parameters:* They are a measure of the irregularities along the surface irrespective of their amplitude.
- *Hybrid Parameters:* They give us details of the combination of amplitude and spacing of the irregularities.

These surface roughness parameters along with their description and the mathematical formulae used for their calculations are presented in Table 34.1 (Reidenbach et al., 1994; Blau, 1992; Creath and Wyant, 1990; Bhushan, 2001; Takadoun, 2008).

The surface topography is commonly measured using contact-type or stylus-type surface profilometers (Blau, 1992; Creath and Wyant, 1990). A stylus profilometer consists of a diamond tip that acts as a mechanical stylus having radius of curvature ranging between 1 and 2 μm . The surface under observation is moved under the stylus orthogonally in micron steps. The vertical movements of the stylus along the surface allows for the topographic defects to be studied using a sensor that generates electrical signals, which are digitized and processed by the computer, giving the topographic measurement of the surface. The primary surface profile is then subjected to various filtering techniques that separate the roughness, waviness lay, and form features of the surfaces. These profilometers can also cover an area by multiple scans. These profilometers are available in various forms ranging from bench-mounted instruments to portable devices of smaller sizes, which are very convenient for in situ measurements of large components.

Also, there are noncontacting three-dimensional profilometers using optical systems for detection of surface heights. The measurement of the state of a surface by optical

TABLE 34.1 Surface Roughness Parameters and the Corresponding Formulae Used for Calculation (Reidenbach et al., 1994; Blau, 1992; Creath and Wyant, 1990; Bhushan, 2001; Takadoun, 2008)

Parameter	Description	Formula
R_a	Arithmetic average of absolute values	$R_a = \frac{1}{n} \sum_{i=1}^n y_i $
R_q, R_{RMS}	Root mean squared	$R_q = \sqrt{\frac{1}{n} \sum_{i=1}^n y_i^2 }$
R_v	Maximum valley depth	$R_v = \min_i y_i$
R_p	Maximum peak height	$R_p = \max_i y_i$
R_t	Maximum height of the profile	$R_t = R_p - R_v$
R_{sk}	Skewness	$R_{sk} = \frac{1}{nR_q^3} \sum_{i=1}^n y_i^3 $
R_{ku}	Kurtosis	$R_{ku} = \frac{1}{nR_q^4} \sum_{i=1}^n y_i^4 $
R_{zDIN}, R_{tm}	Average distance between the highest peak and lowest valley in each sampling length, ASME Y14.36M—1996 Surface Texture Symbols	$R_{zDIN} = \frac{1}{s} \sum_{i=1}^s R_{ti},$ where s is the number of sampling lengths, and R_{ti} is R_t for the i th sampling length
R_{zJIS}	Japanese Industrial Standard for R_z , based on the five highest peaks and lowest valleys over the entire sampling length	$R_{zJIS} = \frac{1}{5} \sum_{i=1}^5 R_{pi} - R_{vi},$ where R_{pi} and R_{vi} are the i th highest peak and lowest valley, respectively

method utilizes a laser or white beam to scan the surface. It is generally conducted using a confocal microscope or an interferometric microscope. These techniques provide very high resolution and rapid data acquisition making them particularly useful in microtopography. These noncontact techniques have become more popular in surface topography, in spite of being expensive, due to the ease of analysis and accuracy in the results obtained along with their ability to scan the surfaces without damaging them.

Other advanced equipments that give significantly improved performance particularly at finer length scales are very high resolution local probe microscopes, which include scanning tunneling microscope (STM) and atomic force microscope (AFM) (Creath and Wyant, 1990). These instruments use very short-range interactions between a fine probe and the sample surface, and hence yield very high resolution as compared to the previously described techniques. The resolution of the final image obtained depends on the size of the probe and the variation of probe-surface interaction with the distance to the surface. While STM is not suitable for nonconducting materials, AFM can be used for any kinds of surfaces to achieve atomic scale resolutions. The resolution of each type of instrument and their applications are tabulated in Table 34.2 (Creath and Wyant, 1990; Bhushan, 2001; Takadoun, 2008).

TABLE 34.2 Resolution and Applicability of Various Surface Roughness Measurement Instruments (Creath and Wyant, 1990; Bhushan, 2001; Takadom, 2008)

Technique	Information Obtained	Vertical Resolution, nm	Lateral Resolution	Types of Specimen
Stylus profilometry	Profilometry, topography, morphology, scar depth, wear volume	0.5	100 nm	Any flat specimen
Scanning tunneling microscopy (STM)	Topographical imaging, profilometry, compositional mapping, morphology, film thickness, spectroscopy, structure, defects	0.03–0.05	Atomic	Conductors
Atomic force microscopy (AFM)	Topographical imaging, friction force mapping, profilometry, morphology, film thickness, scar depth, wear volume, spectroscopy, structure, defects	0.03–0.05	Atomic to 1 nm	Any kind of specimen

34.3 MICROSTRUCTURAL ANALYSIS

Microstructure refers to the microscopic description of individual constituents present in a material. It involves description of composition, phases, crystal structures, crystal orientations, grain sizes, phase/grain size distributions, defects (dislocations, pores, cracks, inclusions), and so on. The microstructure influences the macroscopic behavior in terms of its physical properties, mechanical properties, tribological characteristics, corrosion properties, and so on. One can predict the performance and properties of a material for a given application from the microstructure of the material. Major determinants of the properties of a material come from the corresponding constituent chemical elements present and the processing method. It is very important to determine the structural elements and processing/microstructural defects, which influence the properties of a material.

For performing microscopic analysis, samples are generally prepared using standard metallographic techniques. The technique involves a series of steps such as grinding, polishing, and etching to reveal the microstructure of the sample surface. The sample can be analyzed using optical or electron microscopy after metallographic preparation. It is usually done using a light microscope to observe the morphology of the phases resulting from phase transformations such as solidification and heat treatment. Structural features investigated through the light microscope include surface morphology, size of the precipitates, compositional inhomogeneities, microporosity, corrosion, thickness, structure of surface coatings, and microstructure of defects. Electron microscopy is conducted using scanning electron microscope (SEM) and the transmission electron microscope (TEM) (Bozzola and Russell, 1999). SEM is carried out on the specimen surfaces and the TEM on the electron transparent thin foils made out of bulk specimens. The electron microscope offers a better depth of field and higher resolution than a light microscope. We generally use scanning electron microscope and transmission electron microscope, where the SEM images the surface of the material, while the TEM reveals internal structure (Bozzola and Russell, 1999; Zinkle et al., 2009; Voort, 2004). The only disadvantage of the electron

microscopes is that they operate under vacuum and they cost significantly more than the light microscopes, but the quality of the results obtained using them are unparalleled. The details about each of the microscopic techniques are included in the following section.

34.3.1 Optical Microscopy

Optical microscopy refers to the microscopic inspection of the sample at lower magnification using an instrument known as an optical or light microscope. The specimen is polished carefully using the traditional polishing techniques and is placed perpendicular to the axis of the objective lens for inspection by optical microscopy. Light rays are focused on the sample that reflects some of the rays back to the lens. The image of the sample positioned can thus be observed. The image seen in the microscope not only depends on how the specimen is illuminated and positioned but also depends on the characteristics of the specimen. This image can be captured digitally to generate a micrograph. We can observe the microstructural features such as coating integrity, porosity, voids, oxide layer at different magnifications depending on the capability of the instrument.

34.3.2 Scanning Electron Microscopy

The electron microscopes use a high energy beam of electrons to form an image by illuminating the specimen under high vacuum. To control the electron beam, the electron microscope uses electrostatic and electromagnetic lenses, which aid in focusing the beam to form an image. A schematic view of different components of an SEM and their typical functions is shown in Figure 34.2. SEM is one of the most important characterization

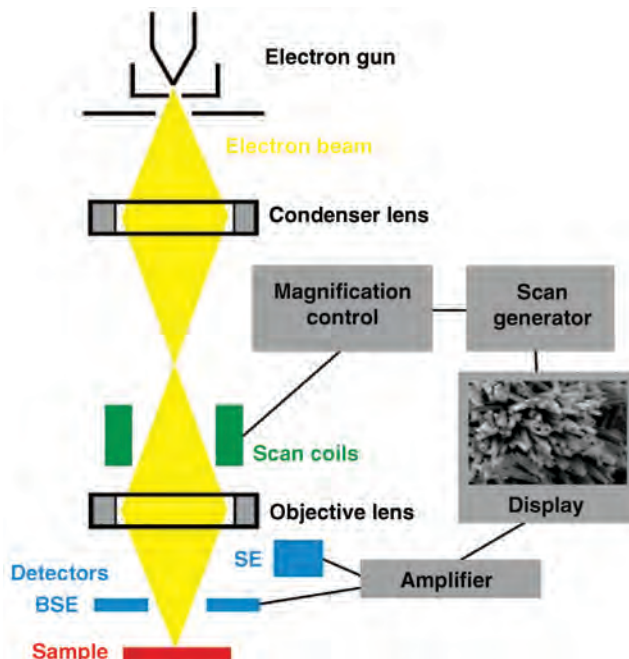


FIGURE 34.2 Schematic representation of an SEM (Bozzola and Russell, 1999).

tools in metallurgy. It is useful in obtaining high resolution images, which are essential for microstructural study of any specimen. Scanning electron microscopy is different from optical microscopy in the way the image is obtained. It creates image by focusing high energy electrons and detecting signals from the interaction of the electrons with the specimen's surface. The signals can be secondary electrons, characteristic x-rays or backscattered electrons. Using an SEM has the advantages of producing images of high resolution and large depth of field. With an SEM, we can produce images of very high magnification up to $50,000\times$ very quickly and easily, which is not possible with a light microscope. SEM has got the added advantages of good compositional contrast and relative ease of affordability not only in terms of the cost but also in terms of reliability and the throughput. Another aspect to be remembered is that the sample has to be conductive in order to dissipate the charge during SEM. Therefore nonconductive surfaces have to be gold or carbon coated to make them conductive so that no charging occurs. Major limitation of SEM is the use of a very high vacuum, which restricts what we can look at. SEM is widely used by biologists, materials, and mineral scientists for microstructural analysis.

34.3.3 Transmission Electron Microscopy

In transmission electron microscopy (TEM), image is formed by a beam of electrons passing through an ultra-thin specimen. TEM allows analysis of the microstructure with atomic scale resolution. The TEM can be divided into three sections: illumination system, specimen stage, and imaging system. The illumination system consists of the electron gun with a set of two or more condenser lens. A schematic representation of different components of TEM and their functions is shown in Figure 34.3. The electron gun generates electrons by thermionic or field emission effects, which are focused on to the sample using the condenser lens. The condenser lens system should consist of at least two electron lenses. The first condenser lens, which is a strong magnetic lens, uses the electron source as object and produces a real image of different sizes depending upon the lens current thereby determining the spot size. The second condenser lens, which is a weak magnetic lens, offers low magnification or no magnification but controls the brightness or the intensity of the beam, directs the beam through a condenser aperture. It restricts the high angle electrons and dismisses any aberration. The condenser lens system also consists of a condenser aperture and condenser stigmator to correct for any astigmatism resulting from the two-condenser lens. The beam then strikes the sample and transmits through it, based on the thickness of the sample. TEM samples are generally made circular with a diameter of 3 mm, and it must be made sure that the specimens are thin enough to transmit the electrons through them to form the magnified image. The imaging system consists of an objective lens, objective aperture, selected area aperture, projector lens, TEM screen, and camera. Objective lens focuses the transmitted beam into an image, whereas the high angle beam is restricted by the objective aperture and selected area aperture. The function of the intermediate lens is to change the image magnification and form diffraction patterns on whole of the TEM viewing whereas the purpose of projector lens is to form an image or diffraction pattern across the TEM screen. The image then hits the phosphor image screen that converts the electron image into a visible form.

Using TEM, one can obtain very fine (as small as single atoms) internal microstructural details of the specimen. Magnifications of the order of $1,000,000\times$ can be obtained very easily using transmission electron microscopy. This technique is widely used to analyze finer details in a range of scientific fields including physical and

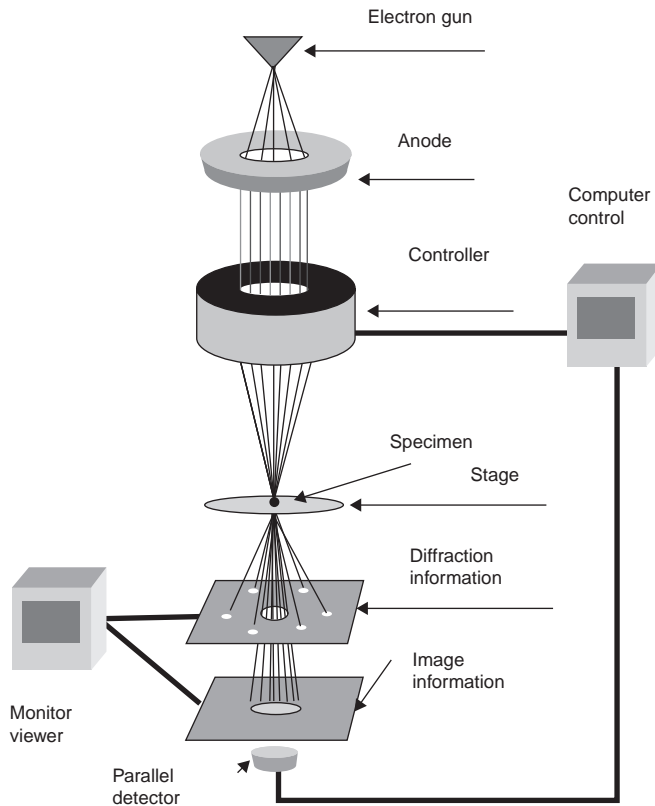


FIGURE 34.3 Schematic representation of a transmission electron microscope (Zinkle et al., 2009).

biological sciences. It finds applications in cancer research, virology, examination of biological specimens, materials science, and semiconductor research. In spite of the advantages, TEM has a number of drawbacks in many concerns. The sample preparation procedure involved is very tedious, extensive, and has to be done very carefully. The sample preparation procedure can also result in a change in the structure. Another limitation is that the field of view of the specimen being very small, it might not represent the properties of the whole surface. There is also a possibility of damage of the sample by the electron beam, especially in the case of biological materials. Regarding the economy point of view, TEM is expensive when compared to other microscopes.

34.4 COMPOSITIONAL ANALYSIS

The chemical composition of surfaces plays a significant role in determining the material properties. Using advanced tools, we can determine the chemical, elemental, and molecular composition of material giving us an idea of the composition of few outer layers of the surfaces under observation. Such an analysis is very important as the properties of surfaces on the outer layers control their electrical, chemical, or mechanical properties. The surface compositional analysis can be done by bombarding a material with ions,

electrons, x-rays, or photons (Voort, 2004). For this analysis, one can use numerous techniques such as x-ray photoelectron spectroscopy, auger electron spectroscopy, and many others for a wide range of materials including photovoltaics, microelectronics, polymers, biological specimens, and many others (Creath and Wyant, 1990).

34.4.1 Energy Dispersive x-ray Analysis

Energy dispersive x-ray spectroscopy technique is used in conjunction with scanning electron microscope to analyze the elemental composition of a specimen. This technique makes use of the x-rays, which are emitted from the sample in response to the bombardment of the electron beam, to characterize the chemical composition of the region of specimen under observation. The microscopic features or phases as small as 1 μm present in the sample can be analyzed using this technique. In this technique, a high energy electron beam is focused onto the region of the specimen to be studied. This beam of charged particles excites the atoms on the surface resulting in release of energy in the form of x-rays. X-ray detector of the EDS measures the number and energy of the x-rays emitted from the specimen. The energy of the x-ray is a characteristic of the element from which it is emitted and hence, helps us in identifying the element associated. The number of x-rays of each characteristic radiation and their relative counts are recorded in the form of a spectrum. This spectrum helps us in determination of the relative amounts of each element present both qualitatively and quantitatively within the analyzed region. This technique facilitates qualitative elemental analysis with sensitivity down to a few atomic percent. Major limitation of this method is that low atomic number elements are detected only by some systems equipped with special detectors.

34.4.2 X-ray Photoelectron Spectroscopy

X-ray photoelectron spectroscopy, also called electron spectroscopy for chemical analysis, is a qualitative spectroscopic technique that involves bombardment of the specimen surface with x-rays and the consequent measurement of photoemitted electrons. These photoemitted electrons have different kinetic energies characteristic of the emitting atoms and their bonding states. XPS measures elemental composition, empirical formula, chemical state, and electronic state of the elements existing in the material. It requires ultra high vacuum conditions for the analysis. It can be used to analyze the surface of few millimeter thicknesses to a depth of 1–10 nm. This technique can detect elements with atomic number greater than 2 (lithium and above). It is generally used to analyze inorganic compounds, metallic alloys, semiconductors, polymers, ceramics, elements, catalysts, paints, glasses, plant parts, medical implants, and many others. Hence, analysis of corrosion layers and oxide films can be done to determine the nature of chemical bonds of the main elements and impurities.

34.4.3 Auger Electron Spectroscopy

Auger electron spectroscopy is one of the widely used techniques for obtaining the chemical composition of solid surfaces. This is a surface specific technique and utilizes the emission of low energy electrons in the auger process to determine the composition of surface layers of a sample. Three steps involved in the characterization are atomic ionization, electron emission, and analysis of the emitted auger electrons. The surface is

bombarded with a focused beam of electrons generating auger electrons that are collected and measured. These auger electrons have discrete kinetic energies that are characteristic of the emitting atoms making this technique widely useful for finding elemental composition. The main advantages of this method include high sensitivity for chemical analysis (in the range of 5–20 Å), rapid data acquisition speed, ability to detect all the elements above helium, and its capability of high spatial resolution. Hence, it can be used to analyze adsorbed layers, thermal or anodic oxide films, thin deposits and segregated layers on surfaces within grain boundaries.

34.5 PHASE ANALYSIS

34.5.1 X-ray Diffraction

X-ray diffraction (XRD) is one of the widely used analytical techniques for identification of phases in a crystalline material and also amorphous phases. It can also be used to determine the size of the unit cell and fractional coordinates of atoms within the unit cell along with other structural information such as degree of crystallinity, preferred orientation, mechanical stresses, lattice strains, and stacking faults. It is possible to characterize the phases qualitatively as well as quantitatively using this technique (Cullity, 1978; Krawitz, 2001; Crankovic, 1986).

X-ray diffraction is a most common technique for the study of crystal structure and atomic spacing associated with the material. It is based on principle of constructive interference of monochromatic x-rays on a crystalline sample (Cullity, 1978). For analysis using this technique, x-rays are produced in a cathode ray tube. First, the filament is heated to produce electrons, which are accelerated onto a target by application of voltage and then bombarded. When these electrons have sufficient energy to displace electrons in the inner shells of the target material, characteristic x-rays are produced. Copper is most common target material used for single-crystal diffraction, with CuK_α radiation = 1.5418 Å. The x-rays, thus produced, are filtered to produce monochromatic radiation, collimated and then directed toward the sample. Testing is carried out by recording the intensity of the reflected x-rays on rotating both the sample and the detector. These incident x-rays interact with the sample surface and once they satisfy Bragg's law, constructive interference occurs resulting diffracted rays, and hence, a sharp peak that can be used for identification of the phase present. Bragg's law ($n\lambda = 2d \sin\theta$) gives the relationship between the peak position (2θ), interplanar spacing (d), and wavelength of the electromagnetic radiation (λ). The diffracted rays satisfying Bragg's law are recorded by a detector and are then converted into count rate, which is then output to a device. The diffractometer design is such that when the sample rotates in the path of the x-ray beam at an angle θ , the x-ray detector, which is mounted on an arm to collect diffracted x-rays, rotates an angle of 2θ . The sample and detector are rotated over a range of 2θ angles to obtain diffracted rays in all possible directions. Schematic of a typical diffractometer is shown in Figure 34.4. The diffraction peaks obtained are converted into d-spacings for identification of the mineral/phase. Each mineral has a set of unique d-spacings, and hence, phase identification and analysis of the sample can be done based on the characteristic radiation emitted.

X-ray diffraction is a most widely used nondestructive technique for identification of unknown crystalline materials (Krawitz, 2001). It finds applications for identification

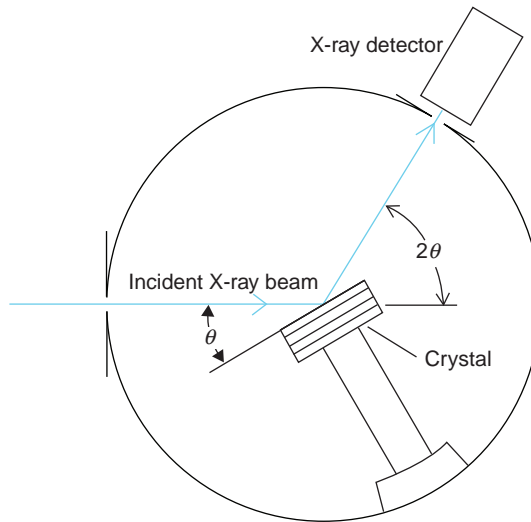


FIGURE 34.4 Schematic diagram of x-ray diffractometer (Cullity, 1978).

of unknown phases in various research areas including geology, engineering, biology, environmental science, and biology. Other applications of this technique include identification of fine-grained materials such as clays, characterization of crystalline materials, unit cell dimension determination, and assessing the sample purity. In combination with specialized techniques, this technique can be used for determination of crystal structures by Rietveld refinement, thin film characterization, texture analysis, and determination of modal amounts of minerals. There is no complex sample preparation procedure involved, and the data obtained can be interpreted with relative ease. There are few limitations due to the manner in which the testing is carried out. If any particular phase is present in minor percentages (below 5%), it may not be detected during the analysis, and it is difficult to differentiate mixture of phases with low symmetry due to large number of diffraction peaks. Although it has few limitations, XRD is a very easy and straightforward method to identify unknown minerals.

34.6 MECHANICAL TESTING

34.6.1 Hardness Testing

The ASM Metals Handbook defines hardness as “*Resistance of metal to plastic deformation*,” usually by indentation (Crankovic, 1986). However, it may also refer to stiffness or resistance of any material to scratch, abrasion, or cutting. For a metal, it may be defined as the ability to resist permanent deformation on application of load. Hence, a metal having higher hardness has greater resistance to deformation.

Hardness measurement can be done on different scales including *macro*-, *micro*-, or *nano-scale* depending on the forces on indenters and depth of indentations. In a metals industry, three types of tests generally carried out for hardness measurement are Brinell hardness test, Rockwell hardness test, and Vickers hardness test (Chandler, 1999). These tests determine the depth to which the ball or cone indenter sinks into the metal at any

given load in a specific period of time, and hence, estimate the materials resistance to deformation. Some of the most common and widely used hardness testing methods in today's industries are explained in the following sections.

34.6.1.1 Rockwell Hardness Test One of the most widely accepted and used hardness test method for bulk materials is the Rockwell hardness test. This is due to its ease, minimum chances of personal error, speed, smaller indentation size, and ability to distinguish small differences in hardness values. It uses the depth of penetration of the indenter at constant load as a measure of hardness.

In this method, a minor load of 10 kg is applied initially in order to seat the specimen followed by the application of major load (60, 100, and 150 kg are used depending on the material). The depth of indentation is recorded automatically in terms of arbitrary hardness numbers. Rockwell hardness testers use various indenters for different scales, and the major loads are applied accordingly. This implies that the Rockwell hardness depends on the load applied and the indenter used, and hence, one must specify the hardness by prefixing the hardness number with a letter indicating the hardness scale employed for testing. Many scales are available in this method, which serves special purposes. An improvement on Rockwell hardness test is the Rockwell superficial hardness test where the preliminary minor load is 3 kgf, and the major loads are 15, 35, and 40 kgf. Different scales are also available in Rockwell superficial hardness test for different purposes.

34.6.1.2 Microhardness Test Microhardness usually refers to the hardness obtained when the load applied does not exceed 1 kgf. The indenter used for microhardness is either Vickers diamond pyramid or Knoop elongated diamond pyramid. The surface to be tested is generally polished using metallographic techniques for precise measurement. The test procedure is similar to the standard Vickers hardness test, and precision microscopes having a magnification of around $500\times$ are used to measure the indentation size. The indentation size can be measured to an accuracy of $\pm 0.5\text{ }\mu\text{m}$ using these microscopes. It should be noted that care should be taken while measuring the size of the indentation to ensure accurate results.

The Knoop hardness number, KHN is mathematically the ratio of the load applied, P (kgf), to the unrecovered projected area, A (mm^2)

$$\text{KHN} = \frac{F}{A} = \frac{P}{CL^2} \quad (34.1)$$

where F is the applied load (kgf), A the unrecovered projected area of the indentation (mm^2), L the measured length of long diagonal of indentation (mm), and $C = 0.07028$ the indenter constant that relates projected area of the indentation to the square of the length of the long diagonal. The Knoop indenter is a rectangular diamond pyramid that produces an indentation having its long and short diagonals in the ratio of 7:1. The depth of a Knoop indentation is about 1/30 of the length of long diagonal.

The Vickers hardness number VPN or HV is given by the applied load (kgf) divided by the surface area of the indentation (mm^2)

$$\text{HV} = \frac{2F \sin(136/2)}{d^2}; \quad \text{HV} = 1.854 \frac{F}{d^2} \quad (34.2)$$

where F is the applied load (kgf), d the arithmetic mean of the two diagonals, d_1 and d_2 in mm, and HV the Vickers hardness. The Vickers diamond pyramid indenter is a square pyramid with an angle of 136° between its faces. The depth of indentation is about 1/7 of the length of diagonal.

34.6.1.3 Nanoindentation Indentation tests are the most commonly applied methods of testing of mechanical properties of advanced materials with microstructural features at much finer (nano/micro) scale. In this context, the term nanoindentation implies the continuous recording of the penetration of the indenter (penetration depth) and the corresponding load, as well as other variables (such as time, frictional force), rather than single-valued measurements of contact area (as in microindentation testing).

Nanoindentation using continuous depth recording (CDR) instruments has both advantages and disadvantages (Fischer, 2004). The advantage of CDR includes a high level of precision, ease of digitization, automation, and data processing. It is ideal for the measurement of creep as well as of plastic and elastic work. Furthermore, each test normally gives a complete loading/unloading cycle, rather than a single reading. As a sole method of measuring indent size, its disadvantage is the need for simplifying assumptions in order to

- separate plastic from elastic effects;
- determine the true zero of the depth measurements;
- allow for the piling-up or sinking-in of material around the indent;
- allow for geometric imperfection of the indenter when deriving absolute hardness values.

In addition to the above applications, the extremely small force and displacement resolutions (as low as 1 nN and 0.2 nm), which can be obtained in this technique combined with large range of applied forces, make this method useful for characterization of all types of material systems. It offers precise control over the data acquisition and sensitivity, which can be used to study the mechanisms of mechanical behavior of materials at submicron length scales. Nanoindentation finds numerous applications in study of dislocation behavior in metals, thin film characterization, fracture behavior of ceramics, residual stress analysis, and time-dependent behavior of polymers and soft metals. It is also an important tool for characterizing tribological behavior of materials including scratch resistance and wear resistance of coatings and bulk materials at nano-scale.

Load Versus Displacement Curve A typical load versus displacement curve obtained from recorded nanoindentation data is shown in Figure 34.5. The loading curve presents a typical parabolic behavior, which is associated with elastic-plastic deformations during the loading. The maximum tip penetration h_{\max} , the contact depth h_c , the residual penetration after unloading h_p , and the measured stiffness S are indicated in Figure 34.5. The elastic recovery after the unloading corresponds to the difference ($h_{\max} - h_p$). The form of these curves, including small changes in the increasing depth rate or in the elastic recovery under unloading, can give information about the surface response to the applied load, and consequently about the mechanical properties at the surface. These curves can be used to extract mechanical properties of the material such as hardness, modulus of elasticity, strain rate sensitivity, and activation volume (Fischer, 2004; Bhushan, 2007; Bhushan, 2000; Li and Bhushan, 2002). Hardness and young's modulus are obtained from this curve using Oliver and Pharr method and the set of formulae for analysis of rigid materials are listed below (for Berkovich indenter) (Fischer, 2004).

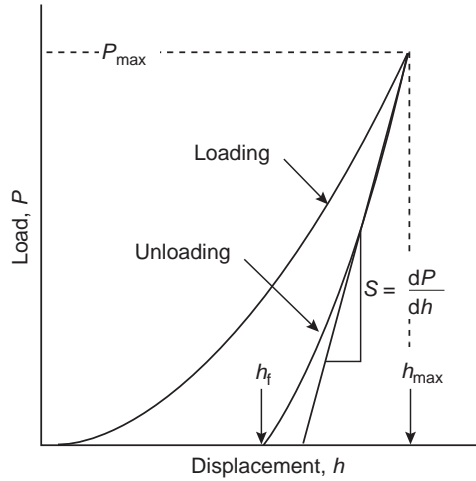


FIGURE 34.5 Schematic illustration of indentation load–displacement data showing important measured parameters (Fischer, 2004).

Reduced modulus

$$\frac{1}{E^*} = \frac{1 - \nu^2}{E} + \frac{1 - \nu'^2}{E'} \quad (34.3)$$

where E^* is the reduced elastic modulus, E is the modulus of the indented specimen, E' is the modulus of the indenter, ν is the Poisson's ratio of the specimen, and ν' is the Poisson's ratio of the indenter.

Contact area (for Berkovich indenter)

$$A = 3\sqrt{3}h_p^2 \tan^2 65.3 = 24.5 h_p^2 \quad (34.4)$$

where A is the contact surface area and h_p is the contact depth.

Hardness

$$H = \frac{P}{24.5 h_p^2} \quad (34.5)$$

where H is the hardness of the specimen and P is the load.

Elastic modulus

$$E^* = \frac{dP}{dh} \frac{1}{2h_p} \frac{1}{\beta} \sqrt{\frac{\pi}{24.5}} \quad (34.6)$$

where dP/dh is the slope of load-displacement curve (called stiffness), and β is a geometrical constant ($\beta = 1.034$ for Berkovich indenter).

Depth and Load-Sensing Equipment The equipment (shown in Figure 34.6) consists of a system of a vertical axis supported by springs to a cell. The indenter is at the end of the axis. The system is composed of a force actuator and a sensor of depth that is generally a

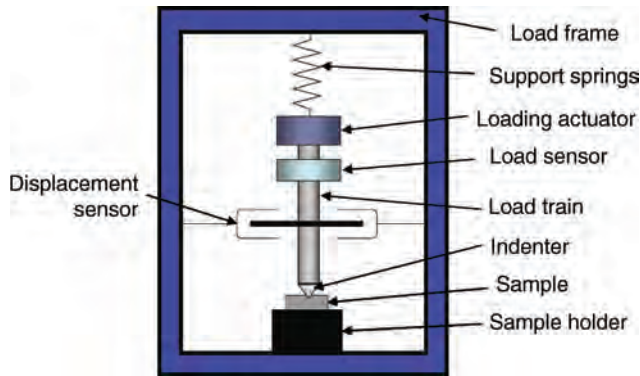


FIGURE 34.6 Illustrative showing different components of a nanoindenter (Bhushan, 2000).

capacitance displacement gauge. The force actuator is capable of applying forces as low as $1\ \mu\text{N}$, and the displacement gauge sensor can give a depth better than $0.1\ \text{nm}$. The maximum load used in this kind of equipment is about $500\ \text{mN}$. On the other hand, if better displacement and load resolutions are necessary, the maximum load is generally lower and depth resolution is increased for shallow penetration. The schematic of sample surface during and after the indentation, showing the parameters definition used in the equation, is as shown in Figure 34.7.

Errors and Limitations The correct measurement of mechanical properties from load and depth-sensing indentation depends on the minimization of errors and comprehension of the methods limitations. However, even in case of a well-calibrated machine, good results are not obtained if the sample presents a high surface roughness. Another problem appears when the material presents pileup at indentation. The pileup at the sides of indentation increases the contact area when the Oliver and Pharr method is used. Then the calculated values of hardness and elastic modulus are lower than the actual values. The errors and limitations in the load and depth-sensing equipment can be divided into three major groups: calibration of equipment and determination of contact surface, calibration of area function for the indenter, and sample effects such as roughness and pileup. Proper understanding of the reasons as to why these errors occur, and the means to minimize or account for these errors is very important to obtain accurate results by nanoindentation.

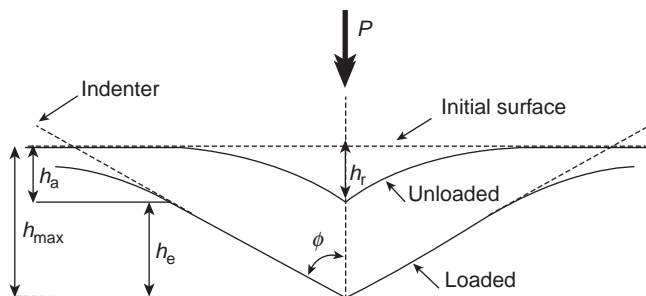


FIGURE 34.7 Schematic illustration of the unloading process during nanoindentation (Li and Bhushan, 2002).

Applications Load and depth-sensing indentation are applied to practically every kind of solid material. Some of the major applications of the nanoindentation technique include (Kuhn and Medlin, 2000)

- (i) determination of the hardness and elastic modulus of deposited thin films and coatings, principally metallic thin films and diamond-like carbon films;
- (ii) determination of mechanical properties of irradiated materials including ionic irradiated surfaces and plasma-based ion implantation; and
- (iii) study of mechanical properties of brittle materials, such as hardness, elastic modulus, toughness, and adhesion properties.

There are certain new areas where the depth and load-sensing indentation is being applied such as determination of plasticity at low dimensions, mechanical properties of nanostructured materials, stress-strain simulation and residual stress determination, viscoelastic properties of polymers, and the creep and strain rate effect on the deformation of materials at low temperature. In these cases, nonberkovich indenters may also be used to characterize mechanical properties other than hardness and elastic modulus.

34.6.2 Fracture Toughness

Fracture toughness is defined as a property that describes the ability of a material containing crack to resist failure. It is one of the most important factors for design considerations of any material (Kuhn and Medlin, 2000). Fracture toughness is denoted as K_c and has the units of $\text{MPa}\sqrt{\text{m}}$ (SI units). It is given by

$$K_c = Y \times \sigma_c \times \sqrt{\pi a} \quad (34.7)$$

where Y is a dimensionless parameter which depends on both crack and specimen sizes and geometries, σ_c is the critical stress for crack propagation (in MPa), and a is the half-crack length (in m). Thus, fracture toughness gives us an idea about the amount of stress required to propagate a preexisting flaw, and in turn, its damage tolerance during service. This is a very important property as there is always a possibility of occurrence of flaws during processing, fabrication or service of a component in the form of cracks, voids, inclusions, and surface defects. These flaws will lead to local concentration of stresses at the tip following which crack growth occurs. The crack propagation can be either unstable or stable, which decides the type of fracture in the specimen. The unstable crack propagation is generally associated with brittle fracture occurring at a well-defined point and can be characterized by a single value of fracture parameter. The stable crack propagation occurs in case of ductile fracture where fracture is an ongoing process and cannot be described by a single point. In case of brittle fractures, this growth will be rapid if the crack tip stresses exceed a critical value leading to failure. The fracture toughness depends on a number of factors such as temperature, environment, loading rate, material composition, microstructure of the material, specimen sizes, and geometries. Although fracture toughness can be obtained from literature, it is preferred to determine the same experimentally for any particular material or joint being designed.

Various measures of toughness are available. Impact toughness, which is widely used method, correlates the energy absorbed during fracture process of with toughness. The

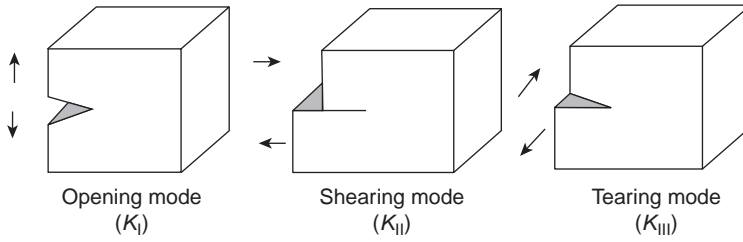


FIGURE 34.8 Figure showing three modes of fracture (Kuhn and Medlin, 2000).

major drawback with this approach is that there is a large degree of uncertainty associated with this kind of empirical correlations. So, it is preferable to determine this property in terms of stress intensity factor (K_C) with the help of different techniques and approaches available. Fracture mechanics approach is widely used for its characterization and is classified into linear-elastic and nonlinear-fracture toughness testing methods (Kuhn and Medlin, 2000; Pabst et al., 1982).

Linear-elastic fracture mechanics approach is commonly used for evaluating the ability of component containing flaw to resist fracture (for design considerations). This approach uses the flaw size, features, geometry of the component and material properties for assessing the same. In general, for brittle materials, the fracture can occur in one of the three modes shown in Figure 34.8, and a roman numeral subscript is used to indicate the mode of fracture. Depending on the mode of fracture, the stress intensity factor is calculated using the corresponding equations taking into consideration the loading, crack size and structural geometry. The thickness of the specimen plays an important role in determination of K_C as the stress states adjacent to flaw changes with the thickness until it exceeds the critical dimension. For thin specimens, the value of K_C is dependent on its thickness, whereas for thicker ones (where specimen thickness is greater than crack dimensions), K_C becomes independent of thickness. This value of K_C is a true material property called plane-strain fracture toughness.

34.6.2.1 Plane-Strain Fracture Toughness Common configurations for the determination of plane-strain fracture toughness are single edge notch bend (SENB or three point bend), and the compact tension (CT) tests. For accurate analysis, thickness of the specimen must exceed the critical thickness (B). It has been proven experimentally that plane-strain conditions prevail when

$$B \geq 2.5 \left(\frac{K_{IC}}{\sigma_y} \right)^2 \quad (34.8)$$

where B is the minimum thickness required to produce a condition where plastic strain energy at the crack tip is minimal, K_{IC} is the fracture toughness of the material, and σ_y is the yield stress of the material. These K_{IC} values can be used to determine the crack length at any given design stress applied to a component using equation:

$$a_c = \frac{1}{\pi(K_{IC}/\sigma_y)^2} \quad (34.9)$$

It can also be used to estimate the critical stress value when a crack of given dimensions is found in a component.

$$\sigma_c \leq \frac{K_{IC}}{Y\sqrt{\pi a}} \quad (34.10)$$

Even though this is a standard procedure adopted for K_{IC} testing of metals, it does not hold good for ceramics as the introduction of precrack in the specimen affects the fracture toughness value considerably leading to erroneous measurement of values. The width of the notch should not exceed $100\ \mu\text{m}$ in order to obtain reliable values of K_{IC} . Also, the notch root has to be sharpened to a radius of less than $10\ \mu\text{m}$ to ensure a reliable toughness measurement. Thus, the sample preparation should be done very carefully to get accurate values of fracture toughness. These limitations led the investigators to find a less labor-intensive technique for quantification of this property, leading to the development of indentation fracture toughness measurement technique.

34.6.2.2 Indentation Fracture Toughness Fracture toughness can also be determined using the Vickers indentation technique in a much simpler and relatively accurate manner. Due to their efficiency, indentation techniques became very popular for toughness measurement of brittle substances (especially ceramics) (Lawn and Marshall, 1979; Anstis et al., 1981;; Strecker et al., 2005). The sample is indented using a Vickers hardness tester at a higher loads leading to the initiation of cracks from the four corners of the impression. Figure 34.9 shows the shape of an impression that is generally obtained on a brittle material. The K_C ($\text{MPa}\sqrt{\text{m}}$) is given by

$$K_C = 0.016 \left(\frac{E}{H} \right)^{1/2} \frac{F}{c^{3/2}} \quad (34.11)$$

where F is the applied load, E is the elastic modulus, H is the Vickers hardness, and c is the length of radial crack emanating from indentation corner. Even though the technique is

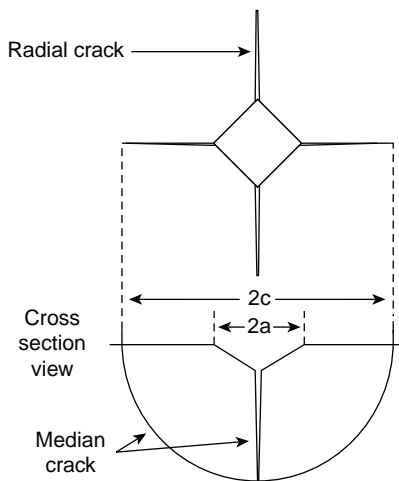


FIGURE 34.9 Fracture toughness measurement by Vickers indentation (Strecker et al., 2005).

easier, considerable care should be taken while measuring the length of the cracks as both equipment error and parallax error can cause considerable uncertainty in the calculated value. These tests are also prone to errors due to inhomogeneities in the material. In order to account for these errors, tests must be carried out repeatedly over a range of loads.

34.6.3 Tribological Properties

Tribology is the science of interacting surfaces in relative motion with each other, which includes the study and application of principles of friction, wear, and lubrication. It is a multidisciplinary field of research aimed at reducing material wear and hence increasing the life time and reliability of mechanical systems and also controlling friction in the same. Wear being one of the main reasons for the failure of the many components, it is very important to have a good understanding of the mechanisms of wear. Wear is defined as a progressive damage or loss of material of a surface caused by relative motion with respect to another surface (Chattopadhyay, 2001). Wear is generally classified into various ways. Based on the appearance of wear scar, damage surface can be considered such as pitted, spalled, scratched, polished, crazed, fretted, and scuffed. The wear can be adhesive, abrasive, or oxidative wear based on physical mechanisms of damage. The wear can be classified based on the conditions surrounding the wear situation such as presence of lubrication, rolling, high temperature, and stress sliding. These classifications are very useful to engineers in predicting different wear situations in different ways. The first way of classification based on appearance helps engineers in recognizing the changes in wear situations at different locations. Hence, one can adopt the corresponding measures of controlling it and thus enhancing the life of the component. The second way of classification based on the physical mechanism is very useful to the engineers in identifying the right mechanism of wear and predicting the wear loss. This approach also helps engineers in identifying the design parameters to be controlled or modified in order to improve the components service life. The third classification is useful from component designer point of view. The description of the wear situations helps the engineers to identify specific set of design rules, equations, and recommendations to be used. While each classification can be correlated with the other, the inter-relationships are not simple, direct, and complete. Due to the complex nature of the wear behavior, it is difficult to establish complete relations between operating conditions, wear mechanisms, and appearance.

The main objective behind performing a wear test is to trace the data that can be applied for a particular application in order to increase its life and reduce cost and maintenance for reliable performance of the particular component. As mentioned before, the nature of wear being complex there is no single, unique parameter to characterize wear behavior of any particular material. This complex nature of wear results in a need to carry out a variety of wear tests each addressing a particular aspect of wear situations. Wear tests are available for assessing the material behavior to different environments and working conditions (Chattopadhyay, 2001; Bayer, 2004; Ludema, 1996; Stachowiak, 2005; Becker and Shipley, 2002; Rabinowicz, 1995). Several standards are available which describe the apparatus, parameters to be varied and sample preparation for each of the tests as detailed in the following section.

34.6.3.1 Dry Sand–Rubber Wheel Abrasion Test This test is developed to predict the wear behavior in situations where low-stress scratching abrasion is the primary mode of wear. In this mode of wear, loose abrasive grains are dragged across a surface under

loading conditions which do not induce fracture of abrasive particles. This test is used to investigate the influence of various parameters such as abrasive particle size, shape, and material parameters on the wear behavior of the component. In this test, a stream of free-falling abrasives between wear specimen and a rotating rubber-coated wheel is trapped. The weight or volume loss in a substance in a very controlled environment is then recorded. This type of testing can be used to assess the scratching abrasion resistance of the material. A standard version of this test was developed by ASTM (ASTM G65), which describes main considerations for the design of apparatus and method of testing in detail.

34.6.3.2 Wet Sand–Rubber Wheel Abrasion Test This test is very much similar to the dry sand–rubber wheel test and also gives us the resistance to low stress scratching abrasion. The basic test concept is same in both the tests, but in the wet test the abrasive is in the form of water slurry. The wear is determined by weight or volume loss similar to that in dry sand–rubber wheel test. Both the tests are used extensively to investigate the scratching abrasion and have been found to correlate well with practical applications. A standard version for materials comparison with dry sand test has been developed by ASTM (ASTM G105). This kind of testing offers an advantage that the procedure can be modified to utilize slurry, which is more representative of the application. Hence, chemical effects associated with the application can be investigated.

34.6.3.3 Erosion by Solid Particle Impingement Using Gas Jets This test method helps the user in investigating the erosion behavior of materials based on the factors affecting it during its service. Gas jets are used to predict the solid particle erosion. It involves the determination of material loss by erosion with jet nozzle type erosion equipment. The conditions and procedures to be followed for this test are outlined in ASTM G76. In actual service, erosion involves particle sizes, velocities, attack angles, environments, and many other factors, which will vary over a wide range. Hence, any single test in one set of conditions might not be sufficient to predict the service life and performance of the component. The investigators should determine carefully the correlation of the results of this test for field performance or consider results from other methods and integrate them to estimate the wear life.

34.6.3.4 Block-on-Ring Wear Test Using Wear Volume It is one of the most commonly used tests for the determination of sliding wear of materials using block-on-ring geometry under different conditions. In this method, a stationary block specimen is pressed with a constant force against a rotating ring specimen at 90° to the axis of rotation of the ring. The resulting friction between the sliding surfaces (i.e., the block and ring) leads to material loss from both the specimens, and this material loss is used for wear analysis. Although both block and ring exhibit wear in this test, it is primarily used to study the wear of block material, in terms of its weight loss and hence wear volume. ASTM G77 describes the procedure to be adopted for this type of test explaining all different conditions. This method can be used for a wide range of materials making it a very flexible test. The test can be carried out using different gaseous atmospheres, lubricants, liquids, or gaseous temperatures in order to get conditions similar to actual service conditions. Rotational speed and load can also be varied to make the service conditions more similar to any particular application. In spite of its flexibility, the method has few disadvantages. As the block experiences more rubbing than the ring, the test is limited in terms of its ability to investigate the relative amount of wear between pairs of materials. Extensive data should

be collected in order to ensure the repeatability of the test. The correlation between the testing conditions and the service conditions varies with change in the test parameters.

34.6.3.5 Pin-on-Disc Wear Test This is also a commonly used test to study and assess the wear behavior of materials. It is a general test that evaluates the sliding wear of pairs of materials. It utilizes a radius-tipped or flat-ended pin to press against a flat rotating disc (specimen to be tested). The relative motion between two bodies results in material loss and thus gives an idea of its sliding wear behavior. The parameters that can be varied in these tests include size and shape of the pin, load, sliding speed, and material pairs. Also, the test can be done in controlled atmosphere and with lubrication. ASTM G99 outlines the standard procedure for this test.

34.6.3.6 Rolling Wear Test This test is used to address the rolling wear. It consists of a pair of driven rollers pressed against one another. The test is carried out by visually monitoring the damage or other condition of the roller surfaces after certain specified time of test duration. The number of cycles for the selected level of damage, which can be judged by the appearance of cracks, surface texture change or spalls, is thus obtained. The longer the test and higher the number of cycles, the greater is its resistance to rolling wear making the tests relatively longer. The main factors for this test are the surface velocities of the rollers, alignment of rollers, and geometrical tolerance of the rollers. This method is mainly used for evaluation of material pairs for rolling applications such as gears, cams, roller bearings, and ball bearings. One main limitation of this test is that the stress levels must remain in the range of the application and also in the elastic range. If plastic deformation occurs, the test becomes invalid. Hence, few materials such as thermoplastics cannot be tested using this method.

34.7 CORROSION PROPERTIES

Corrosion is defined as chemical or electrochemical reaction between a material and its surroundings that leads to deterioration of both the material and its properties at the surface level (Bardal, 2004). Most of the materials undergo some kind of interaction at least to some extent with the environment. The mechanism by which materials deteriorate when attacked by the surrounding is different for different classes of materials (namely, metals, ceramics, and polymers). In metals, it proceeds through material loss by either dissolution or formation of a nonmetallic layer or film (oxidation). Ceramics are resistant to corrosion, but they are also affected at elevated temperatures or adverse environments. In case of polymers, it is more frequently termed as degradation. Polymer can degrade when it absorbs a liquid solvent and swells on exposure.

Corrosive environments include the atmosphere, soils, acids, aqueous solutions, bases, inorganic solvents, salts, liquid metals, and many others. The problem of metallic corrosion is very common and most significant among the three classes of materials in the environments mentioned earlier. Corrosion mechanisms are classified into eight forms namely uniform, galvanic, crevice, pitting, intergranular, selective leaching, erosion-corrosion, and stress corrosion. The problems posed by these forms of corrosion adversely affect the maintenance of critical components. Thus, corrosion is a very important consideration for selection of a material for any application and finds relevance in all fields. For any component to serve the intended purpose at an acceptable level of efficiency, it has to

be properly designed taking into consideration all the factors including properties, performance, life, production, and maintenance cost. The severity of environments encountered by each material is one of the deciding factors for the choice of the material. Corrosion prevention is achieved mainly by careful material selection, environmental alteration, design, application of coatings or films to improve corrosion resistance, and cathodic protection. Some of the methods of corrosion measurement are briefly discussed in the following sections (Bardal, 2004; Shreir, 1963; Bernard and Cramer, 2003; Kruger, 2001; Romaniv et al., 1989).

Corrosion Measurement: Corrosion tests are aimed at determining the corrosiveness of the environment and the rate at which material loss takes place, thus providing a basic structure for the practical control of corrosion. Corrosion measurement is the quantitative method of determination of effectiveness of corrosion control and evaluation of prevention techniques, thus, helping the investigators to optimize the control and prevention methods. Likewise, corrosion monitoring is the practice of measuring and monitoring the corrosivity of any process under practical conditions using probes, which are inserted and continuously exposed to the process stream. These probes may be mechanical, electrical, or electrochemical devices. Some of the techniques measure the corrosion rate or metal loss directly while others give us an indication that a corrosive environment may exist. Corrosion monitoring techniques are used in industries to measure metal loss/corrosion rate in process systems directly. Thus, corrosion measurement, inspection and maintenance, is carried out in industries successfully in order to ensure the effectiveness of any particular component in actual service conditions. The corrosion rate gives an idea of how long any process equipment can be efficiently operated and the corresponding measures should be taken to improve the life of the same.

A large number of corrosion testing procedures exist. The process of corrosion testing and evaluation is subdivided into five steps as given below

- planning corrosion tests and evaluating results;
- performing laboratory corrosion testing;
- performing simulated service corrosion testing;
- performing in-service techniques for damage detection and monitoring;
- evaluating forms of corrosion.

Initially, the surface to be tested has to be prepared in such a way that they duplicate the surface of the material as it would be used, followed by the cleaning to remove any dirt on its surface. The specimen is now ready for corrosion testing, and the appropriate testing procedure has to be decided. This is done depending on the conditions of use by taking into consideration all individual effects of several controlling factors varying one at a time so as to investigate the behavior of the material in the corrosive medium being investigated. Different types of corrosion testing methods available are discussed in the following section.

34.7.1 Electrochemical Methods of Corrosion

This method is mostly used for metallic materials where the corrosion process is normally electrochemical involving a chemical reaction and transfer of electrons from one chemical species to another. During metallic corrosion, oxidation and reduction reactions occur at the metal–electrolyte interface by an electrochemical mechanism.

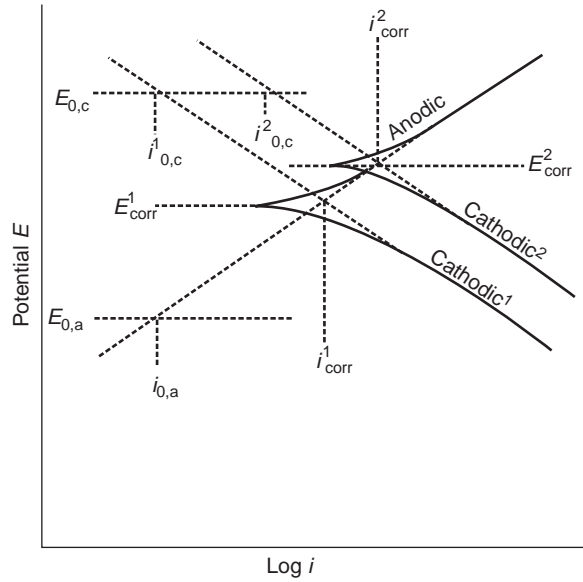


FIGURE 34.10 Corrosion process showing anodic and cathodic current components (Kruger, 2001).

This kind of testing is carried out in an electrochemical cell consisting of a working electrode (the metal under test), counter electrode (platinum), reference electrode (saturated calomel electrode), test cell, potentiostat, voltmeters for monitoring current and potential, and a recording device to record the data obtained during the experiment. In a potentiostatic experiment, a potential is applied and maintained between the reference electrode and working electrode for a specified period of time and the data obtained (tafel plot) is analyzed using software. Metal atoms lose or give up electrons by oxidation reaction, which are transferred to another chemical species by reduction reaction (e.g., hydrogen and oxygen). The rate at which corrosion occurs is determined by the equilibrium between the opposing reactions. The data obtained in this process is recorded in the form of a graph in which the vertical axis is potential and the horizontal axis is the current as shown in Figure 34.10. The sharp point in the curve indicates the point where the current changes sign indicating that the reaction changes from anodic to cathodic at that point.

The equilibrium potential of the metal in the absence of electrical connections is called open circuit potential, E_{oc} , and is a very important parameter, which must be measured. The value of anodic or cathodic current at E_{oc} is called corrosion current, I_{corr} , which is used to calculate the corrosion rate of the metal that is being tested. But, I_{corr} cannot be measured directly and has to be estimated by the electrochemical techniques. A model is generally used for this purpose, and it assumes that the rates of anodic and cathodic processes are both controlled by the kinetics of the electron transfer reaction at the metal surface and is fit using the tafel equation. Tafel equations for both anodic and cathodic reactions are combined for carrying out analysis of a corrosion system as given below

$$I = I_a + I_c = I_a(e^{(2.3(E - E_{oc})/\beta)}) - e^{(-2.3(E - E_{oc})/\beta)} \quad (34.12)$$

where I is the measured cell current (amp), I_{corr} is the corrosion current (amp), E is the electrode potential (volt), E_{oc} is the corrosion potential (volt), β_a is the anodic beta tafel constant (volts/decade), and β_c is the cathodic beta tafel constant (volts/decade).

A log I versus E plot is called tafel plot using which we calculate I_{corr} which is used for corrosion rate calculation using the formula:

$$\text{CR} = \frac{I_{\text{corr}} \times K \times \text{EW}}{dA} \quad (34.13)$$

where CR is the corrosion rate whose units are given by the value of K , I_{corr} is the corrosion current (amp), K is a constant that defines the units for the corrosion rate, EW is the equivalent weight (grams/equivalent), d is the density (gm/cc), and A is the sample area (sq/cm)

But, there are some complications that might lead to nonlinearities in the tafel plot such as concentration polarization, oxide formation, and other effects that alter the surface or occurrence of potential drop. These must be taken care of in order to get accurate analysis of the electrochemical corrosive behavior of metals.

34.7.2 Immersion Testing

This is a laboratory corrosion test most frequently used to evaluate the corrosion behavior of metals in aqueous solutions corresponding to the actual service environment, and hence, the test is adaptable to control important factors that influence results. In this method, the specimen to be tested is totally immersed in a corrosive solution for a period of time, and then removed for examination. A number of factors, such as composition of the solution, temperature, aeration, volume, velocity, method of immersion, duration of test, and method of cleaning of specimens after test, must be considered in order to achieve specific results and ensure adequate reproducibility of the results. It is possible to carry out the test under conditions similar to those encountered in practical situations, but it is very difficult to obtain reliable analysis. Various devices are used to investigate the effect of three main factors namely aeration, temperature, and velocity. Statistical analysis of the data obtained by these apparatus has demonstrated satisfactory reproducibility of the results. In order to achieve better control over the aforementioned parameters, alternating immersion tests are employed where the specimen is immersed in a solution and withdrawn from it in a pre-determined cycle. These tests may simulate certain circumstances of practical corrosion better and are hence preferred. These tests can also be used to evaluate the resistance of the metal to pitting, crevice, galvanic corrosion, hydrogen embrittlement and stress corrosion cracking by employing specifically designed specimens or environments.

34.7.3 Accelerated Tests

In practical situations where materials encounter alternating wet and dry cycles, the life of the system is severely affected as these conditions impose severe corrosion problems on the material. This is mainly due to the formation of partially dry corrosion products that can enhance the accumulation of corrosive agents in terms of absorption of moisture. The intensity of the corrosion attack induced by alternate wetting cycles helps one in estimating the corrosion prevention measures to be undertaken to prevent the accelerated corrosion. Various tests namely electrolytic tests, salt spray tests, corrodokote test, sulphur dioxide tests, and humidity tests are used to demonstrate

the accelerating corrosion damage in a relatively short period. They aim at measuring the accelerating corrosion damage under different conditions including increasing aeration for immersion tests, temperature, acidity of the corrosive medium, relative humidity, and velocity, which would be important in actual service conditions. By examining the above conditions, it is evident that caution must be exercised to get meaningful results in accelerated testing. There are a numerous errors that might result due to multiple failure modes, misapplied models, uncertainty of the model, unquantifiable corrosion factors, unexpected change in the factors during testing, and so on. Hence, one must be extremely cautious while doing these kinds of tests in order to predict the lifetime of a material for any particular application. Accelerated testing is a powerful tool which can be used to examine potential environments for new materials, and thus selecting a material in relation to its corrosion resistance and also determining the best method of corrosion prevention.

34.7.4 Atmospheric Tests

It is very common that every component will be exposed to atmosphere during its service and might be attacked by the atmospheric agents, which has to be studied carefully for improving the performance of the same. The simplest form of atmospheric corrosion measurement is the coupon exposure method. Here, flat panels are mounted and exposed in exposure racks. These coupons are subjected to weight loss measurements, pit density, depth measurements, and other kinds of examination subsequent to their exposure. The main disadvantage with this kind of test is that the specimen has to be exposed for extremely long periods of time to undergo corrosion and obtain meaningful data.

34.8 STANDARDS FOR SURFACE ENGINEERING MEASUREMENT

It is very important to know the standard test methods for any property measurement or material characterization as certified by many researchers to get accurate results and also to avoid ambiguity in the obtained results (Powell, 1978). They give a fairly good idea to an inexperienced person in a specific field to quickly gain knowledge about the measurement technique and parameters involved. There are three ways to identify a standard of interest as follows:

- surface condition based;
- property based;
- standard based.

Listed below are some important standards for surface properties measurements.

ASME B46.1-2002	Surface texture (surface roughness, waviness and lay)
ASME SA-370	Methods and definitions for mechanical testing of steel products
ASTM A 833	Standard practice for indentation hardness of metallic materials

ASTM B 277	Standard test method for hardness testing of electrical contact materials
ASTM B 294	Standard test method for hardness testing of cemented carbides
ASTM B 578	Standard test method for microhardness of electroplated coatings
ASTM B 721	Standard test method for microhardness and case depth of powder metallurgy (P/M) parts
ASTM C 1326	Standard test method for Knoop indentation hardness of advanced ceramics
ASTM C 1327	Standard test method for Vickers indentation hardness of advanced ceramics
ASTM D 5183	EN-standard test method for determination of the coefficient of friction of lubricants using the four-ball wear test machine
ASTM E 3	Standard guide for preparation of metallographic specimens
ASTM E18	Standard test methods for Rockwell hardness and Rockwell superficial hardness of metallic materials
ASTM E 92	Standard test method for Vickers hardness of metallic materials
ASTM E 140	Standard hardness conversion tables for metals relationship among Brinell hardness, Vickers hardness, Rockwell hardness, Rockwell superficial hardness, Knoop hardness, and scleroscope hardness
ASTM E 384	Standard test method for microindentation hardness of materials
ASTM E399	Standard test method for linear-elastic plane-strain fracture toughness K_{IC} of metallic materials
ASTM E 673	Standard terminology relating to surface analysis
ASTM E 1127	Standard guide for depth profiling in auger electron spectroscopy
ASTM G 32	Standard test method for cavitation erosion using vibratory apparatus
ASTM G 65	Standard test method for measuring abrasion using the dry sand/rubber wheel apparatus
ASTM G 73	Standard test method for liquid impingement erosion using rotating apparatus
ASTM G 75	Standard test method for determination of slurry abrasivity (Miller number) and slurry abrasion response of materials (SAR number)
ASTM G 76	Standard test method for conducting erosion tests by solid particle impingement using gas jets
ASTM G 77	Standard test method for wear resistance by block-on-ring wear test

ASTM G99	Standard test method for wear testing with a pin-on-disk apparatus
ASTM G 105	Standard test method for conducting wet sand/rubber wheel abrasion tests
ASTM G 133	Standard test method for linearly reciprocating ball-on-flat sliding wear
ASTM G190	Standard guide for developing and selecting wear tests
ASTM G 171	Standard test method for scratch hardness of materials using a diamond stylus
ISO 3274	Geometrical product specifications (GPS)—surface texture: profile method—nominal characteristics of contact (stylus) Instruments
ISO 3878	Hard metals—Vickers hardness test
ISO 4287	Geometrical product specifications (GPS)—surface texture: profile method—terms, definitions and surface texture parameters
ISO 4288	Geometrical product specification (GPS)—surface texture: profile method—rules and procedures for the assessment of surface texture
ISO 4545	Metallic materials—hardness test—Knoop test part 1: test method
ISO 6507-1	Metallic materials—Vickers hardness test part 1: test method
ISO 9277	Corrosion tests in artificial atmospheres—salt spray tests
ISO 11562	Geometrical product specifications (GPS)—surface texture: profile method—metrological characteristics of phase correct filters
ISO 12085	Geometrical product specification (GPS)—surface texture: profile method—motif parameters
ISO 14577-1	Metallic materials—instrumented indentation test for hardness and materials parameters part 1: test method
SAE J 448a	Surface texture (SAE standard)
SAE J 449a	Surface texture control (SAE recommended practice)

REFERENCES

- Anstis GR, Chantikul P, Lawn BR, Marshall BR. A Critical Evaluation of Indentation Techniques for Measuring Fracture Toughness: Direct Crack Measurements. *Journal of American Ceramic Society* 1981;62:533–538.
- Bardal E. *Corrosion and Protection*. London: Springer Publications; 2004.
- Bayer RG. *Mechanical Wear Fundamentals and Testing*. 2nd ed. New York: Marcel Dekker Inc.; 2004.
- Becker WT, Shipley RJ. *ASM Handbook Volume 11: Failure Analysis and Prevention*. 10th ed. ASM International; 2002.

- Bernard S Covino, Cramer SD. *ASM Handbook Volume 13A: Corrosion: Fundamentals, Testing and Protection*. 10th ed. ASM International; 2003.
- Bhushan B. *Handbook of Micro/Nano Tribology*. 2nd ed. CRC Press; 2000.
- Bhushan B. *Modern Tribology Handbook*. Vol 1, CRC Press; 2001.
- Bhushan B. *Springer Handbook of nanotechnology*. 2nd ed. Springer Inc.; 2007.
- Blau PJ. *ASM Handbook Volume 18: Friction. Lubrication and Wear Technology*. ASM International; 1992.
- Bozzola JJ, Russell, LD. *Electron Microscopy: Principles and Techniques for Biologists*. 2nd ed. Jones and Bartlett Publishers; 1999.
- Callister WD. *Materials Science and Engineering – An Introduction*. 6th ed. New York: John Wiley & Sons; 2006.
- Chandler H. *Hardness Testing*. 2nd ed. ASM International; 1999.
- Chattopadhyay R. *Surface Wear—Analysis, Treatment and Prevention*. ASM International; 2001.
- Crankovic GM. *ASM Handbook Volume 10: Materials Characterization*. 9th ed. ASM International; 1986.
- Creath K, Wyant JC. Absolute measurement of surface roughness. *Applied Optics* 1990;29:3823–3827.
- Cullity BD. *Elements of X-Ray Diffraction*. 2nd ed. Addison-Wesley Publishing Company Inc.; 1978.
- Fischer AC. *Cripps, Nanoindentation*. 2nd ed. Springer Inc.: New York; 2004.
- Gellman AJ, Ko JS. The current status of tribological surface science. *Tribology Letters* 2001;10:39–44.
- Krawitz AD. *Introduction to Diffraction in Material Science and Engineering*. John Wiley & Sons: New York; 2001.
- Kruger J. Electrochemistry of corrosion. *Electrochemistry Encyclopedia*. John Hopkins University, Baltimore, USA; 2001.
- Kuhn H, Medlin D. *ASM Handbook Volume 8: Mechanical Testing and Evaluation*. 10th ed. ASM International; 2000.
- Lawn BR, Marshall DB. Hardness, toughness and Brittleness: An Indentation Analysis. *Journal of American Ceramic Society* 1979;62:347–350.
- Li X, Bhushan B. A review of nanoindentation continuous stiffness measurement technique and its applications. *Materials Characterization* 2002;48:11–36.
- Ludema KC. *Friction, Wear, Lubrication—A textbook in tribology*. CRC Press; 1996.
- Pabst RF, Kromp K, Popp G. Fracture Toughness—Measurement and Interpretation. *Proceedings of the British Ceramic Society* 1982;32:89–106.
- Polak TA, Pande C. *Engineering Measurements—Methods and Intrinsic Errors*. Suffolk, UK: St Edmundsbury Press Limited; 1999.
- Powell CJ. Surface characterization: present status and the need for standards. *Applications of Surface Science* 1978;1:143–169.
- Rabinowicz E. *Friction and Wear of Materials*. New York: John Wiley and Sons; 1995.
- Reidenbach F, Cotell CM, Sparague JA, Smidt FA. *ASM Handbook Volume 5: Surface Engineering*. 9th ed. ASM International; 1994.
- Rok L. Measurements of surface properties in the nano- and microscale using optical, mechanical, and scanning probe methods. *Acta Physica Polonica A* 2009;116:S180–S183.
- Romaniv ON, Tsirulnik AT, Kryskiv AS, Ronchevich I. Electrochemical Methods in Corrosion Monitoring of Metals (review). *Materials Science* 1989;25:1–12.

- Saltiel C, Giesche H. Needs and opportunities for nanoparticle characterization. *Journal of Nanoparticle Research* 2000;2:325–326.
- Shreir LL. *Corrosion, Vol 2: Corrosion Control*. New York: John Wiley & Sons; 1963.
- Stachowiak GW. *Wear—Materials, Mechanisms and Practice*. New York: John Wiley & Sons; 2005.
- Stanley JD. *Surface Engineering Measurement Standards for Inorganic Materials, 960-9*. U.S. Government Printing Office: Washington; 2005.
- Strecker K, Ribeiro S, Hoffmann MJ. Fracture toughness measurements of LPS-SiC: a comparison of the indentation technique and the SEVNB method. *Materials Research* 2005;8:121–124.
- Takadom J. *Materials and Surface Engineering in Tribology*. John Wiley & Sons; 2008.
- Totten GE, Liang H. *Mechanical Tribology—Materials, Characterization and Applications*. New York: Marcel Dekker Inc.; 2004.
- Voort GFV. *ASM Handbook Volume 9: Metallography and Microstructures*. ASM International; 2004.
- Zinkle SJ, Ice GE, Miller MK, Pennycook SJ, Wang XL. Advances in Microstructural Characterization. *Journal of Nuclear Materials* 2009;386:8–14.

35

THERMAL CONDUCTIVITY OF ENGINEERING MATERIALS

JUERGEN BLUMM

- 35.1 Introduction
 - 35.1.1 Thermal conductivity requirements of modern technology
 - 35.1.2 Thermal conductivity—background and basic equations
 - 35.1.3 Modes of heat transfer in materials
 - 35.1.4 Phonon contribution to the heat transfer
 - 35.1.5 Electron contribution to the heat transfer
 - 35.1.6 Radiative heat transfer
 - 35.1.7 Convective heat transfer
 - 35.2 Stationary methods for measurement of the thermal conductivity
 - 35.2.1 Heat flow meter method
 - 35.2.2 Guarded hot plate method
 - 35.3 Transient methods for the measurement of the thermal conductivity
 - 35.3.1 Hot wire method
 - 35.3.2 The flash method
 - 35.4 Test results on various engineering materials
 - 35.4.1 Results on various kinds of materials
 - 35.4.2 Full thermophysical properties characterization of a polymer and a ceramic
- References

35.1 INTRODUCTION

35.1.1 Thermal Conductivity Requirements of Modern Technology

In many fields of modern engineering, physical material properties are becoming more and more important. Modern technical products such as cars, computers, space craft, or modern production machines would not be possible without significant improvements on

the material's side. The physical properties of the materials must be known and the changes of these properties, in case of modifications in composition and structure, must be understood. The thermal conductivity is one of these properties, which was underestimated in the past. In recent decades, however, the thermal conductivity has become more important for a wide range of applications. For modern alloys, the thermal conductivity must be known and optimized if they are utilized in engines or are used as model materials in polymer production machines. In the case where incorrect alloys are utilized or if the selected material does not offer the required thermal conductivity range, the performance of the final parts and products can be significantly compromised. Furthermore, the thermal conductivity is a crucial parameter for the simulation of casting processes. If the thermal conductivity (together with other parameters) is known in the solid and liquid, scientists can simulate the solidification process inside a mold, optimize the mold and cooling accessories, and make sure that the final part is produced without pores in the shortest possible time. Optimization by simulation is not limited to casting processes. Using finite element software, computer-based optimization of parts produced during a sintering process such as powder metals or ceramics is possible. Tailor-made materials with optimized thermal conductivity for a specific application can also be realized. In many cases, characterization of the thermal conductivity of a ceramic part has a crucial impact on the downstream application even if one does not expect it. For example, if the thermal conductivity of a ceramic brake disk is too low, hot spots will be generated on the brake during operation, which can in the worst case cause cracks and mechanical damage of the entire system. Of course, there are other examples, demonstrating the importance of this thermophysical property. Nowadays, reduction of the heating/cooling consumption of buildings is one of the key issues to reduce consumption of primary energy resources. To achieve this, insulating materials have been optimized for decades. The new requirements in some areas, however, made it necessary to develop completely new kinds of insulations, offering minimal thickness and at the same time outstanding heat transfer properties. Vacuum insulation panels are one development resulting from this. Installing a vacuum insulation panel with a thickness of 2 cm on a house wall or in a refrigerator can offer the same insulation performance as a 10- or 20-cm thick conventional foam or fiber insulation.

35.1.2 Thermal Conductivity—Background and Basic Equations

Now the question might be, what is thermal conductivity? Generally, you find definitions in many textbooks or on the Internet (Incropera and DeWitt, 1996, http://en.wikipedia.org/wiki/Thermal_conductivity). The thermal conductivity is generally described as a material property defining the material's capability to transfer heat. In other words, it means that if a temperature difference over the thickness of the material exists, the thermal conductivity describes how much heat is transferred through it. To make things clearer, one can imagine a simple experiment. There is a room inside a building with one wall to the outside. It is winter and the outside temperature is 0°C. The temperature in the room is similar to the entire building at 20°C. To keep the temperature inside the room at 20°C, we need a heating system. Without heating, the temperature will not stay at this temperature but will go down. If the wall is made out of 20 cm of dense concrete, we need to heat a lot. If we build the wall out of a 20-cm thick fiber insulation, the heating will be more than 30 times less. This means, by changing the wall material, I can

significantly influence the heat transfer through it and therefore the consumption of heat to keep the temperature inside the room constant. I can extend this experiment. Increasing the size of the wall will have an impact. Having the room at the corner of the building with two outside walls of the same size will increase the heating consumption by a factor of 2. Increasing the thickness of the wall will also have an impact. Doubling the thickness will reduce the heat flow through it by 50%. Taking into account outside temperature changes will result in another effect. If the temperature goes down to -20°C , the heating consumption will increase by a factor of 2. The reason is that the temperature difference over the wall thickness will increase to 40 K.

Of course, this experiment is greatly simplified. However, it helps us generate a simple equation:

$$\dot{Q} = -\lambda \cdot A \cdot \frac{\Delta T}{\Delta x} \quad (35.1)$$

\dot{Q} is the heat flow through the wall. This is nothing more than the heat going through the wall per time. A is the area of the wall. ΔT is the temperature difference between the outside of the wall and the inside. Δx is the thickness of the wall. λ is a property that depends on the material. We call it thermal conductivity. The standard unit of the thermal conductivity is $\text{W}/(\text{m K})$.

In many cases, the heat flow through the wall per unit area \dot{q} is used in Equation (35.1):

$$\dot{q} = -\lambda \cdot \frac{\Delta T}{\Delta x} \quad (35.2)$$

This is the simplest form of Fourier's law, as it was developed by Jean Baptiste Joseph Fourier in 1822. It describes the heat transfer in one dimension under stationary conditions. This means that the heat flow goes in one direction only, and no temperature change occurs inside or outside of our wall. The general form allowing a heat flow in all directions is as follows:

$$\vec{\dot{q}} = -\lambda \cdot \text{grad } T \quad (35.3)$$

In many cases, this equation is used to describe the heat flow through a material under stationary or quasi-stationary conditions. This means that temperature changes only occur on a long-term scale. For building refrigerators, this is a proper approach. For other processes, it is at least somewhat critical. In many applications we face a problem of temperature changes on the surfaces of a material. Here, we need a more complex differential equation:

$$\frac{\partial T}{\partial t} = \frac{\lambda}{\rho \cdot c_p} \cdot \frac{\partial^2 T}{\partial x^2} \quad (35.4)$$

where T is defined as temperature changes, t the time. ρ and c_p are the density of material and specific heat, respectively. Again, this is the simplest form of transient heat transfer

equation without any heat sources, only considering one-dimensional heat flow. Generally, one has to consider all dimensions. A more general formula is as follows:

$$\frac{\partial T(\vec{x}, t)}{\partial t} = \frac{\lambda}{\rho \cdot c_p} \cdot \Delta T(\vec{x}, t) + \dot{q} \quad (35.5)$$

Here, \dot{q} is the heat generation inside the material, which is in most practical cases 0. Furthermore, the quotient of thermal conductivity and density and specific heat is often defined as the thermal diffusivity a :

$$a \equiv \frac{\lambda}{\rho \cdot c_p} \quad (35.6)$$

The thermal diffusivity is also a material property. It is a material property describing how fast a material reacts to temperature changes. The thermal conductivity and thermal diffusivity are closely related to each other for dense solids. Materials with a high thermal conductivity have a high thermal diffusivity and vice versa.

For the solution of a heat transfer problem, Equation (35.6) has to be solved. Solving this equation is not a straightforward process. There is no general solution available. One has to consider the geometry, the initial boundary conditions, as well as the initial conditions. For a wide range of different heat transfer problems, Carslaw and Jaeger (1959) provided solutions in their book. Even though it was published in 1959, it is still one of the leading books related to the mathematical treatment of heat transfer problems.

Even though Equation (35.6) can be even more complex, one should go into this transient heat transfer equation in more detail. It is a second order partial differential equation describing diffusion of any kind of process. It can be used to carry out a heat conduction analysis as long as the heat transfer is based on the diffusion process. In dense solid materials, this is generally the case. For porous materials, and for gases or liquids, however, the situation can be different. Here, one has to consider other possible means of heat transfer.

Anyhow, let us get back to the thermal conductivity itself. The thermal conductivity is a material property. It can depend on the chemical and structural composition of a material. The best conducting macroscopic materials we know nowadays are diamond (up to 2000 W/(m K)) and highly conducting metals (several 100 W/(m K)). Graphenes and carbon nanotubes also offer high thermal conductivities. However, these materials are only available in very small dimensions. The lowest thermal conducting materials are vacuum insulation panels (VIP). Such structures can offer thermal conductivities in the range of a few mW/(m K). Especially for insulation purposes, such materials are getting increasingly important. Based on VIPs, one can build a refrigerator with a wall thickness of 1 cm. However, the insulation of such a wall is as good as a 10 cm foam insulation generally used today. An overview of the thermal conductivity is given in Figure 35.1. Here, various materials are presented with their thermal conductivity at room temperature. It must be pointed out that some of the materials mentioned can have a large variability in the thermal conductivity. For many materials, it is difficult to estimate the thermal conductivity as impurities, additives or the internal structure can have a crucial impact on the thermal conductivity values.

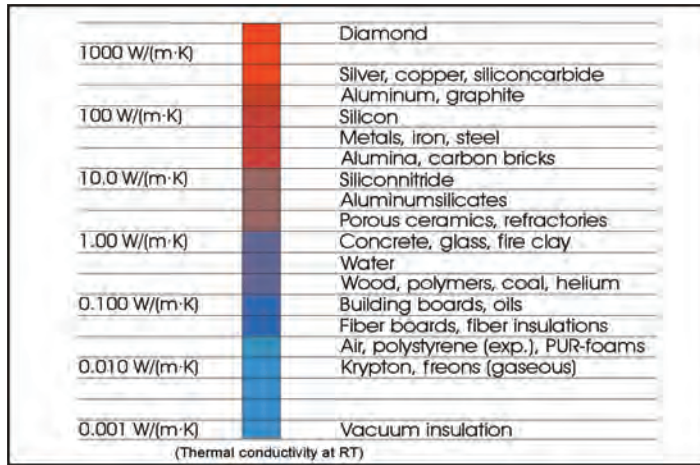


FIGURE 35.1 Overview of the thermal conductivity of different materials at room temperature.

For example, silicon carbide can offer a thermal conductivity of 400 W/(m K) if it is available as a pure single crystal. If it is a polycrystalline material with certain porosity and some additives inside, the thermal conductivity can be 20–50 times lower. To understand this, one has to analyze the heat transfer modes in more detail. If a heat transfer mode is very efficient in a material, it will directly lead to a high thermal conductivity.

35.1.3 Modes of Heat Transfer in Materials

Everything mentioned above is valid for the most important heat transfer mechanisms in solid materials. However, it is not valid for all kinds of heat transfer. In general, heat can be transferred via conduction, radiation, and convection. All these heat transfer modes must be considered to get a full picture of the amount of heat transferred through a material or structure.

35.1.4 Phonon Contribution to the Heat Transfer

Conduction is the most important heat transfer mechanism in solids. In solid materials there are two ways that heat can be transferred. The first one is phononic heat transfer. It is based on lattice vibrations and their movement through a solid structure. A mathematical description requires a thermodynamic approach. Such an explanation is given, for example, in (Reif, 1965; Salmang and Scholze, 1982):

$$\lambda_{\text{Ph}} = \frac{1}{3} \cdot c_V \cdot \bar{v} \cdot l_{\text{Ph}} \quad (35.7)$$

The phononic contribution to the thermal conductivity λ_{Ph} is dominated by the changes in specific heat at a constant volume c_V , the sound velocity \bar{v} , and the mean free path length of the phonons l_{Ph} . The temperature dependence of the heat transfer is easy to explain. At very low temperatures, the temperature dependence is dominated by the changes in specific heat c_V . Here the thermal conductivity is proportional to T^3 . The

sound velocity shows no temperature dependence, while the mean free path length is only limited by the crystal size (in a perfect crystal) or by the crystallite size in a polycrystalline material. At high temperatures, the specific heat is nearly constant. The mean free path length, however, gets shorter. With increasing temperatures, the probability of phonon–phonon scattering effects increase, resulting in a shorter mean free path length of the phonons. This means that the mean free path length l_{ph} of the phonons is proportional to $1/T$. As the sound velocity is still independent from temperature, the changes in the phononic thermal conductivity follow the changes in the mean free path length. This is the dominating factor for temperatures above room temperature. Therefore, most crystalline insulating materials show a decrease in thermal conductivity with increasing temperature proportional to $1/T$.

35.1.5 Electron Contribution to the Heat Transfer

A similar approach can be carried out to analyze the contribution of free electrons in electrically conducting solids. In metals, electrons can move more or less freely through the solid structure. Therefore, they can also carry heat through the structure. In good electrically conducting metals, this heat transfer mechanism is by far the dominant one. This strong correlation between electrical and thermal conductivity was the basis of an empirical law, the Wiedemann–Franz law. It states that the quotient of thermal conductivity and electrical conductivity is a constant multiplied by temperature:

$$\frac{\lambda}{\sigma} = L \cdot T \quad (35.8)$$

In this equation, σ is the electrical conductivity and L is the Lorenz number. The Lorenz number is equal to $2.44 \times 10^{-8} \text{ W}\Omega\text{K}^{-2}$ (http://en.wikipedia.org/wiki/Wiedemann%E2%80%93Franz_law). This number, however, is only valid for highly conducting metals. For low electrically conducting metals, the number requires some adjustments to offer reliable correlations. For practical estimations, it is still a good approach. For accurate data, however, a separate measurement of both parameters would be recommended.

There might be the question of the temperature dependence of the electrical contribution to the thermal conductivity. This is a bit more complex. Explanations are generally based on the analysis of the scattering effects inside the material. First of all, one has to consider the number of electrons inside a material which generally increases versus temperature. However, the electrons can be scattered on defects inside the crystal structure. As long as there are no changes in the impurity content or in the defect structure, this scattering mechanism has no temperature dependence. There is another scattering mechanism, which is related to electron–phonon scattering. This depends on the number of phonons and shows a temperature dependence equal to T^{-3} . As a result, the temperature dependence of the electronic contribution is as follows:

$$\lambda_{el} \propto \frac{T}{a + b \cdot T^3} \quad (35.9)$$

There is even a more mathematical approach to develop and explain the temperature dependence of the electronic contribution that is given elsewhere (Tritt, 2004). This is generally referred to as the Drude-model.

35.1.6 Radiative Heat Transfer

Of course, heat can be transported via radiation from one place to another. This heat transfer mechanism is the dominant factor for bringing solar energy to earth. Anyhow, one has to consider that radiation is generally described by the Planck's law or Stefan-Boltzmann's law. Radiation is different than the other mechanisms mentioned prior. It does not require any media. It also works in vacuum. Furthermore, it is not a diffusion process and therefore, it is not considered in the equations in Section 35.1.2. One should consider that radiation can also have an effect on the effective heat transfer in solids. In translucent materials such as glasses or in semitransparent or translucent materials such as ceramics or foam and fiber structures, radiation can be the dominant heat transfer mechanism at high temperatures. Descriptions of the heat transfer generated by radiation can be found in various books such as (Siegel and Howell, 2002) and are not explained here in detail. Especially in insulating materials, this factor has to be considered.

35.1.7 Convective Heat Transfer

Convection is a heat transfer which does not occur in solid materials. However, in liquids or gases, this mechanism can be dominant. We can visualize convection with a simple experiment. By putting a pot of water on a kitchen stove and simply watching what happens after switching on the stove. After a few minutes, one can see that the hot water from the bottom is moving upwards and the cooled water is going down. We generated convective heat transfer inside the pot. It simply refers to the transport of heat by material transport. The mathematical treatment of convective heat transfer can be quite complex. As it generally does not occur in engineering materials, we refer to the corresponding literature related to this topic such as (Bejan, 2004).

In solid materials, generally the phononic and the electronic contributions are the dominant heat transfer mechanisms. In insulating materials, the situation gets more complex. Such materials generally have a solid structure and a lot of pores inside. Here, we do not have only the solid contribution. The heat transfer by the gas inside the pores has a significant impact as well. Furthermore, the materials are often translucent. Therefore, the radiation has a contribution as well. Around room temperature, this is generally in the range of a several percent of the entire heat transfer. In addition, there can be a coupling contribution resulting from the fact that several heat transfer mechanisms act parallel to each other. Therefore, we generally work with the effective thermal conductivity which is the sum of all contributions:

$$\lambda_{\text{effective}} = \lambda_{\text{solid structure}} + \lambda_{\text{gas}} + \lambda_{\text{radiation}} + \lambda_{\text{coupling}} \quad (35.10)$$

35.2 STATIONARY METHODS FOR MEASUREMENT OF THE THERMAL CONDUCTIVITY

Over the last few decades, various methods have been developed for characterizing the thermal conductivity of materials. The first methods employed for the measurement of this material property were based on the Fourier's equation (35.1). Nowadays, such stationary methods are used to analyze building materials and insulations. Generally, there are two standardized methods available for characterizing the effective thermal conductivity or the thermal resistance of low-conducting materials.

35.2.1 Heat Flow Meter Method

The heat flow meter technique (Blumm, 2002) is a fast and reliable tool for measurements around room temperature. The heat flow meter method (ASTM C 518, ISO 8301, EN 13163, etc.) is the most commonly used method for quality control and R&D. Modern heat flow meters are designed to offer high accuracy, fast evaluation, and ease-of-use. A typical heat flow meter system is shown in Figure 35.2.

In a heat flow meter, the sample which is generally several 10's of centimeters' long and wide and several centimeters' thick is sandwiched between a hot and a cold plate. In most instruments used nowadays, the temperature of the plates is set via a peltier system. Such systems offer homogeneous plate temperatures and a low thermal mass of the plates. This is one of the requirements to improve the testing speed. In the surface of the plates, thermocouples are embedded to measure the temperatures and to determine and control the difference between the hot plate and the cold plate. Additionally, heat flux transducers are installed in the central area (metering section) to determine the heat flow from the hot plate to the cold plate. The LVDT system (linear variable displacement transducer) measures the distance between the plates and therefore the sample thickness. By knowing the sample thickness, the temperature difference between the plates and the heat flow, the thermal conductivity can be determined according to Equation (35.1). For this, another point needs to be ensured. The heat flow should be one-dimensional. Therefore, the sample is generally made larger and the metering section is surrounded by a heated area (guard section). The schematic design is illustrated in Figure 35.3. For the general application in research and quality control, instrument specifications and design details must be prepared in such a way that the requirements of the current standards such as ISO 8301 or DIN EN 12667 are fulfilled. For example, in all instruments, the plate temperatures are measured at three different positions and therefore, requirements for plate flatness and emissivity are met.



FIGURE 35.2 Netzsch model HFM 436/3 Lambda Heat Flow Meter system.

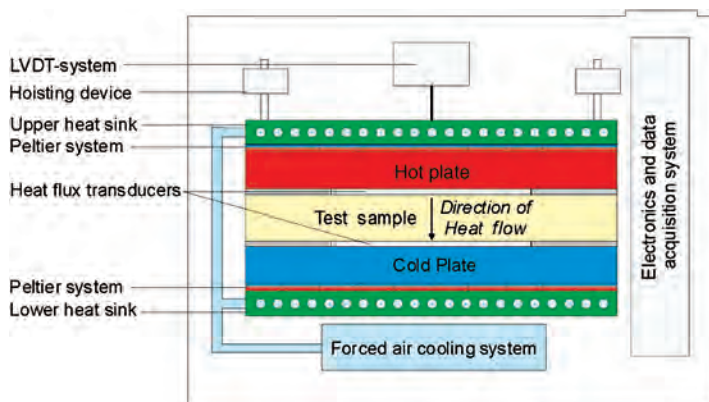


FIGURE 35.3 Schematic design of the NETZSCH Heat Flow Meter HFM 436/3 Lambda.

Depending on the instrument version, tests are possible around room temperature (at a fixed mean sample temperature between 10 and 30°C). With the extended temperature versions, temperature-dependent tests can be carried over a broader temperature range. Samples are generally 30 by 30 cm in square with a thickness of up to 10 cm. One drawback of heat flow meter systems is that they are not absolute measurement devices. Prior to a sample test, the system must be calibrated. The calibration is necessary to correlate the output of the heat flux transducers (generally a voltage) and the heat flow going through it. This is done by putting a sample with known thermal conductivity inside the machine, carrying out a test and determining a calibration factor:

$$N = \frac{\lambda_{\text{cal}}}{V_{\text{hft}}} \cdot A \cdot \frac{T_u - T_l}{d} \quad (35.11)$$

The calibration factor N is given by the known thermal conductivity of the calibration standard λ_{cal} , the output of the heat flux transducer V_{hft} , the metering area A , the temperature difference between the upper and lower plate and the thickness of the calibration standard. If this is done, the calibration factor can be used to for any subsequent sample test. Of course, it is recommended to check the calibration factor from time to time. The speed of a heat flow meter test is one of the main reasons for the success of this method. Exchanging samples can be done within seconds and the test itself lasts generally no more than 10–20 min. A quasi-equilibrium can, in many cases, be reached in a matter of minutes, as the heat flux stabilizes and the measured thermal conductivity settles into its steady-state value even before the sample reaches total steady-state conditions. The dual heat flux transducer configuration—one on each plate—together with the low thermal mass of the plates and the fast control algorithms is the key for the fast test speed. The measurement on 10 different EPS specimens depicted in Figure 35.4 requires approx. 15 min per sample. Two test cycles were carried out. One test series was carried out at a mean sample temperature of 24°C according to ISO 8301 and DIN EN 12667. A second test series was carried out at $10 \pm 0.3^\circ\text{C}$ according to DIN EN 13163, a standard established specifically for those kinds of insulating polystyrene foams. In all cases, the

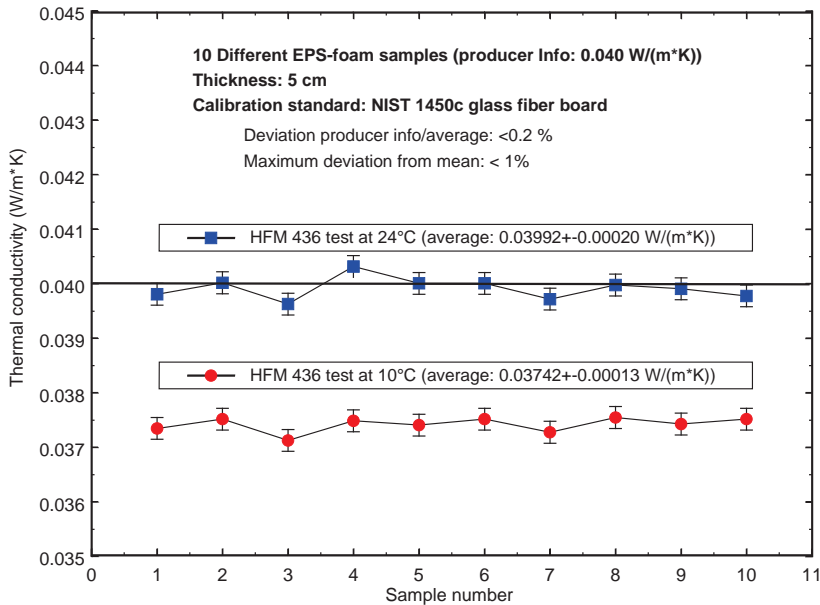


FIGURE 35.4 Quality control runs by means of an HFM 436 at mean sample temperatures of 24 and 10°C. Ten different specimens were used for this test cycle. The average testing time/sample was 15 min.

thermal conductivity shows a scattering of less than 1%. Furthermore, the results deviated by less than 0.5% from the manufacturers' specification.

Sometimes, it is useful to have data available not only at one specified temperature but over a broader temperature range. Here, the heat flow meter test can also be used. However, going from temperature to temperature requires longer testing times. An example for temperature-dependent tests on two PUR foams is presented in Figure 35.5. This test was carried out at a mean sample temperature between 0 and 40°C. Here, the reliability of the method even on temperature-dependent test can easily be seen. Modern instruments can cover an even broader temperature range. However, it is generally limited to a range between −20 and 90°C.

35.2.2 Guarded Hot Plate Method

The limited temperature range, the need for a calibration standard, and the limitation regarding tests in controlled atmospheres are the main drawbacks of the heat flow meter method. The guarded hot plate overcomes these problems. No system calibration is required as long as the uncertainty of the components and measurement devices offers a high accuracy. However, guarded hot plate tests often require long measurement times (days). Furthermore, the sample change is more complex compared to heat flow meter systems.

Modern guarded hot plate systems such as the new NETZSCH model GHP 456 *Titan* overcome most of the existing problems. The system is vacuum-tight by design. Therefore, measurements are possible under well-defined atmospheres. Even gas pressure-dependent tests are possible. This allows, for example, determination of the influence of the gas inside the pores on the effective thermal conductivity. The system is designed to

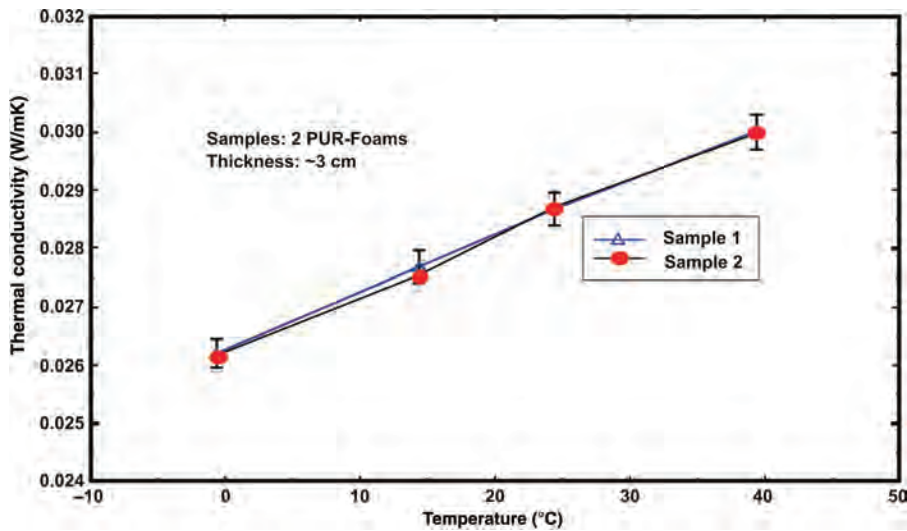


FIGURE 35.5 Low-temperature test on two different PU foams from the same batch. The thermal conductivity of the material was measured between 0 and 40°C on two different samples. The differences were less than 1%.

measure square samples 300 by 300 mm in size and up to 100 mm thick. Depicted in Figure 35.6 is a schematic design of a typical guarded hot plate system.

The hot plate, 150 by 150 mm in size, is sandwiched between two samples of the same material. The hot plate is surrounded by a 75 mm wide guard ring which is controlled to the same temperature as the hot plate and therefore prevents radial heat transfer laterally from the hot plate. Two auxiliary heaters are placed above and below the specimens. The

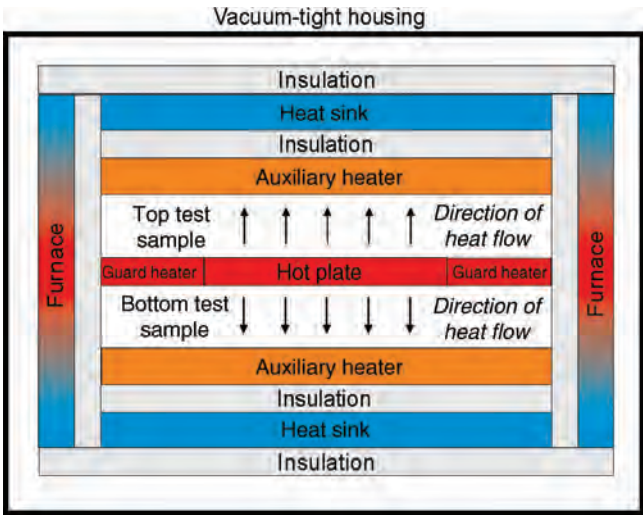


FIGURE 35.6 Schematic design of a state-of-the-art guarded hot plate system.

auxiliary heaters are generally brought to a temperature of 10–20 K below the hot plate temperature. Heat sinks with integrated cooling tubes are positioned above and below the auxiliary heaters. Temperature measurement of the plates is made with 28 sheeted PT100 temperature sensors with a temperature resolution of 0.001 K. The entire plate stack is surrounded by a sectional furnace (artificial environment). The furnace is controlled in such a way that it simulates a comparable temperature profile that occurs in the plate stack. This avoids radial heat loss from the plate stack. Furthermore, all wires from the plates (power cables, temperature sensor wires) are connected to the furnace. Therefore, no significant temperature differences occur. This prevents any heat flow via the wiring and avoids possible wire heat losses. The system can be equipped either with aluminum alloy plates for the temperature range between -160 and 250°C (mean sample temperatures) or with special high-temperature alloy plates for the range between -160 and 700°C (mean sample temperature). The temperature difference can generally be selected between 5 and 40 K. The system is designed for square samples 300 by 300 mm in size. The thickness can be between 5 and 100 mm.

The working principle is quite simple. The guard ring and the artificial environment ensure a one-dimensional heat transfer from the hot plate to the auxiliary heaters. By measuring the sample thickness, the temperature difference between the plates and the power input into the electrically heated hot plate under stationary conditions, the thermal conductivity can be directly determined:

$$\dot{Q} = -\lambda \cdot 2 \cdot A \cdot \frac{\Delta T}{\Delta x} \Rightarrow \lambda_{\text{eff}} = -0.5 \cdot (I \cdot U) \cdot \frac{1}{A} \cdot \frac{\Delta x}{\Delta T} \quad (35.12)$$

It must be considered that there is a slight modification compared to Equation (35.1) due to the fact that the heat flow from the hot plate that goes up and down. Therefore, the hot plate area must be considered twice.

A more realistic view into the instrument is presented in Figure 35.7. Here, the hoisting device can be seen on both sides of the instrument. The hoist opens the vacuum chamber, sectional furnaces, and the plate stack. In less than a minute, the operator has access to the test section for changing samples.

A photo of a modern system including the vacuum-tight housing, and the power supplies is shown in Figure 35.8. The complexity of a guarded hot plate system is clearly seen.

Depicted in Figure 35.9 are the results for the thermal conductivity measured on the NIST-certified standard reference material SRM 1450c high-density glass fiber board (Gills, 1997). The material is certified between 280 and 340 K (7 – 67°C). The measurements with the guarded hot plate system were carried out between 25 and 60°C . It can clearly be seen that the measured results are easily within the stated uncertainty of the standard material (2.6% according to (Gills, 1997)).

Presented in Figure 35.10 is the result of a measurement on a rigid polyurethane (PU) foam material carried out with the guarded hot plate system (solid triangles). The tests were performed with a guarded hot plate system between 40 and -160°C in steps of 20 K in a nitrogen atmosphere. Also shown is the result of a heat flow meter test on the same material (red circle) at room temperature. It can clearly be seen that the measurement results of the two techniques employed agree quite well around room temperature. The thermal conductivity of the PU-foam decreases almost linearly with decreasing temperatures around room temperature. Between -40 and -120°C , two steps were detected in the measurement results caused by the condensation of cell gases on the walls of the

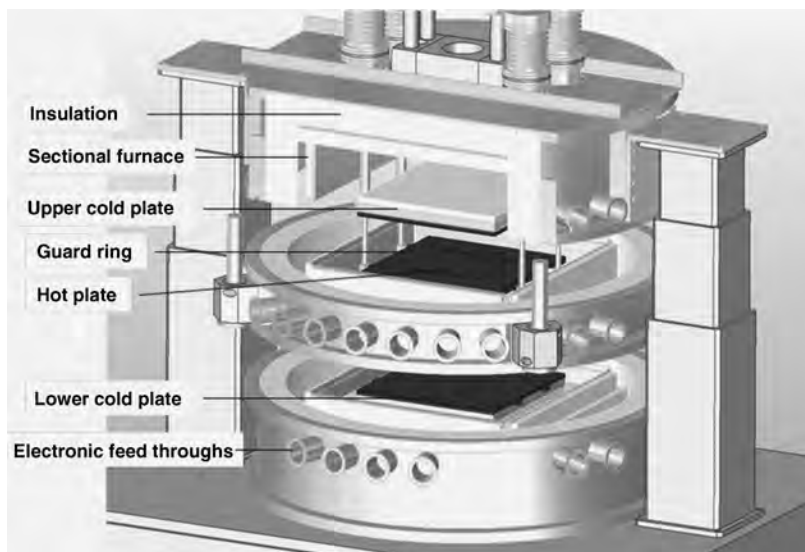


FIGURE 35.7 Detailed schematic of the hoisting device, plate stack and feed through of the GHP 456 Titan.

foam structure. This causes a significant increase in the solid contribution to the effective thermal conductivity of the foam material. Below -120°C , the linear decrease of the thermal conductivity continues.

Depicted in Figure 35.11 is a result of a mineral fiber insulation measured up to the maximum test temperature of the system. The test was carried out between room temperature and 700°C (mean sample temperature). The radiative impact on the temperature dependence of the thermal conductivity can clearly be seen here. Up to 200°C , the increase is nearly linear. At higher temperatures, the radiative contribution to the thermal conductivity leads to a higher increase rate.

There are various other methods based on the above-mentioned techniques, allowing for measurement of the thermal conductivity. Generally, the stationary methods are limited in the thermal conductivity range they can cover. Heat flow meters and guarded hot plate systems can measure large samples up to a few $\text{W}/(\text{m K})$. With some modifications, those methods can be modified for tests on smaller samples and slightly higher thermal conductivities. However, for metals, ceramics, and other more conducting materials, other methods are generally employed.

35.3 TRANSIENT METHODS FOR THE MEASUREMENT OF THE THERMAL CONDUCTIVITY

For samples above $10 \text{ W}/(\text{m K})$, transient methods are generally used to determine the thermal conductivity. These methods are based on a completely different working principle. Here, the sample is not sandwiched between the hot and cold plates. The sample is brought to a certain temperature. After the sample reaches thermal equilibrium, a certain heating power is brought inside the sample or on its surface. From the temperature response, the thermal diffusivity or thermal conductivity is determined. This means that



FIGURE 35.8 NETZSCH GHP 456 Titan.

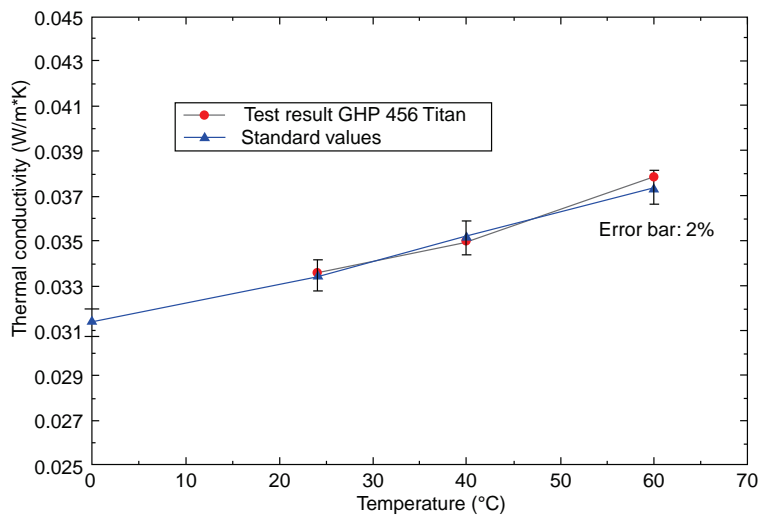


FIGURE 35.9 Thermal conductivity of NIST 1450c (test result and standard values from the certificate).

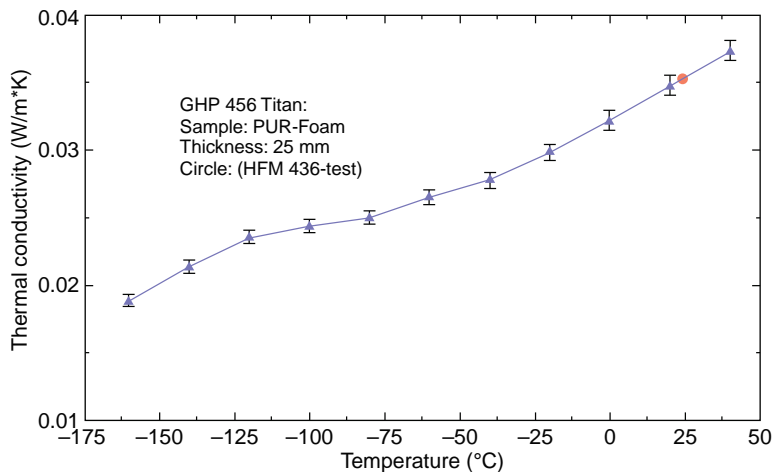


FIGURE 35.10 Thermal conductivity of a rigid polyurethane foam material.

for the analysis of the results, solutions of the transient heat transfer equation (35.4 and 35.5) must be found. A traditional way of measuring the thermal conductivity is the so called hot wire technique (Davis, 1984).

35.3.1 Hot Wire Method

In a hot wire system, a linear heat source (hot wire) is embedded in the sample. Generally, two sample bodies are prepared and a metallic wire is sandwiched between the two bodies. During the test, a constant heating power is generated inside the sample by the hot wire. The resulting temperature rise at the hot wire, or at a well-defined distance to it, is measured versus time. The experiment is illustrated in Figure 35.12:

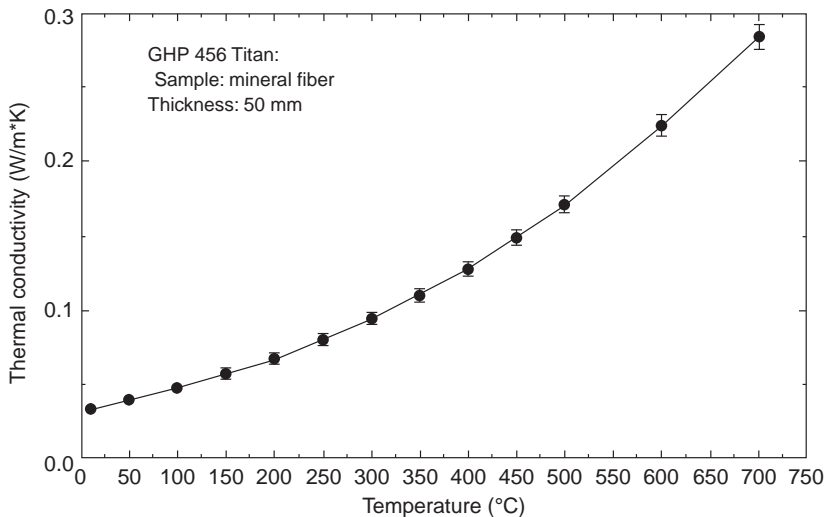


FIGURE 35.11 Thermal conductivity of a mineral fiber insulation up to 700°C.

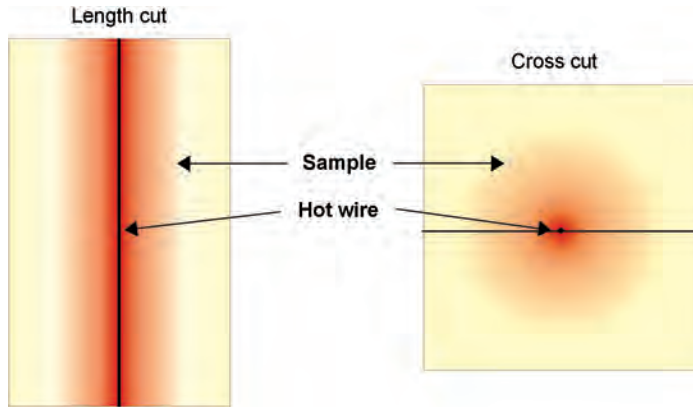


FIGURE 35.12 Heating of a sample by the hot wire embedded in the center.

Assuming the hot wire is infinitely thin and generates a homogenous heating of the sample, the mathematical treatment of the resulting data is related to a solution of the transient heat transfer equation. The solution depends on the point at which the temperature rise is measured. In a cross wire arrangement, a thermocouple is placed at the hot wire and the temperature rise is measured there. A second possibility is to place a temperature sensor parallel to the hot wire at a well-defined distance. Another possibility is to use the hot wire itself as a temperature sensor if the resistance change versus temperature is known (T(R)-technique). All three methods have advantages and drawbacks. An overview is presented in Figure 35.13.

Nowadays, the T(R)-technique is most commonly used. This technique can be adapted to different shapes of the heat source. Even sensor disks or plates with the hot wire directly embedded can be used.

The hot wire technique is often used to characterize larger sample sizes. Refractory bricks are often measured using this method. The main drawbacks of the method are that

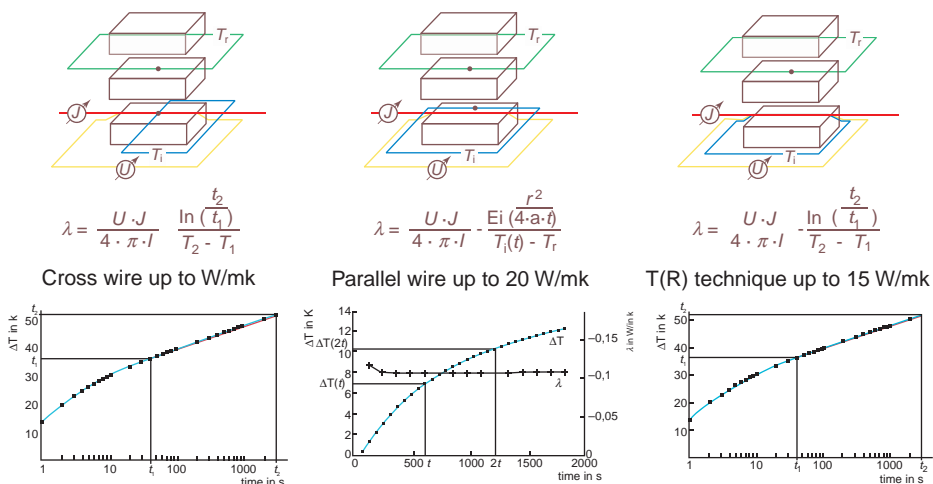


FIGURE 35.13 Different measurement methods used in a hot wire system.

tests on small samples are difficult and the preparation of the sample takes some effort and has a direct impact on the measurement accuracy. Where the sample preparation is perfect, errors of more than 15% are possible. Furthermore, tests on electrically conducting materials or materials reacting with the hot wire are difficult and often reduce the possible accuracy. Here, the hot wire must be insulated or protected to offer reasonable test results. Therefore, more and more people select other methods offering faster results and which do not require as much effort in sample preparation. One method used more and more often is the flash or laser flash method.

35.3.2 The Flash Method

For measurement of the thermal conductivity of highly conductive materials and small sample dimensions, respectively, the laser flash method is commonly used (Parker et al., 1961). This method is an absolute technique for determination of the thermal diffusivity. However, the specific heat can also be determined by laser flash. Generally, the measured raw data signals between a calibration standard and the sample are compared and the specific heat can be measured. Together with the density, which in many cases only has a weak temperature dependence, the thermal conductivity can be determined. Of course, many researchers use other techniques such as differential scanning calorimetry (DSC) (Blumm and Kaisersberger, 2001) and dilatometry (Blumm, 1999) to determine the specific heat and density change. By using those other methods, additional information such as the temperatures and enthalpy changes resulting from phase transitions and the thermal expansion or the coefficient of thermal expansion can be gained. The flash method has established itself to be the most widely used method for measurement of medium and highly conducting materials. Short measurement times, easy sample preparation, high accuracy and the simple possibility of doing temperature dependent tests are only some of the advantages of this contact-less, non-destructive measurement method. In a flash test, a plan-parallel sample is heated in a furnace to the required measuring temperature. The front face of the sample is then heated by a short light pulse (typical light pulse length < 1 ms). The light pulse is often generated by a laser system or a high power Xenon flash lamp. The heat injected on the front surface by the absorbed light spreads in the sample and leads to a temperature rise on the rear face of the sample. This temperature increase is measured with an infrared detector versus time. The schematic design of the vacuum-tight laser flash system according to Bräuer et al. (1992), used for the measurements, is shown in Figure 35.14. The head of an Nd:YAG-Laser is located in the lower part. With this laser, laser pulses with lengths between 0.3 and 1.2 ms and energies of up to 25 J can be generated. Additionally located in the lower part of the system is a sensor with which the pulse shape of the laser is detected. This allows for a correction of any pulse length influences. The sample is located on a sample carrier in the center of the tube furnace. The sample temperature is measured with thermocouples, being in contact with the sample or placed close to it. The temperature rise on the rear face of the sample is generally recorded with infrared detectors (e.g. InSb or MCT) looking directly at the rear face of the sample from above. Flash systems generally allow accuracies in the range of 3–5%.

A photo of a modern laser flash system is depicted in Figure 35.15. Here, the setup with electronics and power supply (left side) and the base unit furnace and detector housing (right side) can be seen in detail.

The typical detector signals in a flash test look pretty straightforward. At the time zero, the light pulse hits the sample surface. The resulting temperature increase on the surface

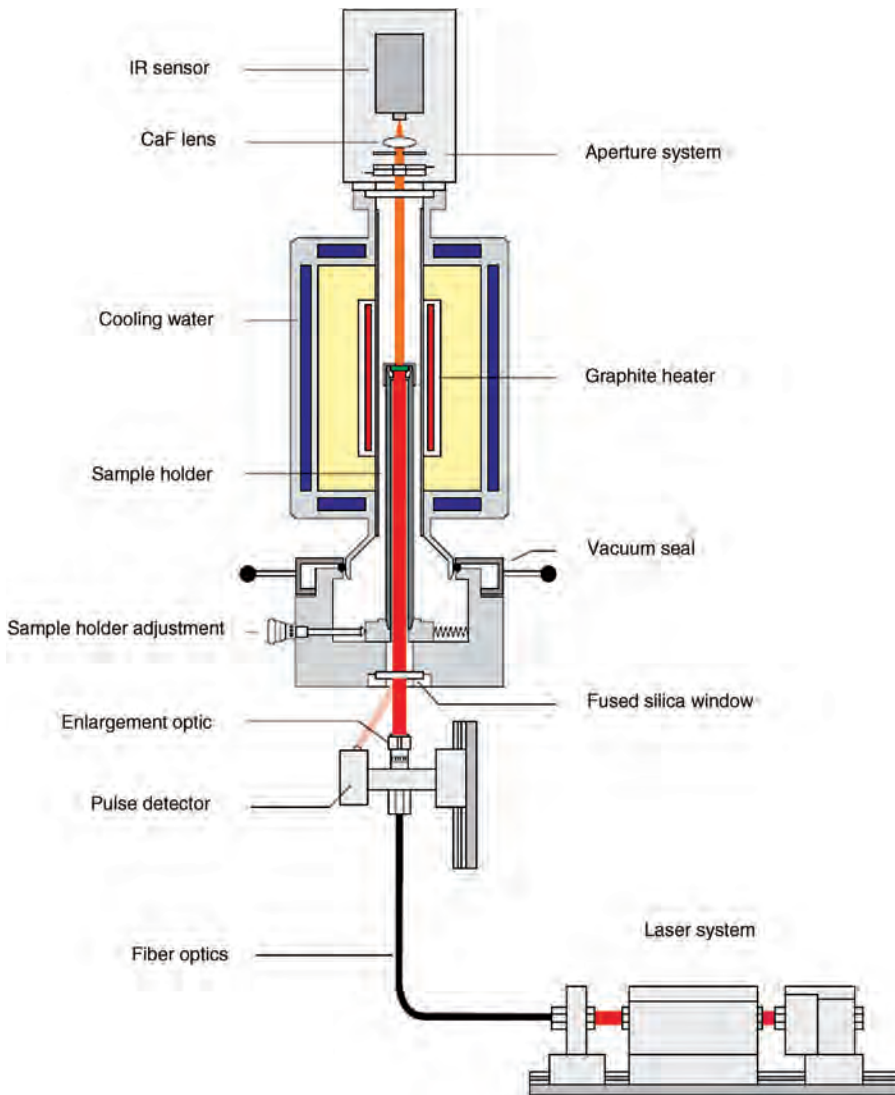


FIGURE 35.14 Schematic design of the NETZSCH LFA 427.

propagates through the sample. Thus, the temperature on the back surface starts to increase as well. It reaches a maximum before the temperature of the sample slowly goes back to ambient temperature. Three typical detector curves are depicted in Figure 35.16.

One might question how a thermophysical property can be achieved from these detector curves? There are two properties which can be determined here. The specific heat is related to the height of the measured signal. Considering the same sample dimensions and the same energy input by the light source, the specific heat is higher if the detector signal is lower and vice versa. The thermal diffusivity can be determined by analyzing the temperature change versus time. Again, a solution of the transient heat transfer equation under certain initial and boundary conditions is required. Such a solution is presented here.



FIGURE 35.15 Photo of the NETZSCH model LFA 427 laser flash system.

In the following theoretical description, a cylindrical sample body with thickness d was assumed. The origin of the coordinate system is placed in the front and center of the sample face. The x direction should run parallel to the sample surface. The basis of our analysis is the transient heat transfer equation in one dimension as mentioned before (4). It must be pointed out the T is the temperature change and not the absolute temperature for the following analysis:

$$\frac{\partial T(x, t)}{\partial t} - a \cdot \frac{\partial^2 T(x, t)}{\partial x^2} = 0 \quad (35.13)$$

Neglecting the heat loss on the surfaces of the sample body (adiabatic approach), the boundary conditions for the sample front and rear faces can be specified as follows:

$$\frac{\partial T(0, t)}{\partial x} = 0; \quad \frac{\partial T(d, t)}{\partial x} = 0 \quad (35.14)$$

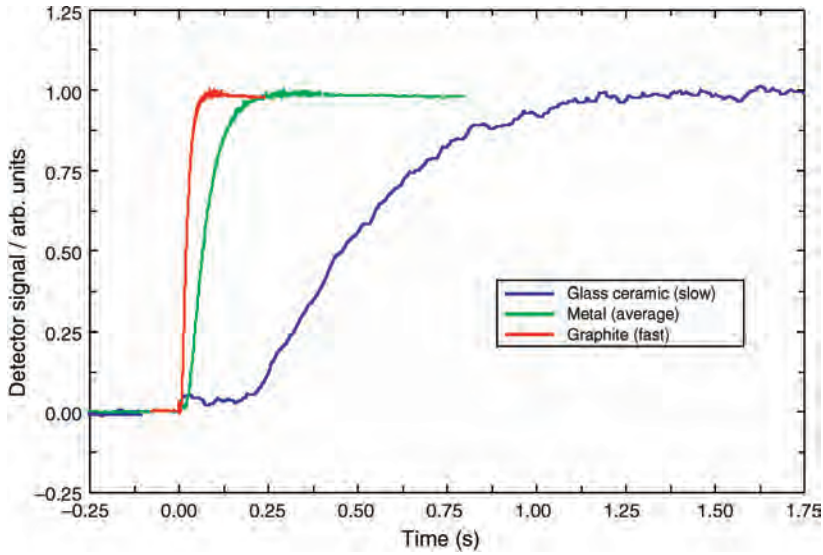


FIGURE 35.16 Typical detector curves in a flash experiment. The three curves show the results of different materials with a different thermal diffusivity/conductivity.

For initial conditions, a delta-shaped light heat pulse is used. This can be described as follows, whereby Q_0 is the energy per area brought in by the laser:

$$T(0,0) = \frac{Q_0}{\rho \cdot c_p \cdot g} \quad (35.15)$$

Here g is a thickness of a very thin layer on the front surface of the sample. The above-mentioned differential equation can be solved under the mentioned initial and boundary conditions as follows (Parker et al., 1961):

$$T(d,t) = T_\infty \left[1 + 2 \sum_{n=1}^{\infty} (-1)^n \cdot \exp\left(-\frac{n^2 \pi^2 a t}{d^2}\right) \right] \quad (35.16)$$

This solution is relatively simple, but yields limited accuracy in practice. On the one hand, a merely one-dimensional heat flow is assumed. Heat losses, which are inevitable at high temperatures and materials with a low thermal diffusivity, are not considered. Finally, a delta-shaped heat pulse can technically not be realized. Corrections are therefore essential to offer a high level of accuracy under all possible measurement conditions.

35.3.2.1 Evaluation Considering Pulse Length Influences and Heat Losses For measurements at very high temperatures or with low conducting materials, a mathematical treatment considering heat losses from the surfaces is required. An example of such a measured curve (detector signal) for a low conducting sample material at 100°C is shown in Figure 35.17. In the adiabatic case, a decrease in the measuring signal should not occur at any point of time. The measuring signal should increase as a function of time until an asymptotical critical value is reached. In the presented, real measured curve, however a

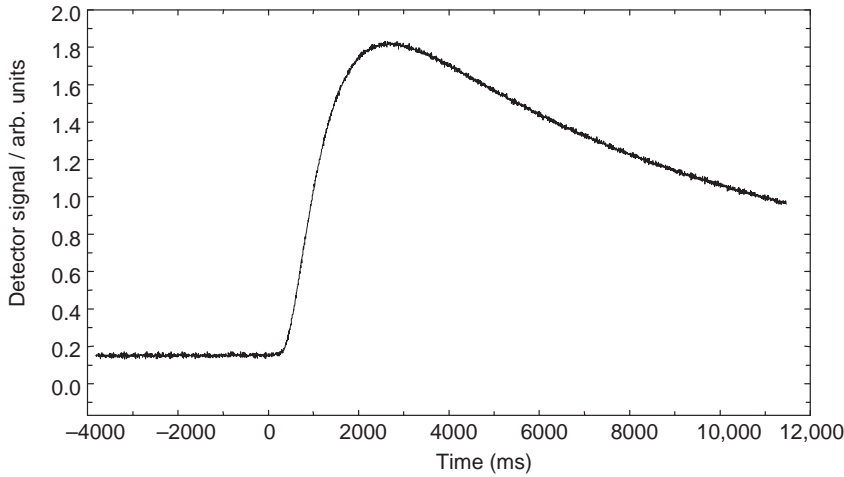


FIGURE 35.17 Laser flash measuring curves (detector signal) for a silicon carbide green body at 100°C.

clear increase for times of more than 3000 ms can be seen. The evaluation according to the adiabatic model yields values which are 13.7% too high.

A model, first introduced in 1963 by Cape and Lehman, overcomes issues like this. An evaluation considering the real geometry (cylindrical samples), that is, the transient heat transfer equation in cylinder coordinates, is recommended there:

$$\frac{\partial T(x, r, t)}{\partial t} - a \cdot \nabla^2 T(x, r, t) = \frac{1}{\rho \cdot c_p \cdot d} \cdot q(x, r, t) \quad (35.17)$$

The right part of the equation describes an original term for the heat development corresponding to the light pulse in a flash test. For the consideration of heat losses from the surfaces, radiation losses were assumed. Apart from losses from the sample's front and rear faces, radiation from the side faces of the sample cylinder may also occur. The description of the boundary conditions requires three equations (35.18) in this case:

$$\begin{aligned} \frac{\partial T(0, r, t)}{\partial x} &= \frac{1}{\lambda} \cdot 4\sigma\epsilon_x T_0^3 \cdot T(0, r, t) \\ \frac{\partial T(d, r, t)}{\partial x} &= \frac{1}{\lambda} \cdot 4\sigma\epsilon_x T_0^3 \cdot T(d, r, t) \\ \frac{\partial T(x, R_s, t)}{\partial r} &= \frac{1}{\lambda} \cdot 4\sigma\epsilon_r T_0^3 \cdot T(x, R_s, t) \end{aligned} \quad (35.18)$$

ϵ_i are the emissivities of the different surfaces, T_0 is the sample temperature prior to the light pulse injection, σ is the Boltzmann constant, and R_s is the radius of the sample body.

The same initial conditions for a delta-shaped heat pulse are used as described before:

$$T(0, 0) = \frac{Q_0}{\rho \cdot c_p \cdot g} \quad (35.19)$$

The above-mentioned differential equation can be solved under the given initial and boundary conditions according to Cape and Lehmann (1963) as follows:

$$T(r, t) = T_{\infty} \cdot \sum_{m=0}^{\infty} C_m X_m \sum_{i=0}^{\infty} D_i(r, Y_r) \cdot \exp[\omega_{im}(t - \tau)/t_c] \quad (35.20)$$

Taking the real pulse shape of the laser into consideration, this can be integrated into the following equation:

$$T(r, t) = T_{\infty} \cdot \sum_{m=0}^{\infty} C_m X_m \sum_{i=0}^{\infty} D_i(r, Y_r) \int_0^t d\tau \cdot W(\tau) \cdot \exp[\omega_{im}(t - \tau)/t_c] \quad (35.21)$$

Thereby X_m are the positive roots of the transcendent Equation (35.22).

$$(X_m^2 - Y_x^2) \cdot \tan(X_m) - 2 \cdot X_m Y_x = 0 \quad (35.22)$$

Y_x is the facial biot number which defines the contribution of the axial (facial) heat loss by means of the following equation:

$$Y_x = 4 \cdot \sigma \cdot \varepsilon_r \cdot T_0^3 \cdot \lambda^{-1} \cdot d \quad (35.23)$$

If X_m and Y_x known, $C_m X_m$ from Equation (35.21) can be calculated as follows:

$$C_m X_m = (-1)^m \cdot \frac{2a}{d} \cdot \frac{X_m^2}{X_m^2 + 2Y_x + Y_x^2} \quad (35.24)$$

The coefficient $D_i(r, Y_r)$ determines the contribution of the radial heat loss. They are described by the following equation:

$$D_i(r, Y_r) = \frac{2 \cdot Y_r}{Y_r^2 + z_i^2(Y_r)} \cdot \frac{J_0\left(z_i \cdot \frac{r}{r_0}\right)}{J_0(z_i)} \quad (35.25)$$

J_0 are the integer Bessel functions. Y_r describes the contribution of the real heat losses and is defined by the following equation:

$$Y_r = 4 \cdot \sigma \cdot \varepsilon_r \cdot T_0^3 \cdot \lambda^{-1} \cdot R_s \quad (35.26)$$

z_i are the positive roots of the following transcendent equation:

$$Y_r \cdot J_0(z_i) = z_i \cdot J_1(z_i) \quad (35.27)$$

The term ω_{im} in the exponential function is calculated with Equation (35.28), in which only known sizes occur:

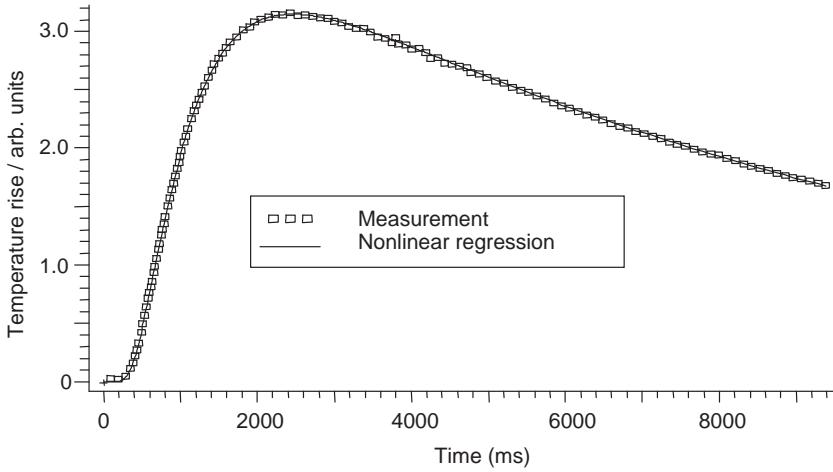


FIGURE 35.18 Detector signal during a measurement of Pyroceram 9606 at 900°C as well as the fit by means of the model curve.

$$\omega_{im} = -\left(\frac{d}{\pi}\right)^2 \cdot \left(\frac{X_m^2}{d^2} + \frac{z_i^2}{r_0^2}\right) \quad (35.28)$$

The characteristic time t_c is defined as follows:

$$t_c = \left(\frac{d}{\pi}\right)^2 \cdot \frac{1}{a} \quad (35.29)$$

$W(\tau)$ depicts the standardized, light pulse shape.

The mathematics sound complex. However, modern flash devices are computer controlled and the analysis is fully automated. Using the above mentioned mathematical equations embedded in a non-linear regression routine, the thermal diffusivity is determined within seconds and all possible influencing parameters are considered. An example for such a fitting process is presented in Figure 35.18.

35.4 TEST RESULTS ON VARIOUS ENGINEERING MATERIALS

35.4.1 Results on Various Kinds of Materials

An example of a thermal conductivity result of a modern insulation material (Styrodur C) is shown in Figure 35.19. As can be seen, the thermal conductivity increases linearly from 0.019 W/(m K) at -100°C to 0.033 W/(m K) at room temperature. Explanation of the temperature dependence of the thermal conductivity is difficult. All heat transfer mechanisms have a certain temperature dependence. The composite of all results include the typical linear dependence around room temperature. The results from the two tests are nearly the same and within a typical uncertainty of a well-designed, guarded hot plate system.

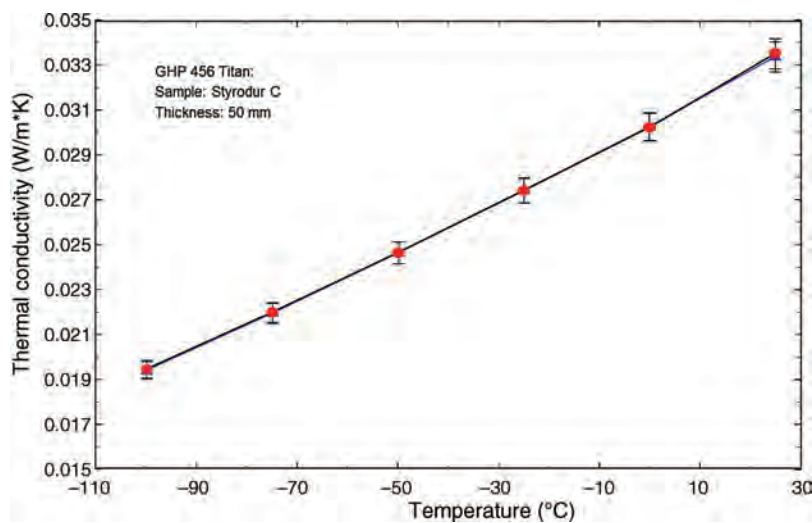


FIGURE 35.19 Thermal Conductivity of an XPS foam insulation (Styrodur C) between -100 and 25°C (two different tests).

Presented in Figure 35.20 are the results measuring purified water. The tests were carried out between 25 and 50°C using an LFA system equipped with aluminum container systems for liquids. Literature values (Incropera and DeWitt, 1996) for the density and specific heat were employed for the analysis. Additionally, literature values (different sources) for the thermal conductivity of water are shown in the plot. It can clearly be seen that the results for the thermal conductivity are in the typical range for water. Both, the

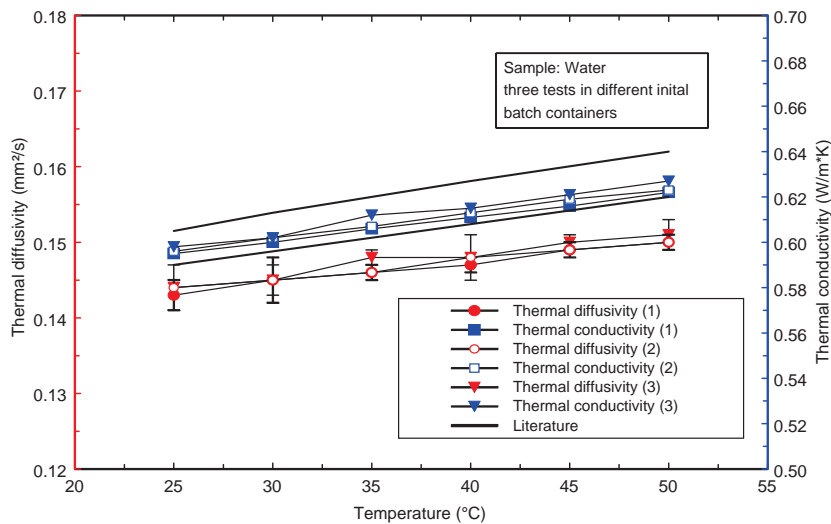


FIGURE 35.20 Thermal diffusivity and thermal conductivity of water measured in a sample holder for liquids and pastes.

thermal diffusivity and thermal conductivity increase slightly with temperature. The difference between the individual results and literature values for the thermal conductivity are generally less than $\pm 2\%$.

Presented in Figure 35.21 are the thermophysical properties (thermal diffusivity, specific heat, and thermal conductivity) of a polycarbonate material between room temperature and 300°C . The sample is an amorphous solid at room temperature. Therefore, the material was first heated above the glass transition (up to 200°C). At this temperature, the material is soft/liquid. At 200°C , the lid was pressed on the crucible to bring the sample to the required shape. After that, the sample container was cooled down to room temperature and the actual test was started. This procedure ensured a plane parallel disk between the aluminium walls of the container and good thermal contact between the container and the sample. As can be seen in Figure 35.1, the thermal diffusivity shows a nearly linear decrease between room temperature and 130°C . Between 130 and 150°C , a step to lower values was obtained. This effect is caused by the glass transition in the material. Above the glass transition, only a slight temperature was detected in the thermal diffusivity. The specific heat (measured by differential scanning calorimetry) shows a nearly linear increase versus temperature. During the glass transition, the typical step to high values was obtained. The resulting thermal conductivity shows a continuous increase versus temperature. No significant influence of the glass transition was obtained.

Presented in Figure 35.22 are the measurement results of an NR/BR rubber mixture between -125 and 75°C . The specific heat used for calculating the thermal conductivity was measured by an additional DSC test. It can clearly be seen that the thermal diffusivity decreases over the entire temperature range. Between -75 and -50°C , a step is visible in the thermal diffusivity. The specific heat increases over the entire temperature range. In addition, a step is visible in the same temperature range as the step in the thermal diffusivity results. Both steps can be explained by a glass transition in the rubber material. The thermal conductivity does not show any effects of the glass transition; a nearly linear increase was obtained over the entire temperature range.

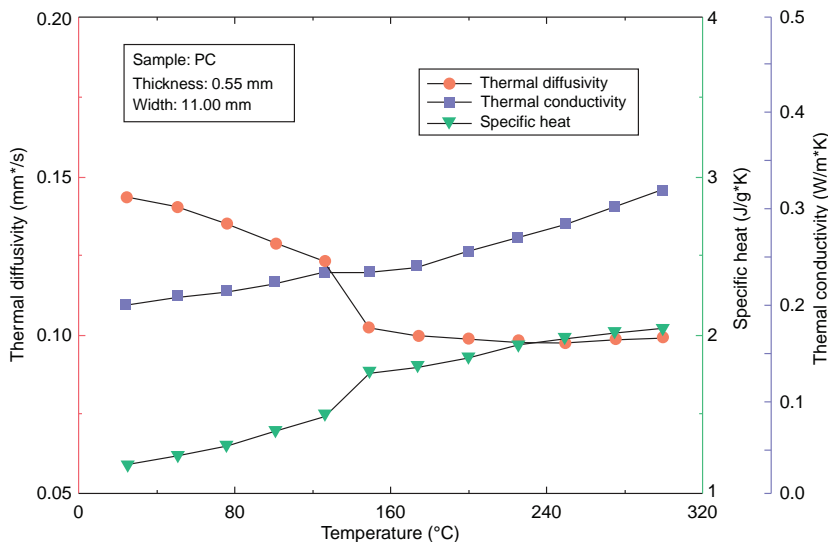


FIGURE 35.21 Thermal diffusivity, specific heat, and thermal conductivity of polycarbonate (PC).

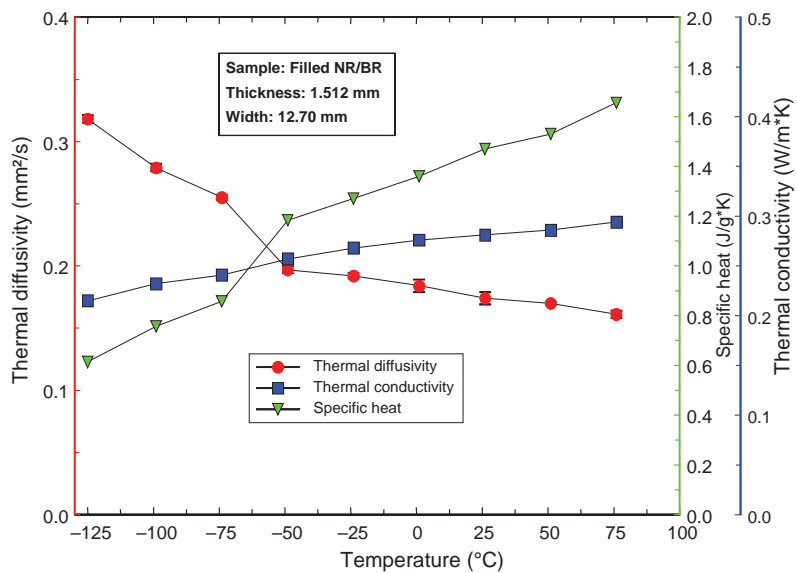


FIGURE 35.22 Specific heat, thermal diffusivity and thermal conductivity of an NR/BR rubber mixture between -125 and 75°C .

Presented in Figure 35.23 are the specific heat, thermal diffusivity and thermal conductivity of a polycrystalline graphite material. Such materials are quite interesting for several reasons: The density only shows a weak dependence versus temperature. Also the heat transfer is mainly based on the contribution of the lattice structure. The specific heat, however, shows a large increase versus temperature below room temperature. This can be

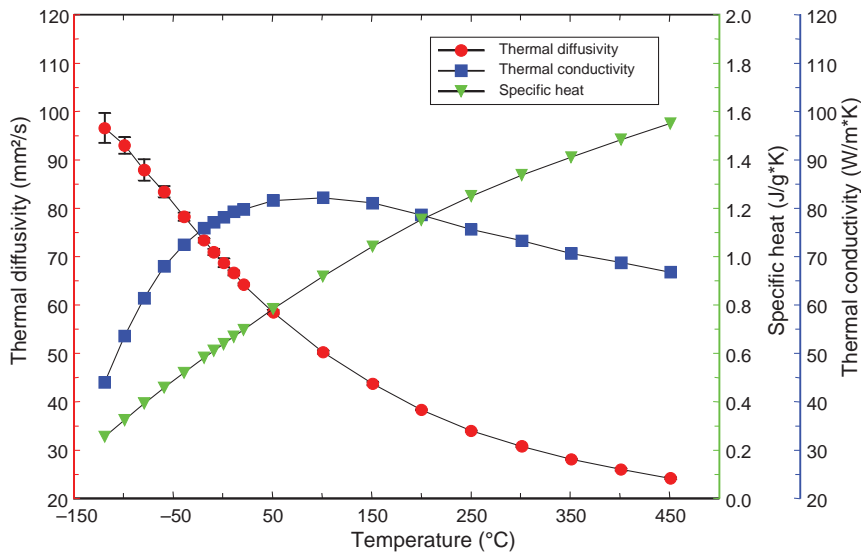


FIGURE 35.23 Specific heat, thermal diffusivity and thermal conductivity of a polycrystalline graphite material between -125 and 450°C .

explained by the Debye theory (Kittel, 2005), which describes the temperature dependence of the specific heat. At low temperatures, the specific heat should increase proportionally to T^3 (temperature in Kelvin). At high temperatures, the specific heat should converge to a constant value. The temperature range in which the material starts to differ from the T^3 depends on the Debye-temperature, which is a material-specific property. This temperature is comparably high for carbon materials (above 2000 K). Therefore, one can expect that the specific heat below and around room temperature still closely resembles the T^3 behaviour. The thermal diffusivity depends mainly on the mean free path length of the phonons. This mean free path length decreases versus temperature. Therefore, the thermal diffusivity should be proportional to T^{-1} (Kittel, 2005).

Determining the thermal conductivity according to Equation 35.6 should therefore show an increase versus temperature proportional to T^2 at low temperatures and a decrease proportional to T^{-1} at high temperatures. Considering such a temperature dependence, a maximum should occur in the thermal conductivity in the range between. For most ceramic materials, this maximum is generally in the low-temperature range. Due to the high Debye temperature of graphite, however, its maximum should be around or slightly above room temperature. As can be seen from Figure 35.23, the test results follow the expected theoretical behavior almost exactly. Only at temperatures below -50°C is the thermal diffusivity slightly below the expected temperature dependence. However, one should keep in mind that the mean free path length of the phonons cannot increase infinitely when the temperature is decreasing. The mean free path length cannot be longer than the size of a crystallite inside the sample. This, of course, results in the fact that the thermal diffusivity cannot increase infinitely but should converge to a constant value which is dependent on the structure of the material itself. This explains the low-temperature behavior of the thermal diffusivity. The thermal conductivity shows the expected maximum which was measured around 100°C .

Depicted in Figure 35.24 are the thermophysical properties of an SiC (silicon carbide) ceramic between room temperature and 1000°C . The thermal diffusivity and specific heat were determined with a laser flash system. Using the measured data, the thermal conductivity was calculated by multiplying the thermal diffusivity, specific heat and room-temperature bulk density. The thermal diffusivity values decrease over the entire temperature range. The specific heat increases, as can be expected from the Debye theory. The thermal conductivity also decreases over the entire temperature range. However, it is at a very high level (nearly $150\text{ W}/(\text{m K})$) at room temperature, which is typical for polycrystalline SiC ceramics.

Presented in Figure 35.25 are the test results (thermal diffusivity, specific heat, and thermal conductivity) measured on a magnesia–carbon refractory material with a carbon content of approx. 20%. The specific heat increases versus temperature as expected from the Debye theory. Furthermore, the results are between the typical values for pure magnesia and pure carbon (Incropera and DeWitt, 1996). The thermal diffusivity decreases versus temperature over the entire temperature range. Between 300 and 500°C , a step was detected in the results which can be explained by the decomposition of the organic binder. The decomposition of the binder yields a slightly higher porosity and carbon black between the grains of the refractory. This results in an increased thermal resistance between the grains, causing a decrease in the thermal diffusivity of the entire material. The thermal conductivity shows a slight increase up to 100°C , which can be explained by the strong increase in specific heat in this temperature range. This increase compensates for the decrease in thermal diffusivity. At higher temperatures, the changes in thermal

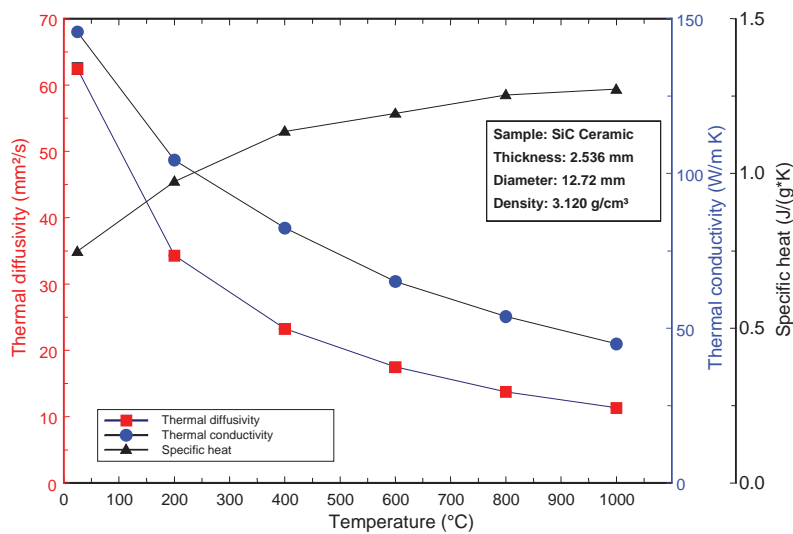


FIGURE 35.24 Specific heat, thermal diffusivity and thermal conductivity of a silicon carbide (SiC) ceramic between room temperature and 1000°C.

conductivity are mainly controlled by changes in thermal diffusivity. The slight drop in thermal conductivity above 800°C is most likely due to contact/surface reactions between the carbon and the oxide material inside the refractory. The value changes from ~25 W/(m K) at room temperature to ~16 W/(m K) at 1000°C, typical for magnesia refractories with a significant carbon content.

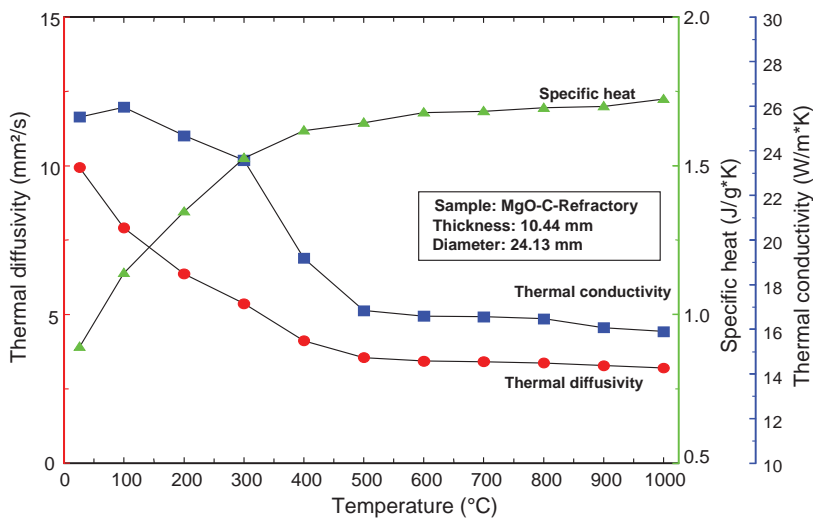


FIGURE 35.25 Thermophysical properties of a magnesia carbon refractory material (20 wt.% carbon content).

35.4.2 Full Thermophysical Properties Characterization of a Polymer and a Ceramic

35.4.2.1 Test Results on PTFE The thermophysical properties of polytetrafluoroethylene (PTFE) were determined between -130 and 150°C . For this, various test techniques were employed. For the thermal expansion measurements on the PTFE, a pushrod dilatometer was employed. The specific heat of PTFE was measured using a differential scanning calorimeter. The thermal diffusivity was measured using laser flash. Using the measurement results, the thermal conductivity was calculated according to Equation 35.1.

$$\lambda(T) = \rho(T) \cdot c_p(T) \cdot a(T) \quad (35.30)$$

Presented in Figure 35.26 are the measured linear thermal expansion and the expansivity of PTFE. The expansivity or physical coefficient of thermal expansion is defined as the rate-of-expansion divided by the original sample length:

$$\text{Expansivity} = \frac{1}{L_0} \cdot \frac{dL}{dT} \quad (35.31)$$

Starting at -130°C , the sample length increases over the entire temperature range with a slight increase in the rate-of-expansion versus temperature. Beginning at 19.2°C , two overlapped steps were detected in the thermal expansion curve. The two expansion steps are due to the solid–solid transitions (Villani, 1990). From the well-ordered to the partially ordered phase, an expansion step of approximately 0.4% was measured. For the transition from the partially ordered to the fully disordered phase above 35°C , a smaller step of approximately 0.1% was measured. Maxima in the expansivity were detected at 23.6 and 32.3°C . Those temperatures represent the points of the strongest expansion of the material during the phase transition. Above the phase transition range, the thermal expansion continuously increases with an increasing rate-of-expansion.

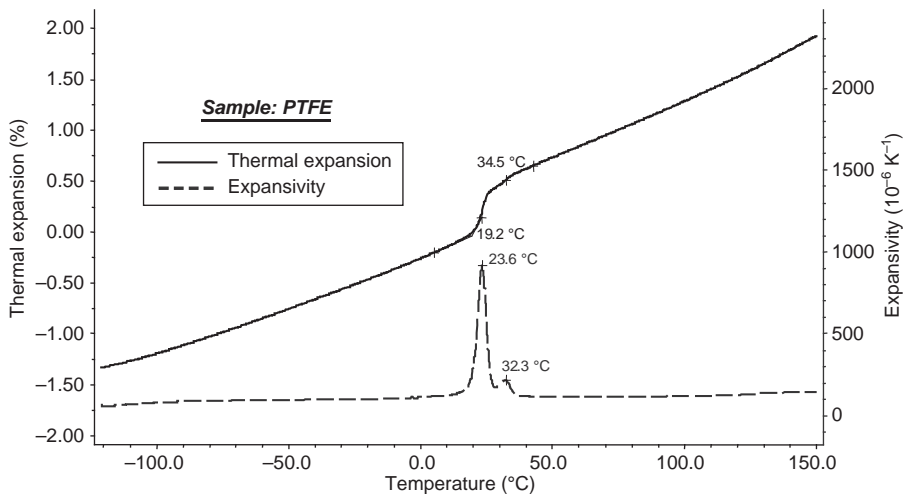


FIGURE 35.26 Thermal expansion and expansivity of the PTFE material.

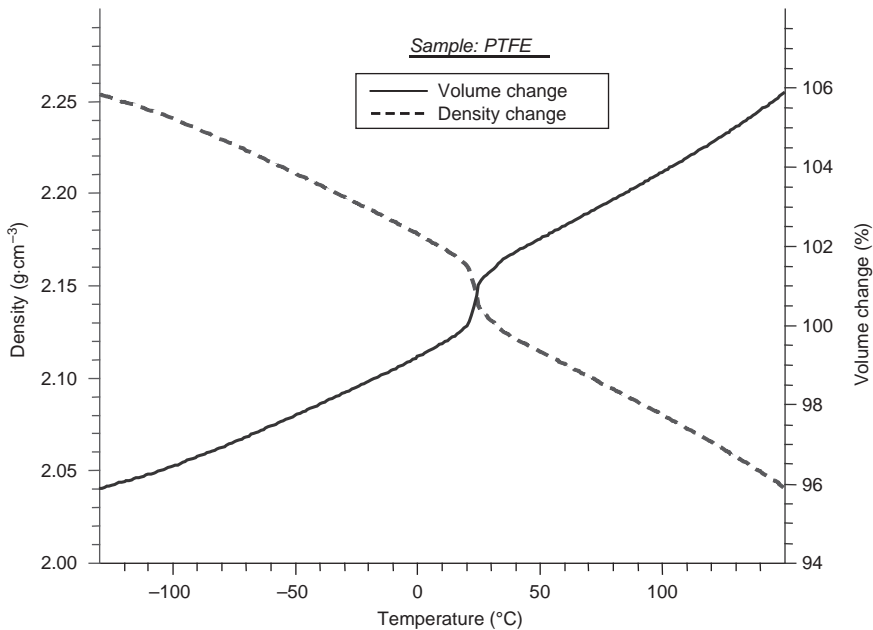


FIGURE 35.27 Volumetric expansion and density change of the PTFE material.

Presented in Figure 35.27 are the volume and density change of the PTFE material. The volume at room temperature was set to 100%. The volume increases over the entire temperature range. The steps in the measured thermal expansion of course show up in a similar way in the volume change. The density decreases from $2.254 \text{ g} \cdot \text{cm}^{-3}$ at -130°C to $2.041 \text{ g} \cdot \text{cm}^{-3}$ at 150°C . As expected, the upward steps in the volume cause a step downward in the density curve.

Depicted in Figure 35.28 is the apparent specific heat (specific heat and overlapped transition enthalpies) of the PTFE material. At low temperatures, the specific heat increases versus temperature as can be expected from the Debye-theory (Kittel, 2005). At 19.0°C (onset temperature), an endothermal peak overlaps the specific heat. The peak shows two separate maxima at 23.5 and 31.6°C , indicating that two overlapped transitions occur in this temperature range. The two transitions are due to structural changes in the material (well-ordered to partially ordered hexagonal structure and partially ordered to disordered structure). The structural changes are related with an entire enthalpy change of 7.76 J/g . Above the solid–solid phase change region, no significant phase transition was obtained in the measured specific heat until the melting range of the material was reached (240 – 360°C , peak temperature at 337.2°C). The heat of fusion was measured to be 40.56 J/g .

Presented in Figure 35.29 is the thermal diffusivity of the PTFE material versus temperature. As can be seen from the results, the thermal diffusivity decreases continuously with temperature outside the phase change region. This can be explained by solid-state physics (Kittel, 2005). PTFE is a partially crystalline material. The heat transfer inside the material is dominated by phonon conduction (by the lattice structure). For such materials, the temperature dependence of the thermal diffusivity is mainly related to changes in the mean free path length of the phonons. Due to more phonon–phonon interactions at

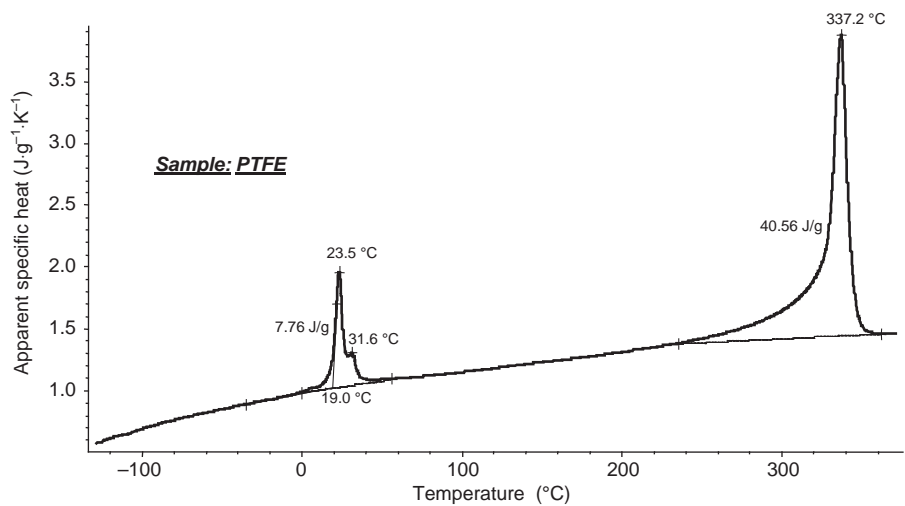


FIGURE 35.28 Apparent specific heat of the PTFE material.

higher temperatures, the mean free path length gets shorter with increasing temperature and is proportional to $1/T$. Therefore, the thermal diffusivity shows a $1/T$ dependence as well. Around room temperature, significantly lower values were measured. This effect is most likely due to the structural changes in the material mentioned earlier. During the phase change, the material structure becomes more disordered, causing stronger scattering of the phonons and therefore a shorter mean free path length. This results in a reduced thermal diffusivity. In any case, it must be mentioned that measurement of the thermo-physical properties in a phase change region is critical, as the phase transition enthalpy can have an impact on the measurement and on the resulting data. In the fully disordered

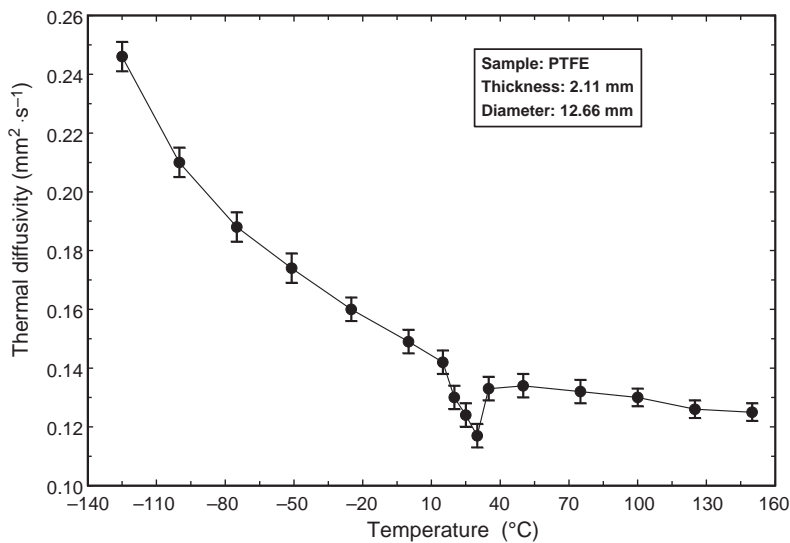


FIGURE 35.29 Thermal diffusivity of the PTFE material.

structure, the values are lower than in the ordered structure. The temperature dependence of the thermal diffusivity above 25°C is small. Only a very weak step was measured around 125°C. This is the typical temperature range of the glass transition of the amorphous content inside PTFE.

Depicted in Figure 35.30 are the thermal diffusivity, specific heat and density change of the PTFE material versus temperature. The enthalpy changes caused by phase transitions are removed from the specific heat results by a linear interpolation of the values measured before and after the phase transitions. It can clearly be seen that all measured thermophysical properties show a slight step or slope change in the temperature around room temperature. The strongest impacts of the structural changes on the measured result were detected in density and thermal diffusivity, which show significantly different behavior before, during and after the phase change.

Figure 35.31 shows the thermal conductivity of the PTFE material calculated from the measured results by multiplying the thermal diffusivity, specific heat and density. In the low-temperature range, the thermal conductivity is nearly constant. The values are around 0.32 W/(m·K). At room temperature, significantly lower values (around 0.26 W/(m·K)) were found compared to the results prior to and after the phase change. Obviously, the structural changes in the material reduce the thermal conductivity by more than 10%. Above 35°C, the thermal conductivity shows a slight increase versus temperature. Constant or slightly increasing thermal conductivities are typical for such partially crystalline polymer materials. No further transitions were measured in the temperature range above 100°C. Typical literature values for the thermal conductivity of pure PTFE materials are around 0.25 W/(m·K) (data sheet Polytetrafluoroethylene, Goodfellow, <http://www.goodfellow.com/csp/active/static/G/Polytetrafluorethylen>, 2007). The results measured in this work are slightly higher (approximately 6%) at room temperature and significantly higher at low or high temperatures. However, most of the literature data refer to tests carried out at room temperature and therefore in the phase change region.

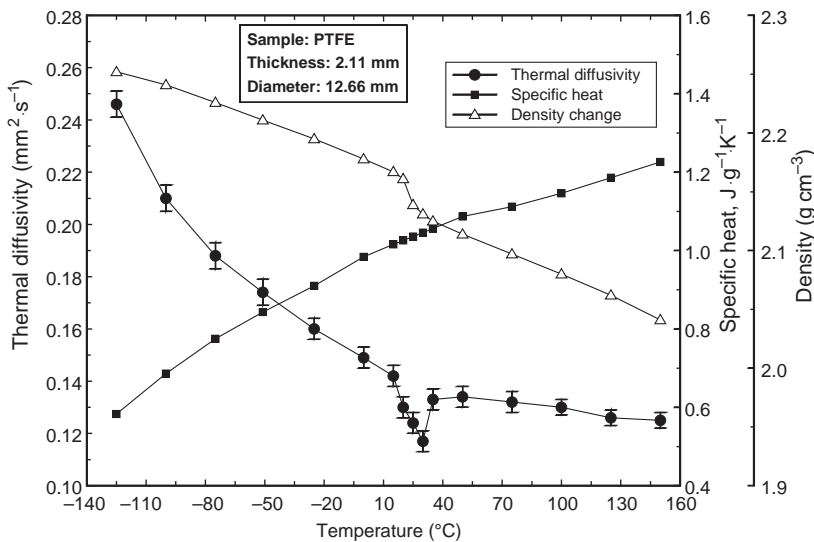


FIGURE 35.30 Thermal diffusivity, specific heat, and density change of the PTFE material.

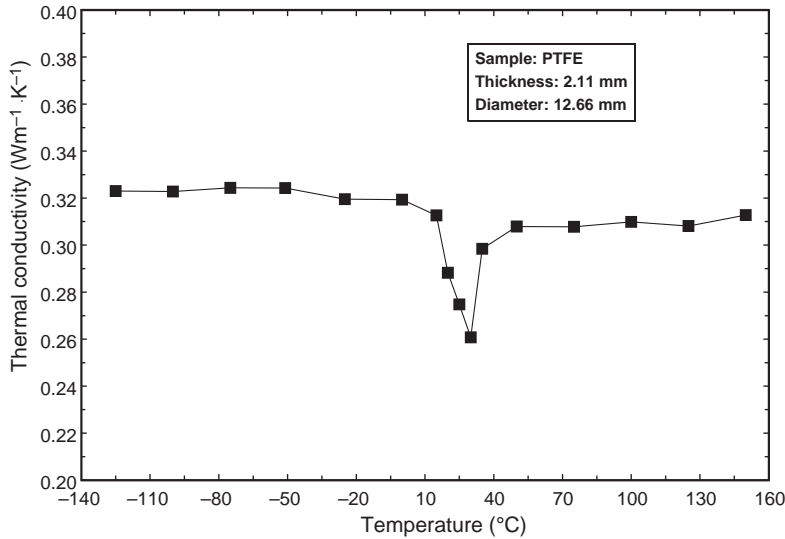


FIGURE 35.31 Thermal conductivity of the PTFE material.

35.4.2.2 Thermophysical Properties of Zirconia Before, During, and After Sintering In the production of high-tech ceramics, a powder is generally mixed with additives and a binder, pressed to a green body, and then sintered at elevated temperatures. For optimization of the production process and an optimum temperature profile during sintering, the thermophysical properties of the material must be known prior to, during and after the sintering process.

Laser flash, differential scanning calorimetry, and pushrod dilatometry were applied to an yttria stabilized zirconia ceramic prior to, during, and after the sintering process. Using the test results, the thermophysical properties required for simulation of the temperature profiles occurring in real ceramic green bodies during the production/sintering process were determined.

Presented in Figure 35.32 are the specific heat, density change, and thermal diffusivity of the zirconia green ceramic. The density change was determined from the measured length change with the dilatometer and a room-temperature bulk density of 3.089 g/cm^3 . For the density change, the mass change during the binder burnout was also taken into consideration. The influence of the binder burnout results in a slight density decrease between 300 and 500°C . During sintering, the density increases to approx. 5.8 g/cm^3 at 1600°C . The specific heat was measured on an already debinded material. Therefore, no influence of the binder burnout overlaps the measurement result. The specific heat increases over the entire temperature range. At high temperatures, nearly no temperature dependence was obtained which is in agreement with the well-known Debye theory. The thermal diffusivity was first measured on the green body with binder. The green body with binder component shows a thermal diffusivity of approx. $0.24 \text{ mm}^2/\text{s}$ at room temperature. The values slightly increase with increasing temperature. Above 300 K, a decrease in the measured data of approx. 50% can be seen. This step is caused by the binder burnout. The binder reduces the thermal contact resistances between the particles. Due to the decomposition of the binder, this contact medium is eliminated and the influence of the thermal contact resistance increases. The measurement on the debinded material between room

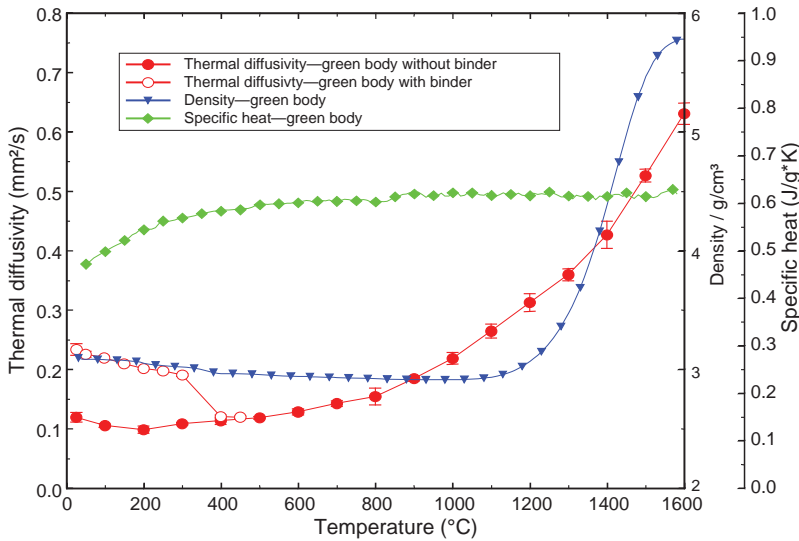


FIGURE 35.32 Density change, specific heat and thermal diffusivity of a zirconia green body between room temperature and 1600°C.

temperature and 500°C clearly shows the influences of the binder on the transport properties of the material. The test results over the entire temperature range below the binder burnout are significantly lower than for the sample with binder. Between 500 and 800°C, a continuous increase in the thermal diffusivity was evident; above 800°C, a further increase was determined. A possible reason for this rise could be the beginning of the formation of sintering necks between the powder particles. In the range of the main shrinkage (volume sintering), an increase in the temperature dependence of the thermal diffusivity can be seen.

Presented in Figure 35.33 are the measured thermophysical properties of the sintered zirconia material. As can be expected, the density shows a nearly linear decrease versus temperature. The specific heat was not influenced at all by the sintering process. The same values were obtained. The thermal diffusivity shows a decrease versus temperature. The temperature dependence is more or less proportional to $1/T$. This can be expected for ceramic materials since the thermal diffusivity only depends on the mean free path length of the phonons. Due to increased scattering phenomenon, this path length is getting smaller for higher temperatures.

Figure 35.34 depicts the thermal conductivity of zirconia for the green and sintered zirconia material. The thermal conductivity results were computed from the measured data by multiplying the density, specific heat and thermal diffusivity. It can clearly be seen that the green body with binder has a higher thermal conductivity than the green body without binder. This is mainly due to the influence of the binder on the thermal diffusivity. Comparing the data for the sintered material with that for the debinded green body, a difference of a factor of 20 was determined in the range around room temperature. During sintering, the thermal conductivity shows a large increase caused by the increase in thermal diffusivity and density.

The temperature dependence of the thermal conductivity of the sintered zirconia shows a slight decrease versus temperature. Comparing the data with literature (Schlichting

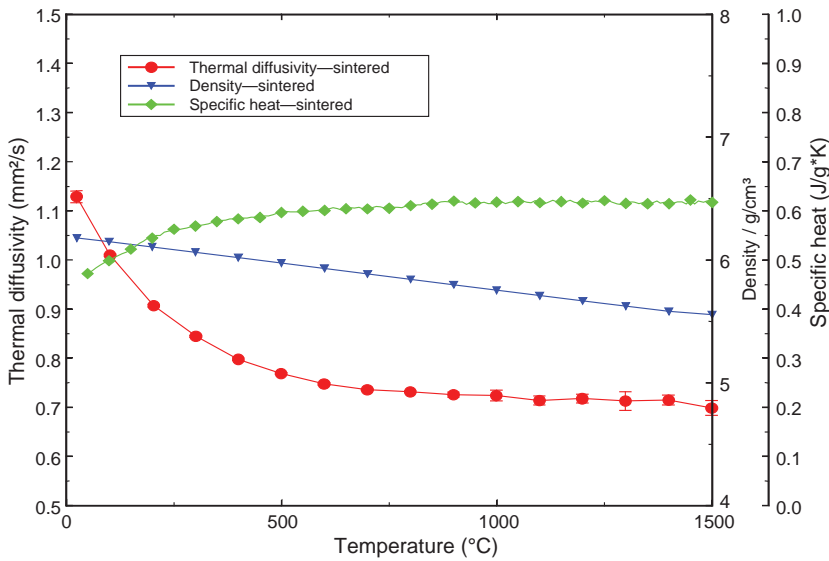


FIGURE 35.33 Density change, specific heat and thermal diffusivity of sintered zirconia between room temperature and 1500°C.

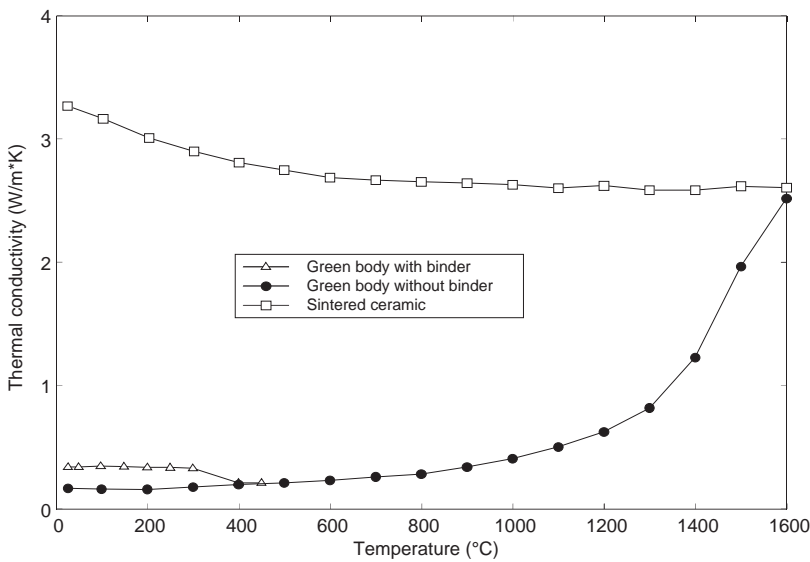


FIGURE 35.34 Thermal conductivity of a zirconia green body (with and without binder) and the sintered ceramic between room temperature and 1600°C.

et al., 2001), a good agreement was found. This source indicates room temperature values of approx. 3.2 W/(m K) for zirconia stabilized with 5.2% yttrium oxide. Up to 1273 K , these values decrease to approx. 2.5 W/(m K) .

With the help of the measured data, simulation of the temperature profile in a zirconia part was made by means of Finite-Element-Calculations (ANSYSTM). For the calculations, the measurement data of the debinded zirconia green body during heating was used. For the part geometry, an infinitely expanded cylinder with an initial diameter of 30 mm was employed. The temperature program for the outer edge layer consisted of a constant heating rate of 5 K/min . The heat energy was correspondingly adjusted so that this heating remained constant over the entire simulation range between room temperature and 1600°C .

Shown in Figure 35.35 are the results for a surface temperature of 103°C . Under the assumed conditions, a temperature difference of approx. 50 K occurs between the surface and center. Such a large temperature gradient combined with the high coefficient of thermal expansion can cause cracks and structural damage and larger sintering parts. Therefore, a lower heating rate should be considered for the real production of such a ceramic material.

Simulation of the temperature distribution at a surface temperature of 1602°C is depicted in Figure 35.36. The difference between the central and surface temperature is only approx. 4 K . This can be explained by the considerably higher thermal conductivity of this almost entirely sintered ceramic. Further, it must be mentioned that the diameter of the cylinder was reduced to 24 mm ; this was taken into account in this calculation.

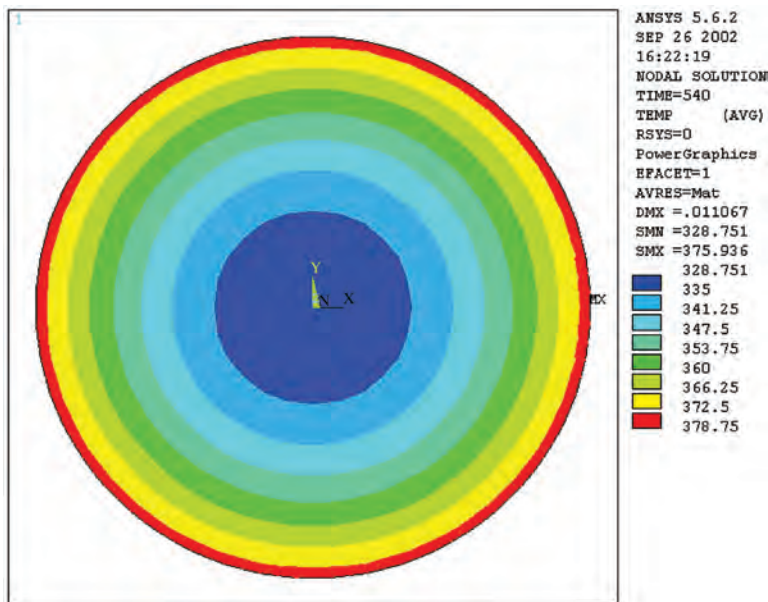


FIGURE 35.35 Calculation of the temperature distribution in a debinded zirconia green body (cylinder with a diameter of 30 mm) at a surface temperature of 103°C .

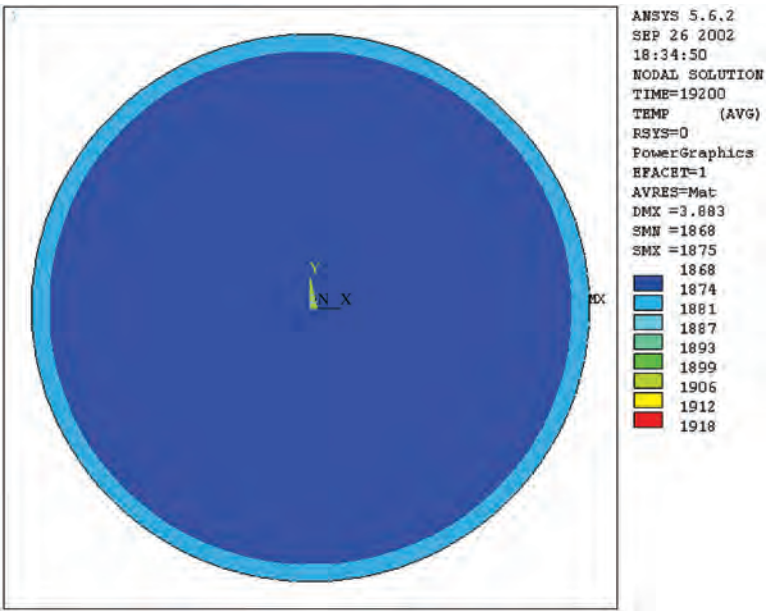


FIGURE 35.36 Calculation of the temperature distribution in a zirconia cylinder (with a diameter of 24 mm) at a surface temperature of 1602°C.

Figure 35.37 shows the results for the temperature distribution at different surface temperatures between 100 and 1600°C over the radius. The change of diameter during sintering of the component from initially 30 to approx. 24 mm can clearly be seen. The temperature differences continuously decrease with increasing temperature.

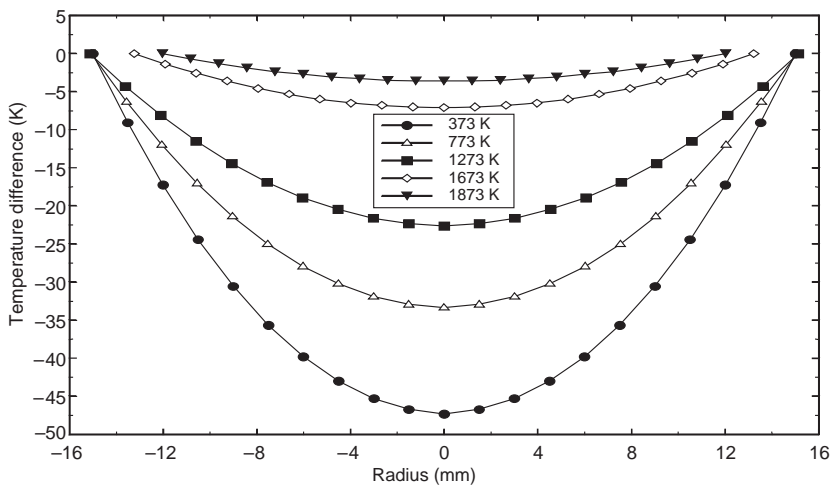


FIGURE 35.37 Simulation of temperature differences and shrinkage in a debinded zirconia green body (cylinder with an initial diameter of 30 mm) during heating at 5 K/min at different temperatures.

REFERENCES

- Incropera FP, DeWitt DP. *Introduction to Heat Transfer*. New York, Chichester, Brisbane, Toronto, Singapore: John Wiley & Sons; 1996.
- http://en.wikipedia.org/wiki/Thermal_conductivity
- Carslaw HS, Jaeger JC. *Conduction of Heat in Solids*. New York: Oxford University Press; 1959.
- Reif F. *Fundamentals of Statistical and Thermal Physics*. Berkley: McGraw-Hill; 1965.
- Salmang H, Scholze H. *Keramik, Teil 1 Allgemeine Grundlagen und wichtige Eigenschaften*. Berlin Heidelberg New York: Springer-Verlag; 1982.
- http://en.wikipedia.org/wiki/Wiedemann%E2%80%93Franz_law
- Tritt T. *Thermal Conductivity: Theory, Properties and Applications*. New York: Kluwer academic Publishers/Plenum; 2004.
- Siegel R, Howell JR. *Thermal Radiation Heat Transfer*. New York: Taylor and Francis; 2002.
- Bejan A. *Convection Heat Transfer*. New Jersey: John Wiley & Sons, Hoboken; 2004.
- Blumm J. Measuring Thermal Conductivity. *Ceramic Industry* 2002; 53.
- Gills TE. Standard Reference Material, the apparent thermal conductivity and thermal resistance of a NIST SRM 1450c high density fiberglass specimen, NIST special publication 1997.
- Davis WR. Hot-Wire Method for the Measurement of the Thermal Conductivity of Refractory Materials. in Maglic KD, Cezairliyan A, Peletsky VE, editors. *Compendium of Thermophysical Property Measurement Methods*. Vol. 1 *Survey of Measurement Techniques*, New York: London, Plenum Press; 1984, 161.
- Parker JW, Jenkins JR, Butler PC, Abbott GI. Flash method of determining Thermal Diffusivity, Heat Capacity and Thermal Conductivity. *Journal of Applied Physics* 1961;32:1679.
- Blumm J, Kaisersberger E. Accurate measurement of the transformation energetics and specific heat by DSC in the high-temperature region. *Journal of Thermal Analysis* 2001;64:385.
- Blumm J. Dilatometrie an keramischen Werkstoffen, *Das Keramiker Jahrbuch 2000*. Baden-Baden: Göller Verlag; 1999, 46.
- Bräuer H, Dusza L, Schulz B. New Laser Flash Equipment LFA 427. *Interceram* 1992;41:489.
- Villani V. A Study on the thermal behavior and Structural Characteristics of Polytetrafluoroethylene. *Thermochima Acta* 1990;162:189.
- Kittel C. *Introduction to Solid State Physics*. 8th ed. John Willey & Sons; 2005.
- data sheet Polytetrafluoroethylene, Goodfellow, <http://www.goodfellow.com/csp/active/static/G/Polytetrafluorethylen> 2007.
- Schlichting KW, Padture NP, Klemens PG. Thermal conductivity of dense and porous yttria-stabilized zirconia. *Journal of Material Science* 2001;36:3003.

36

OPTICAL METHODS FOR THE MEASUREMENT OF THERMAL CONDUCTIVITY

PRABHAKAR R. BANDARU AND MAX S. AUBAIN

- 36.1 Thermal boundary resistance may limit accuracy in contact-based thermal conductivity (κ) measurements
- 36.2 Optical measurements of κ may avoid contact-related issues
 - 36.2.1 Probing κ in lower dimensional structures (thin films and nanostructures)
 - 36.2.2 Probing the anisotropy in the κ
- 36.3 Thermoreflectance (TR)
 - 36.3.1 Principles
 - 36.3.2 Thermoreflectance to measure the κ of thin films
- 36.4 Characteristics of thermoreflectance from Si thin films—modeling and calibration
- 36.5 Experimental procedures
- 36.6 Results and discussion
 - 36.6.1 Determination of the specific heat, C , and the phonon group velocity, v_g
 - 36.6.2 Finite element modeling
 - 36.6.3 Experimental measurements of the lateral temperature variation
 - 36.6.4 Estimation of the in-plane thermal conductivity of Si thin films
- 36.7 Summary and outlook
- Acknowledgments
- References

36.1 THERMAL BOUNDARY RESISTANCE MAY LIMIT ACCURACY IN CONTACT-BASED THERMAL CONDUCTIVITY (κ) MEASUREMENTS

The issue of heat transport in materials has a long and venerable history ranging from the early investigations of Fourier. The efforts have been mostly focused on the

estimation of the thermal conductivity (κ) based on the widely accepted relationships below, which relate the heat flux (Q) to the temperature difference (ΔT), with the thermal conductance (K) as the constant of proportionality, and the heat flux through an equivalent cross-sectional area (A) to the temperature gradient ($\nabla \cdot T$) across a specified length (L).

$$Q = K\Delta T \quad (36.1a)$$

$$Q = -\kappa A \nabla \cdot T \quad (36.1b)$$

From geometrical considerations ($K = \kappa[A/L]$). The conventional method of the κ determination then use steady-state or transient/time-dependent measurements, incorporating the analysis of the heat conduction equation at any position, r , and time, t , i.e.,

$$\frac{\partial T(r, t)}{\partial t} = \nabla \cdot [\kappa(r, t) \cdot \nabla \cdot T(r, t)] \quad (36.2)$$

Equation (36.2) is solved with respect to specified boundary conditions (Carslaw and Jaeger, 1986; Ozisik, 1968), and the isotropic value of the κ ($\equiv \kappa(r, t)$) are the values that are widely reported.

However, a major issue in such methods is the use of physical contacts, which serve for heat transfer from the heater (which generates the Q) into the material that needs to be characterized. Now, it is well known that the thermal boundary resistance (R_{TBR}) (Swartz and Pohl, 1989) due to contact-material interfaces could limit the thermal conductivity/conductance (Chen, 1998). An additional thermal resistance ($= 1/K$)/unit area, R_{TBR} would then contribute to obtain an apparent value of the thermal conductivity (i.e., κ_{app}), which is really the value being measured/reported and not the true/intrinsic value (i.e., $\kappa_{\text{int}} \equiv \kappa$), that is, through

$$R_{\text{app}} \left(= \frac{L}{\kappa_{\text{app}}} \right) = R_{\text{TBR}} + R_{\text{int}} \left(= \frac{L}{\kappa_{\text{int}}} \right)$$

From a physical point of view, the heat transport across the interface of the contact and the material could be significantly affected by the thermal boundary resistance (TBR), which mainly arises due to the reflection, transmission, or absorption of heat-carrying phonons (collective lattice vibrations) (Kittel, 1996) at the interface. In addition, the surface/interface beneath the thermal energy sensors (e.g., thermocouples and thermometers) would also modulate heat flow and lead to errors/spurious readings, and so on. Consequently, much effort has been expended to provide an explanation and to account quantitatively for the influences of the TBR on thermal conductivity measurements. At the very outset, since interface conditions are mostly unpredictable (unless extreme care is taken to test with materials under specialized conditions, vacuum, cryogenic temperatures, etc. the used phenomenological models are at best approximate and serve only as a guide or provide limits for the effects of the TBR.

Such models were first quantitatively discussed in detail for solid-liquid Helium interfaces, which were posited to have an associated Kapitza resistance (Pollack, 1969).

The corresponding Kapitza resistance/TBR between materials was then found to depend on the ratio of the materials' Debye temperatures (Stoner and Maris, 1993). It was proposed that a ratio close to one could result in a closer match of the phonon wavelengths and enhanced phonon transmission, compared to when the ratio was much different than unity, cf., 0.05 for a lead–diamond interface. In this context, two basic models, that is, (1) the acoustic mismatch model (AMM) and (2) the diffuse mismatch model (DMM) have been used (Swartz and Pohl, 1989) to understand the κ in the context of conduction across an interface/boundary. In the AMM, the constituents on either side of the interface are treated as bulk solids, in a continuum approximation, with an associated acoustic impedance, $Z (= \rho v)$, where ρ is the bulk material density and v the associated acoustic velocity. The heat flux across the interface is then determined (Little, 1959) by the product of the incident number of phonons and the transmission probability, which in turn is inversely proportional to the relative contrast in the Z values of the materials constituting the interfaces, for example, the maximum transfer of heat occurs when the two materials have identical Z values. Due to the materials being modeled as continua, the detailed nature of the interface is ignored in the AMM approximation, that is, the phonon wavelength (λ_{ph}) > interface roughness. However, at increasing phonon energy/decreasing λ_{ph} , the phonon mean free path (l) could be comparable to the scale of surface roughness and imperfections, and the AMM would be inappropriate (Swartz and Pohl, 1989; Little, 1959).

Alternately, the diffuse mismatch model (DMM) implicitly considers the detailed nature of the interface through considering individual phonon traversal. The phonons impinging on the interface lose memory of their original state (direction, polarization, etc.) subsequent to scattering. The transmission of any phonon is now determined by whether there is a corresponding phonon, on the other side, of same energy to which scattering can occur, that is, by the phonon distribution or the density of states (DOS), $g(\omega)$. Although both the AMM and the DMM have been applied to simulate experimental situations, their simplified formulation considering primarily elastic scattering, generally precludes universal agreement (Swartz and Pohl, 1989) with practical situations. For example, in the case of SiO_2 and SiN_x films deposited on Si substrates, the DMM was shown to yield good agreement at low temperatures (<20 K) while it differed from experimental observations by an order of magnitude at room temperature (Lee and Cahill, 1997). Typically the DMM, which considers a greater number of phonons elastically scattering would be thought to predict a higher thermal conductance/conductivity compared to the AMM (Lee and Cahill, 1997). Indeed, the highest K , corresponding to the phonon radiation limit, is manifested through the DMM (Swartz and Pohl, 1989). However, inelastic scattering processes, for example, due to point defects (Koh et al., 2009), interfacial roughness, tunneling, excitation of surface phonons, phonon down-conversion (Kelly, 1985), and influence of electrical carriers, or strain at the interface are not considered in either model and make detailed prediction difficult.

To understand the interface scattering in more detail, molecular dynamics (MD) simulations, which serve to check the diffusive approximations inherent in the Boltzmann transport formalism underlying the Fourier relationships have been attempted (Chen et al., 2004). In MD, mutual atom–atom interactions, for example, modeled through Lennard–Jones (Kittel, 1996; Chen et al., 2004) or other inter-atomic potentials (Volz and Chen, 1999) are considered, and the heat flux computed from the product of the atomic forces and velocities.

36.2 OPTICAL MEASUREMENTS OF κ MAY AVOID CONTACT-RELATED ISSUES

Considering the unpredictable influences of the contacts, it was recognized that non-contact methodologies for measurement of κ , for example, using optical beams to provide both heating and subsequently probe temperature gradients/distributions, could provide an alternative. In this regard, a “flash method” was proposed in 1961 (Parker et al., 1961) whereby a discharge from a flash lamp (of 400 J in energy) was incident on the front surface of a sample (1–3 mm in thickness) and analysis of the time (t) variation of temperature (T) at the back surface of the sample correlated to the thermal diffusivity ($\alpha = \kappa/C\rho$), where C is the specific heat and ρ the material density—Figure 36.1. It was noted that the C can be determined through the maximum temperature rise at the back surface. Generally, such transient-based techniques have an advantage over steady-state methods in that the T - t curves can be generated, which implicitly contain the diffusivity parameter, and a single experiment can be used to deduce the values of κ , C , and α . The flash method, now using laser pulses, has since been subject to extensive development, for example, by NETZSCH[®] Instruments (N. T. Analysis, <http://www.netzsch-thermal-analysis.com/>), in the commercial sector. It is then indeed remarkable that the determination of α could be reduced to $(= 0.1388 \times l^2/t_{1/2})$ where the l is the sample thickness and $t_{1/2}$ the “time at 50% of the temperature increase measured at the rear of the test sample in seconds”—taken from the laser flash apparatus (LFA) 427 brochure at NETZSCH.

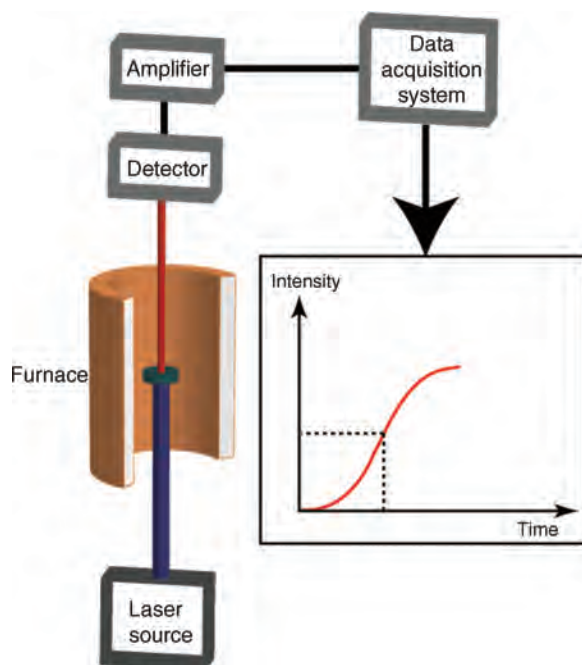


FIGURE 36.1 Schematic of laser flash apparatus (LFA) used to measure the thermal conductivity (adapted from the LFA 427 brochure at NETZSCH).

However, issues with nonuniform heating over a finite heating spot area, the accuracy of the temperature monitoring system, and the proper fitting of the T - t curves (considering possible heat losses) all have to be considered (Baba and Ono, 2001) and calibrated to yield a precise value of the thermal diffusivity. Correspondingly, monitoring the heat pulse propagation (originating, say, from an electrical current pulse) at locations on the sample away from the contacts could also be used to determine κ and C through computation of the statistical moments of the T - t curve ($\int_0^\infty \Delta T(x, t) t^n dt$ for $-\infty < n < \infty$) (Arriagada et al., 2009).

36.2.1 Probing κ in Lower Dimensional Structures (Thin Films and Nanostructures)

While measurements on bulk materials (defined by a thickness adequate to permit the diffusion of the heat from the pulse and establish a T - t profile where the temperature at the surface, T_s decays to at least a value of $(1/e) T_s$, (where e is the base of the natural logarithm) at the back surface of the sample) are enabled adequately through the laser flash-based methods, new challenges arise in the measurement of thin films. An estimate of the thickness at which the flash methods, as described previously, may not be adequate can be deduced through the diffusion length $L \sim \sqrt{\alpha t}$ [The $(1/e) T_s$ and the $L \sim \sqrt{\alpha t}$ are rules of thumb given that a common time-dependent solution for $T(r, t)$ from Equation (36.2) varies as $\exp(-cL^2/\alpha t)$, where c is a numerical constant). As an average value of α for solids is of the order of $1 \text{ cm}^2/\text{s}$ and that the time resolution can be 0.1 ns (with state-of-the-art, fast sampling oscilloscopes at 10 GS/s), the minimum thickness is of the order of 100 nm . As thin films of around this thickness are used extensively, for example, in the semiconductor industry (Ohring, 2002; Sze and Ng, 2006), and as the thin films are deposited on substrates, an alternate technique is required to measure the intrinsic κ of the film, free of the influence of the substrate.

Higher time resolution measurements would then be required and are provided through transient thermorefectance (TR)-based techniques (Paddock and Eesley, 1986). A “pump-probe” methodology is invoked, where initially a pump mode-locked Ti-sapphire laser (Saleh and Teich, 2007) with pulse width $< 1 \text{ ps}$ and pulse intensity $1\text{--}10 \text{ nJ/pulse}$ corresponding to mW powers (Capinski et al., 1999; Schmidt et al., 2008) is incident onto and heats the thin film sample. The heating modifies the refractive index, and hence changes the reflectivity of the film which to a first approximation is proportional to the temperature change, ΔT (also see Section 36.2.2). The reflectivity is subsequently probed, after a calibrated time delay, through a weaker intensity (of the order of 0.01 nJ/pulse) probe laser. The probe laser beam can either be distinct or can be referenced to the pump laser (through an attenuator) and needs to be focused onto the heated spot, quite accurately (Capinski and Maris, 1996). The spatial resolution of the incident laser spot is diffraction limited and is typically $5\text{--}10 \mu\text{m}$ and the laser beam penetration is of the order of the electromagnetic skin depth (Fox, 2001), which is $\sim 20 \text{ nm}$ (at a laser wavelength $\sim 633 \text{ nm}$ and material resistivity of $\sim 0.1 \Omega \text{ cm}$). Such a large aspect ratio, between the spot size and the skin depth, essentially ensures one-dimensional heat flow from the surface downward and could be thought to measure cross-plane thermal conductivity. The reader can consult additional references on the discussion of optical methods such as Raman spectroscopy (Christofferson et al., 2008) and charge-coupled device (CCD)-based image acquisition methods (Christofferson et al., 2008), which have also been used to map temperature distributions.

36.2.2 Probing the Anisotropy in the κ

Generally, while material thermal conductivity (κ) is traditionally defined from the ratio of the heat flux to the temperature gradient as a second-rank tensor, it is typically regarded as a scalar and an isotropic materials property. Anisotropy in the thermal conductivity is not generally considered even in thin films or even in one-dimensional nanostructures, such as nanowires (Boukai et al., 2008) or nanotubes (Bandaru, 2007) where anisotropy could be thought to be obviously present, for example, intuitively the κ values along and perpendicular to a strictly two-dimensional film or a one-dimensional wire would not be expected to be the same (a structure is effectively n -dimensional if the length in the $(3-n)^{\text{th}}$ dimension/s is less than the λ_{ph} —in the case of thermal conductivity). Limits to κ should then be considered, for example, in terms of the mean free path, l and particular phonon wavelength, λ_{ph} , especially when the relevant length scales in the film approach the carrier mean free path (Chen, 1998; Sondheimer, 1952; Baillis and Randrianalisoa, 2009; Narumanchi et al., 2004).

Although substantial progress has been made in the past few years in understanding heat conduction in lower dimensional structures (Cahill et al., 2003; Chen and Shakouri, 2002) many issues are still unresolved. There is still difficulty in accurate measurement (Cahill et al., 2002), and precise understanding of the phonon interactions with interfaces/boundaries has not yet been obtained. In case of thin films, a few attempts have been made to correlate possible anisotropies to film and measurement geometry (Borca-Tasciuc et al., 2001), where for example, a variation dependent on the direction heat flow, cross-plane (κ_{cp}) or in-plane (κ_{ip}), was noted. The discussion has then largely been framed on the basis of the classical (Kittel, 1996) interpretation of the κ , as the product of the specific heat capacity (C), the phonon group velocity (v), and the mean free path (l) that is, $\kappa \sim Cvl$. Although the assumptions of classical mechanics are inadequate to explain electron motion, due to the spin degree of freedom, classical modeling is typically considered adequate for phonons.

It then naturally follows that a reduction of any of the constituent terms would reduce the κ . As such, a reduction cannot be surmised from an elementary formulation of the Fourier heat conduction problem, and various phonon transport models have then been considered (Narumanchi et al., 2005)—ranging from the “gray” approximation where all the phonons (both acoustic and optical) are taken as equally contributing to the heat transport to “semi-gray” models (Chen and Shakouri, 2002; Chen, 2000) where mostly longitudinal acoustic (LA) phonons contribute while optical phonons are relatively stationary due to their small dispersion and group velocity. A relative partitioning of the C values between the acoustic and optical phonons also results. A further improvement on the previous considerations arises through consideration of the phonon dispersion in the first Brillouin Zone (BZ), typically considered isotropic (Dolling and Cowley, 1966), through polynomial fits to the experimental spectra (Baillis and Randrianalisoa, 2009). The resulting $\omega - k$ fits can then be utilized to determine the frequency dependent v , while the C can be estimated through taking the temperature derivative of the total energy obtained by integrating the fits over the BZ (Kittel, 1996). However, the remaining factor in the thermal conductivity expression, l is not amenable to analytical calculation/fitting as it depends on the influencing (Peierls, 1955) intimate details of the underlying material, such as anharmonic interactions, defects and impurities, and surface corrugation, which are rarely the same in any two particular material structures. The sensitivity of the l to such details is quite difficult to determine theoretically and, due to its importance for

the κ , needs to be experimentally probed. For example, when one considers the l as a vector, with decomposition into three orthogonal components in a rectangular coordinate system, l_x , l_y , and l_z , there would be three corresponding values of the κ , κ_x , κ_y , and κ_z . Limits to l could be also considered, for example, in terms of the mean free path and particular phonon wavelength, especially when the relevant length scales approach the carrier mean free path (Chen, 1998; Sondheimer, 1952; Baillis and Randrianalisoa, 2009; Narumanchi et al., 2004). Presently, there are very few reports on the experimental determination of the thermal conductivity tensor (Quelin et al., 1993), and typically the isotropic value is suggested (Che et al., 2000). It is then of interest to explore experimental methods for the determination of the κ tensor, which could then yield insight into the relevance of the isotropy assumption. It could be reasonably expected that the effects of anisotropy would be increasingly manifest in lower symmetry crystal structures as well as lower dimensional materials, such as two-dimensional thin films or one-dimensionally oriented nanowires.

We then consider exploring anisotropy in the thermal conduction through exploring a prototypical system incorporating thin films of silicon arranged on an insulating substrate. In addition to its immense technological usage, silicon can be configured in silicon-on-insulator (SOI) structures, where Si thin films can be prepared with varying thickness (from an atomic layer upward) on an underlying oxide of low thermal conductivity (~ 1.4 W/mK). The advantage of the SOI manifold is that the Si thin films can be considered to be approximately thermally independent of the underlying structure. Such a configuration permits the characterization of thin films, say < 100 nm in thickness, that would be too difficult to handle in practice. However, it should be noted that while the Si films are typically close to single-crystalline, processing could introduce random defects/impurities. In addition to enabling easier experimentation (the alternate would be to suspend the Si through complicated etching procedures), the SOI structure is commercially used in electronic devices, where it has been indicated that up to a 25% increase in speed concomitant with a 50% reduction in the consumed power is achievable (Sze, 2003). However, enhanced electronic switching performance is coupled with enhanced heat production (Pop, 2010), the dissipation of which is a major issue with SOI-based structures.

It is then of much scientific and technological interest to investigate thermal conduction issues through Si thin films in SOI structures. This involves first, the determination of both the κ_{cp} and κ_{ip} (assuming that the in-plane conduction is isotropic). At present, the majority of the experiments involve periodic surface heating—at an angular frequency— ω , though an electrical resistor, for example, as in the 3ω method (Cahill, 1990) where a metal line serves as both the heater and the thermometer, or through the use of a pump laser, for example, as in time domain thermoreflectance (TDTR) (Paddock and Eesley, 1986; Capinski et al., 1999) or frequency domain thermoreflectance (FDTR) (Schmidt et al., 2009). The thermal conductivity is deduced through an analysis of the measured signal, for example, the third harmonic of the voltage in the 3ω method or through the change in reflectance of the surface from a probe laser in TDTR, the physical basis of which is the spread of the thermal wave in the underlying films/layers. Although heat from a point source would diffuse radially, the use of metal lines of finite width or pump laser spot diameters of much greater than the thermal penetration depth (TPD) $\sim \sqrt{\kappa/\omega C}$, results in quasi-one-dimensional heat transfer and the probing of the κ_{cp} . A few corrections in the measurement of the κ_{cp} , to account for the degree of orthogonal heat flow have been described (Borca-Tasciuc et al., 2001) to understand the relative influences of the TPD, heater width, underlying layer thickness, and thermal conductivity

anisotropy, and other geometry dependent correction factors in the 3ω method. Alternatively, in TDTR/FDTR methods, the effects of heat accumulation (e.g., when the material does not reach its unperturbed state between two successive laser heating pulses) on radial heat transfer have been probed (Schmidt et al., 2008) and used to extract the in-plane and cross-plane thermal conductivity values of highly oriented pyrolytic graphite (HOPG). Generally, a greater sensitivity to lateral heat spreading would be achieved with an underlayer of low κ due to the slower diffusion of heat from the heating spot.

Although the above state-of-the-art methodologies may yield some measure of the anisotropy of the thermal conductivity, they nevertheless provide an indirect measure. It would be desirable to develop a simpler method for measuring the anisotropy in any material. One possible methodology, in the spirit of the above discussion, involves correlating the changes in the optical reflectance to the temperature (Rosencwaig et al., 1985), which would be most suitable for a noncontact method and which mitigates issues such as boundary resistances and heater capacitances (Borca-Tasciuc et al., 2001). The concomitant difficulties are the sensitivity of the measured signal-to-surface conditions and wavelength (Tessier et al., 2003). As mentioned earlier, the SOI structure would be an ideal platform to consider thermal conductivity anisotropy through TR measurements, the principles of which will now be discussed in detail.

36.3 THERMOREFLECTANCE (TR)

36.3.1 Principles

TR involves experimental methods which allow for noninteractive interrogation of the relative or absolute temperature of a surface or interface of a metallic or semiconducting material, enabled through the temperature dependence of the material's electronic energy levels and macroscopically manifested through the dielectric constant (ϵ) and change in the measured reflectance.

A little more specifically, the response incorporates the effects of temperature increase/heating through a shift of the electron/hole Fermi energy in metals or semiconductors, which in turn modulates the light absorption characteristics and reflectivity (Fox, 2001). Hence, the heating of an electrically conducting material can be measured optically and the extent of heating would be a function of the inherent thermal conductivity. The thermal modulation of reflected light from semiconductors was initially used to study band structure (Berglund, 1966), and such a method was concomitantly exploited for the study of heat transport. One advantage was that it allowed for temperature mapping of arbitrary sample surface geometries. This feature has great utility in understanding thermal transport in microelectronics and MEMS where "hot spots" can easily develop. In addition, as this optical method is noninteracting, it may allow for the thermal measurement of sensitive nanoscale devices such as nanowires or suspended structures.

The TR signal, with respect to the total reflected amplitude, is typically two to five orders of magnitude smaller and requires highly sensitive detection techniques. The response is also dependent on the wavelength of light used (Tessier et al., 2001, 2003), surface condition, and the existence of transparent overlayers (Ju and Goodson, 1998; Maize et al., 2008; Tessier et al., 2006). Using modified microscopes or other custom-built CCD-based setups, temperature resolutions of 10 mK and diffraction-limited spatial resolutions are fundamentally possible and have nearly been achieved (Mayer et al., 2006).

TR relies on the dependence of the reflectivity of a material, which for example, can be described through the change in refractive index or, equivalently, the change in dielectric constant due to temperature. The reflectivity of the surface of a material with respect to a normal, incident electromagnetic wave is defined as

$$R = \left(\frac{\tilde{n} - 1}{\tilde{n} + 1} \right)^2 \quad (36.3)$$

where \tilde{n} is the wavelength-dependent complex refractive index, $\tilde{n} = n + ik$, with the real part, n proportional to the bending of the light, while k represents the light attenuation. Differentiating and normalizing the R gives the contribution due to the change in both Δn and attenuation constant, Δk :

$$\frac{\Delta R}{R} = \frac{4(n^2 - k^2 - 1)\Delta n + 8nk\Delta k}{[(n + 1)^2 + k^2][(n - 1)^2 + k^2]} \quad (36.4)$$

The above can be expressed equivalently through the variation of the complex dielectric constant, $\tilde{\epsilon}(=\tilde{n}^2) = \epsilon_1 + i\epsilon_2$, as:

$$\frac{\Delta R}{R} = \frac{2A}{A^2 + B^2} \Delta \epsilon_1 + \frac{2B}{A^2 + B^2} \Delta \epsilon_2 \quad (36.5)$$

where $A = n(n^2 - 3k^2 - 1)$ and $B = k(3n^2 - k^2 - 1)$.

In a semiconductor, a temperature change could be considered to be manifested through the modulation/change of the fundamental energy band gap, E_g , whereby $\Delta \tilde{\epsilon} = (\partial \tilde{\epsilon} / \partial E_g) (\partial E_g / \partial T) \Delta T$. Such a relation considers that the material characteristics (e.g., thermal expansion and density) are relatively unchanged implying a low ΔT and from the above, the relative change in reflectance, $\Delta R/R$, is linearly proportional to the temperature, that is,

$$\frac{\Delta R}{R} = \frac{1}{R} \frac{dR}{dT} \Delta T = C_{TR} \Delta T$$

where C_{TR} is defined as the TR coefficient. The C_{TR} then incorporates various materials parameters and is quite sensitive, for example, to the surface (which changes the R) making published values (typically in the range 10^{-4} to 10^{-5} K^{-1} for metals and 10^{-3} for semiconductors) more useful as a guideline, rather than a reference. Consequently, the determination of C_{TR} needs calibration for an individual sample surface.

36.3.2 Thermoreflectance to Measure the κ of Thin Films

We now illustrate, through a specific case study, the practical usage of TR to measure thermal conductivity of thin films. Such a study is also aimed at a preliminary understanding of the issues of κ anisotropy that were discussed earlier, through the consideration of lateral/in-plane conductivity.

Previous measurements of the lateral thermal conductivity of Si thin films in SOI structures mainly used electrical resistance thermometry in the steady state (Asheghi et al.,

1997, 1998), where temperature variation at two distinct points (via *in situ* fabricated highly doped areas in the Si films) was measured using an electrical resistance change. Subsequently, the κ was fit using two-dimensional heat conduction models. In another effort (Liu and Asheghi, 2006), measurements on suspended Si thin film membranes, fabricated through wet etching techniques, were performed. However, this methodology requires comparison of a metal heater deposited on a suspended bridge, with and without the Si device layer present, and thermal contact issues related to the sensors could still be significant in the characterization of κ_{ip} . A scanning TR technique was also suggested (Ju and Goodson, 1998) to monitor the transient temperature distribution along the drift region of a SOI power transistor. However, the discrepancy in the trend of values in the earlier data (Asheghi et al., 1997, 1998; Ju and Goodson, 1999) with reduced values in later measurements and theoretical predictions (Liu and Asheghi, 2006) was noted.

We now consider the utility of the TR technique to monitor the κ_{ip} of Si thin films (in the submicron range) in SOI-based structures—Figure 36.2a. The experimental method

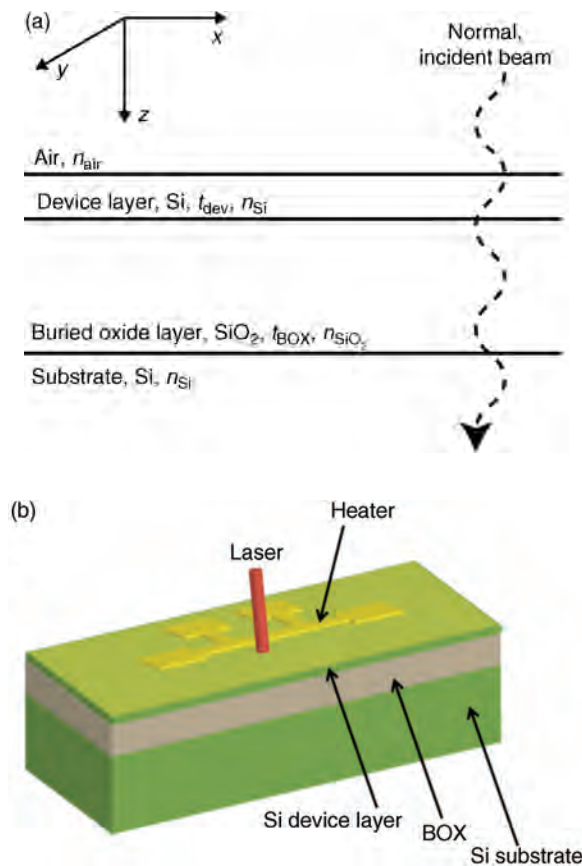


FIGURE 36.2 (a) Schematic of the SOI samples, which provides a test bed for the study of the thermal conductivity of thin Si films. The optical parameters governing the beam–material interactions, n : refractive index, and t : the material thickness, are indicated. (b) A diagram of the experimental principles in the measurement of the temperature profile of a heated SOI sample by rastering a laser beam (e.g., He–Ne laser), as a function of distance from the heater, on the sample surface.

was aimed to avoid thermal contact-related issues through the use of the TR-based temperature sensing. Although the details will be exemplified later the surface temperature gradient in the thin film induced through on-chip heating was monitored—Figure 36.2b. Based on the sensitivity of the TR to sample conditions, it was generally observed that the TR in the visible wavelength range on SOI structures must be carefully calibrated and understood considering the optical transparency and interference effects due to reflections from multiple interfaces, that is, air–Si, Si–SiO₂ (the buried oxide layer, BOX), and SiO₂–Si substrate, all of which are involved in the correlation of the measured TR intensity to the actual sample temperature. The optical response of the sample structure, as manifested through an optical characteristic matrix (OCM) (Born and Wolf, 1964), was then modeled and calculated to predict the device layer thickness at which the TR response is optimal, considering the probe wavelengths (Tessier et al., 2001, 2003) and overlayer thicknesses (Tessier et al., 2001, 2003, 2006; Maize et al., 2008; Ju and Goodson, 1998). A finite-element physics solver was used to calculate the temperature profile in the SOI structure, with the κ_{ip} of the device layer used as the only free parameter to fit the calculated surface temperature to the measured data. Finally, the determined κ_{ip} are compared to theoretical predictions (Chen, 1998; Sondheimer, 1952; Holland, 1963), and the potential application of the method to other film-on-substrate systems are discussed.

36.4 CHARACTERISTICS OF THERMOREFLECTANCE FROM Si THIN FILMS—MODELING AND CALIBRATION

The general principle of the thin film thermal conductivity measurement is that when the surface is heated in a localized region, the temperature distribution in the sample away from the heated region would be a sensitive function of the material properties, such as the κ_{ip} . In addition to surface roughness and contamination effects, any overlayer could also modify the spectral response of the TR, due to internal reflections and interference effects, as mentioned previously. Consequently, it is often preferred, in practice, to predict the C_{TR} using analytical methods, through knowledge of the temperature variation of the refractive index and thermal coefficient of expansion. For example, in the pertinent case of a Si substrate with a silicon dioxide overlayer, the C_{TR} has been derived as a function of thickness of the overlayer (Ju and Goodson, 1998). The reasonable agreement of the calculated values with experimentally measured results suggested that the former can be sufficient for estimation of the overlayer and spectral dependence of the C_{TR} .

The TR contribution of the top Si film/device layer was deconvoluted from the underlying layers/substrate through calculating the sensitivity of the TR intensity to the thickness and temperature of the device layer. The purpose was to show that the TR signal intensity could be maximized through deliberate selection of the device layer thickness through evaluating the reflection matrix of the SOI structure. An interrogation wavelength (λ) corresponding to visible light, that is, at $\lambda = 633$ nm, and geometry, as specified in Figure 36.2 was chosen. The top layer is a single crystalline Si device layer (typically with thickness < 250 nm), with an underlying 1- μ m SiO₂ (BOX) layer, supported by a 675- μ m thick single crystal substrate. From an optical standpoint, the electromagnetic skin depth (Born and Wolf, 1964), d_{Si} , of Si, at $\lambda = 633$ nm, is ~ 2.3 μ m which suggested that incident radiation would penetrate through the Si device layer. The BOX layer was

optically transparent and was modeled with only a real component to the refractive index, while the substrate is much thicker than d_{Si} and is optically opaque.

If the incident radiation modeled as an electromagnetic plane wave and represented through

$$Q_0 = \begin{bmatrix} E_0 \\ H_0 \end{bmatrix}$$

with E_0 and H_0 as the free-space amplitudes of the electric and magnetic fields, is incident upon a slab of material with refractive index \tilde{n}_s , the resultant amplitude at depth z within the slab is given by

$$Q = \begin{bmatrix} E(z) \\ H(z) \end{bmatrix}$$

where $Q_0 = M_s Q$. M_s is the optical characteristic matrix (OCM) of the layered slab, comparing the amplitude of the propagated wave to that of the initial state, and is

$$M_s(z) = \begin{bmatrix} m_s^{11}(z) & m_s^{12}(z) \\ m_s^{21}(z) & m_s^{22}(z) \end{bmatrix} = \begin{bmatrix} \cos\left(\frac{2\pi\tilde{n}_s z}{\lambda}\right) & -\frac{i}{\tilde{n}_s} \sin\left(\frac{2\pi\tilde{n}_s z}{\lambda}\right) \\ -i\tilde{n}_s \sin\left(\frac{2\pi\tilde{n}_s z}{\lambda}\right) & \cos\left(\frac{2\pi\tilde{n}_s z}{\lambda}\right) \end{bmatrix} \quad (36.6)$$

The electric and magnetic fields are taken to be of the form, $E_x = E(z)e^{i(kz - \omega t)}$ and $H_y = H(z)e^{i(kz - \omega t)}$, respectively. For SOI structures, the M_{SOI} is equal to the product of the individual OCMs of each optically active layer, that is, the top Si device layer (dev) and the SiO_2 buried oxide (BOX).

$$M_{\text{SOI}}(z) = M_{\text{dev}}(z)M_{\text{BOX}}(z) \quad (36.7)$$

It can then be shown (Born and Wolf, 1964) that the total reflectance of the SOI structure, R_{SOI} , is

$$R_{\text{SOI}} = \left| \frac{(m_{\text{SOI}}^{11} + m_{\text{SOI}}^{12}\tilde{n}_{\text{air}})\tilde{n}_{\text{sub}} - (m_{\text{SOI}}^{21} + m_{\text{SOI}}^{22}\tilde{n}_{\text{sub}})}{(m_{\text{SOI}}^{11} + m_{\text{SOI}}^{12}\tilde{n}_{\text{air}})\tilde{n}_{\text{sub}} + (m_{\text{SOI}}^{21} + m_{\text{SOI}}^{22}\tilde{n}_{\text{sub}})} \right|^2 \quad (36.8)$$

The above equations can be generally used to find the reflectance of any layered structure of arbitrary dimensions, allowing for advanced measurements such as selective probing of nonsurface layers.

A plot of R_{SOI} as a function of the device layer thickness (t_{dev}) at $\lambda = 633 \text{ nm}$ is shown in Figure 36.3a, along with obtained experimental results, and indicated the accuracy of the modeling and experimental calibration. The peaks and troughs in the R_{SOI} are due to interference effects of the incident radiation from boundaries of the device layer. The shape of the variation, that is, broad peaks and narrow troughs, were due to a large refractive index contrast ($\tilde{n}_{\text{dev}} - \tilde{n}_{\text{BOX}} \sim 2.5$) between the Si device layer and the underlying oxide, and the periodicity was determined by one quarter of the optical path length ($t_{\text{dev}}\tilde{n}_{\text{dev}}$), where interference interactions will be most influential. It can then be inferred from the $R_{\text{SOI}} - t_{\text{dev}}$ variation, that the C_{TR} (or dR/dT) could be increased at an optimal

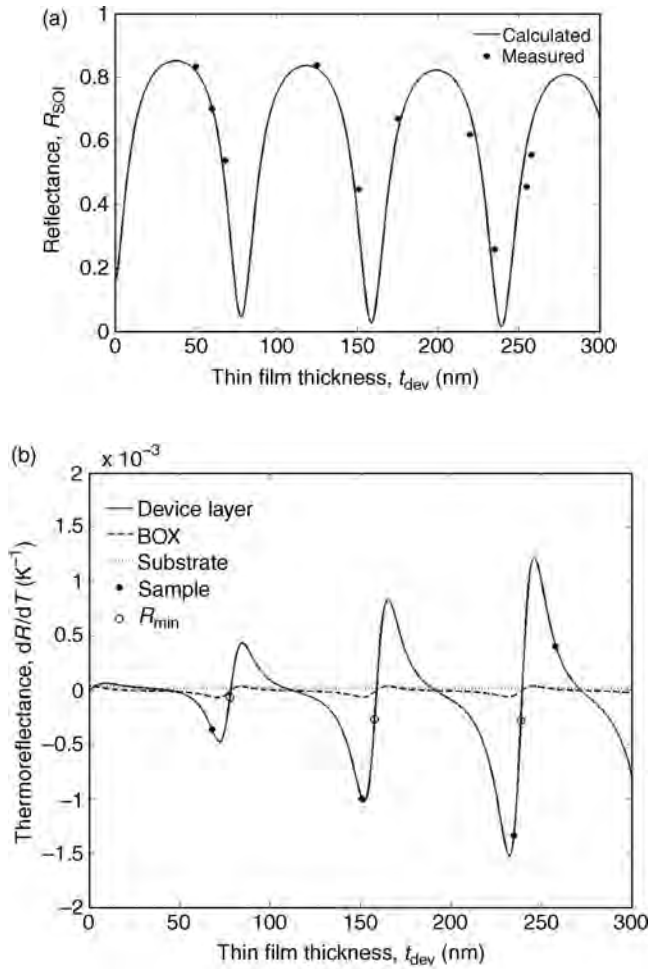


FIGURE 36.3 (a) The calculated and measured optical reflectance (R_{SOI}) of the SOI samples, with varying Si thin film/device layer thickness (t_{dev}). (b) The variation of the temperature derivative of the reflectance (dR/dT) with t_{dev} for the individual layers of the sample, including the top device layer, underlying BOX and the substrate. The values corresponding to the t_{dev} of the measured samples (Sample) and the minima (R_{min})—from Figure 36.3a, are indicated.

t_{dev} . Generally, when any layer of thickness, t , has a temperature variation ΔT , the change of the refractive index and thickness would be $\Delta \tilde{n} = d\tilde{n}/dT \Delta T$ and $\Delta t = \xi \Delta T$, respectively, where ξ is the linear thermal coefficient of expansion. For a given SOI sample with a specified device layer thickness, t_{dev} , and an initial temperature, T^0 , with reflectance $R(T^0, t_{dev})$, the change in the net reflectance, ΔR , can be written as:

$$\Delta R = \frac{dR}{dT} \Delta T = R(T^0 + \Delta T, t_{dev} + \Delta t) - R(T^0, t_{dev}) \quad (36.9)$$

For a unit rise in ΔT ($= 1$ K), and using Equations (36.7) and (36.8), a plot of ΔR versus the t_{dev} was formulated—as in Figure 36.3b, to show the individual dR_i/dT ($i = dev, BOX,$

or substrate) due to the uniform temperature rise for the i th layer. From an experimental point of view, the observed dR/dT of the SOI structure is modeled as arising from a linear superposition of the individual layers, as follows

$$\frac{dR_{\text{SOI}}}{dT} = \frac{dR_{\text{dev}}}{dT} + \frac{dR_{\text{BOX}}}{dT} + \frac{dR_{\text{han}}}{dT} \quad (36.10)$$

Such a model is necessary as the individual contributions of the layers of the SOI are not measurable and was justified on the basis of observations, through computational simulations using MATLAB[®], that the difference of dR_{SOI}/dT between (i) that found by considering and summing the individual contributions from each layer (each with a $\Delta T = 1$ K) and (ii) assuming that SOI structure as a whole has $\Delta T = 1$ K, is less than 1%.

The following were then observed from the plot: (1) there is a pronounced modulation of the dR_i/dT which could be either positive/negative, and which is proportional to the slope of R_{SOI} , (2) the dR_{dev}/dT is dominant over that of the other layers by almost two orders of magnitude ($10^{-3} \text{ K}^{-1} \text{ vis-à-vis } 10^{-5} \text{ K}^{-1}$) resulting from both the position of the device layer, and the much larger $d\bar{n}/dT$, and would be the chief contributor to the TR intensity, and that (3) there is a pronounced modulation of the dR_{dev}/dT increasing in amplitude with t_{dev} and exhibiting a maxima whenever there is a minimum in the R (cf., Figure 36.3a), and has the (4) experimental implication that there are select values of t_{dev} where the dR_{dev}/dT is maximum/minimum with a corresponding effect on the observed TR intensity. Consequently, the Si thin film thickness (t_{dev}) must be chosen carefully for maximal signal-to-noise ratio in the TR measurements, enabling more accurate determination of the thermal conductivity. Si thin films with $t_{\text{dev}} = 68, 151, 235$, and 258 nm were then chosen, corresponding to thicknesses near the maxima of the absolute dR_{dev}/dT shown in Figure 36.3b.

36.5 EXPERIMENTAL PROCEDURES

Si thin films of optimal t_{dev} thickness, following the discussion in Section 36.4, could be fabricated on SOI substrates through reactive ion etching (RIE) of the Si film. A measure of the electrical dopant density should be taken to look at possible effects on the variation of κ (in the work reported here, the density of $\sim 10^{15}/\text{cm}^3$ (Taur and Ning, 1998) was not expected to affect the thermal conductivity significantly (Asheghi et al., 2002)). Measurement and calibration of the film thickness was done through spectral reflectance, and error in the measurement was estimated to be less than 5 nm. The surface roughness introduced during the RIE was of the order of a few nanometers. A metal line was then deposited on the Si film surface to serve as an electrical resistance-based heater, for example, see Figure 36.2b. The heater line was constituted of a multilayer of Cr (10 nm)/Au (200 nm) deposited through electron beam evaporation and was typically 10-mm long and 8- μm wide, ensuring a sufficiently high aspect ratio to probe lateral conductivity. To characterize the possible current leakage from the heater into the substrate, 30 nm of SiO_2 was deposited underneath the heater line. However, measurements did not indicate any leakage effects precluding the need for such an oxide. Subsequent to the fabrication of the optimal t_{dev} films, the samples were mounted and wire-bonded to ceramic chip-carriers. Thermal epoxy was used to bond the substrate bottom to the sample holder. The overall sample configuration is schematically shown

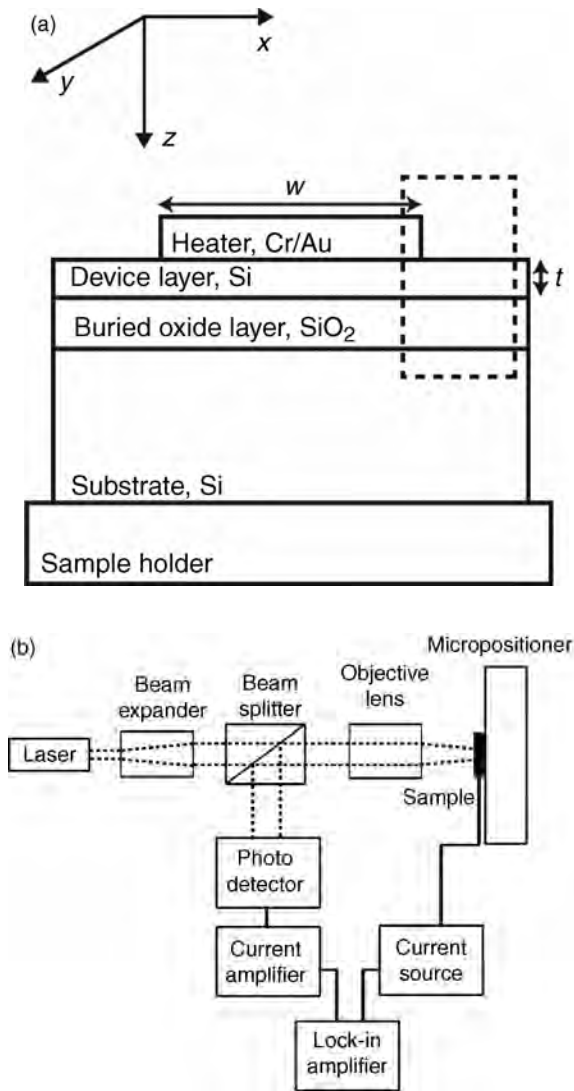


FIGURE 36.4 (a) Schematic of the cross-section of the measured samples (mounted on a chip carrier) with the heater. The lateral heat flow, in the x -direction, in the region delineated by the dotted box, is sensitive to the in-plane thermal conductivity (κ_{ip}) and the corresponding temperature variation along the surface can be estimated. (b) Schematic of the arrangement of apparatus used for the scanning TR thermometry.

in Figure 36.4a, and the setup for the measurement of the TR intensity is indicated in Figure 36.4b. It was assumed that the micropositioner (with $<0.5\text{ }\mu\text{m}$ resolution), to which the sample holder was mounted, was an adequate thermal sink and fixed the sample holder bottom to ambient temperature.

A sinusoidal current $I(f)$, at a given frequency f , was passed through the deposited metal line and induced Joule heating (with a harmonic component of $2f$, as derived from

the I^2 component of the heating). AC-based modulation techniques enable accurate lock-in-based detection of the effects of the induced heating with high signal-to-noise ratio. The thermal losses due to convection or radiation can be ignored through the use of time constants (and f values) not overlapping with the characteristic time scales associated with such loss mechanisms. Using lumped thermal analysis (Mills, 1999), it was estimated that, for the device layer thicknesses considered in our experiments, such time scales were of the order of magnitude of 0.01–0.1 s. Consequently, at a heating current frequency, f_h , of ~ 2.5 kHz, one can safely consider only the conductive heat transfer. Also, the estimated change in the surface temperature due to convective/radiative heat loss was estimated to be less than 1%, through elementary calculations using Newton's law of cooling and the Stefan–Boltzmann Law.

At the chosen f , the thermal wave propagates into the Si substrate, through the top device layer and the BOX. However, due to the orders of magnitude lower κ and much larger thickness of the oxide compared to that of the device layer, it can be assumed that there is a much larger temperature drop across the oxide. Hence, the top device layer is essentially isothermal through the thickness at any given distance from the heater. However, there seems to be significant lateral heat conduction in the BOX near the heater, which modifies the temperature profile of the device layer, and precludes the use of analytic expressions as in earlier studies (Asheghi et al., 1998). Consequently, finite element modeling was necessary to understand such variations, as will be discussed later.

A normally incident, linearly polarized He–Ne laser was then focused onto a heated sample through an objective lens. The spot size could be determined through a knife-edge technique, where the beam was scanned over the metal heater edge onto the Si film surface. Assuming a Gaussian beam profile, with intensity variation given through $I(x) = I_0 \exp(-x^2/2r^2)$, the spot radius, r , was found to be as small as $\sim 2 \mu\text{m}$. The sample was translated, with $0.2 \mu\text{m}$ resolution, rastering the focused laser spot across the surface—see Figure 36.2b. The reflected intensity was diverted through a nonpolarized beam splitter onto a diode detector connected to a lock-in amplifier synchronized to $2f$. Both the amplitude (proportional to the temperature fluctuations of the surface) and the phase (related to the sign of the C_{TR}) of the signal were recorded as a function of distance from the heater. The obtained values in the TR measurement were limited by the dark current noise of the detector, which was of the order of 4 nA.

36.6 RESULTS AND DISCUSSION

36.6.1 Determination of the Specific Heat, C , and the Phonon Group Velocity, v_g

A suitable assumed value of the density (ρ) and the specific heat (C) constituting the thermal diffusivity, $\alpha (= \kappa/\rho C)$ of the device layer was necessary before solving/modeling the transient heat conduction equation. Although the density may be assumed to be that of the bulk (CRC Handbook of Chemistry and Physics, 2004), $\rho_{\text{Si}} = 2329 \text{ kg/m}^3$, the bulk value of $C \sim 1.67 \times 10^6 \text{ J/m}^3\text{K}$ associated with the “gray” approximation of phonons (Chen, 2005), in which all phonons are treated identically, may not fully characterize their frequency dispersion. A more appropriate value of C which accounts for the phonon dispersion was derived (Chen, 1998; Sondheimer, 1952; Holland, 1963). To illustrate, it was assumed that only the acoustic phonons contribute to heat conduction, while the optical phonons do not, due to their small

group velocity, and that the specific heat capacity associated with heat conduction would be that appropriate for the former group.

The full phonon dispersion in Si, consisting of one longitudinal and two degenerate transverse modes (i.e., LA, TA and LO, TO modes) was considered. Using polynomial fitting functions to analytically describe the experimentally measured acoustic phonon dispersion (Brockhouse, 1959), a $\bar{C}_{\text{avg}} \sim 0.95 \times 10^6 \text{ J/m}^3\text{K}$, was calculated under the Debye model formulation (Kittel, 1996). An average phonon group velocity, \bar{v}_g , was also estimated by including the dispersion and normalizing as follows

$$\bar{v}_g = \frac{C_{\text{LA}} v_{g,\text{LA}} + 2C_{\text{TA}} v_{g,\text{TA}}}{\bar{C}_{\text{avg}}} \quad (36.11)$$

The resulting analysis yields $\bar{v}_g = 2274 \text{ m/s}$, which is notably smaller than the value typically considered for bulk, that is, $\sim 6000 \text{ m/s}$. Taking a κ value for Si ($\sim 140 \text{ W/mK}$), \bar{v}_g , and \bar{C}_{avg} , we obtain an average value of the mean free path, $\bar{l}_{\text{avg}} \sim 200 \text{ nm}$, implying more phonon-boundary interaction in thin films than traditional thermal analysis suggests. Thus, a reduction of κ_{ip} at film thickness near or less than \bar{l}_{avg} would be expected. Such a framework to understand the reduction of the in-plane thermal conductivity, κ_{ip} , in Si thin films (Narumanchi et al., 2004; Liu and Asheghi, 2006; Mazumder and Majumdar, 2001) and the device layer in SOI substrates (Asheghi et al., 1998; Ju and Goodson, 1999) has been previously established and agrees well with the above calculations.

36.6.2 Finite Element Modeling

As noted earlier, the SOI geometry along with the large aspect ratio of the heater (~ 40) *vis-à-vis* the device layer thickness (typically, $t_{\text{dev}} < 250 \text{ nm}$) implies a decaying temperature profile from the heater edge—in the x -direction, following Figures 36.2b and 36.4a. Consequently, a two-dimensional analysis of the heat conduction along the cross-section of the sample, perpendicular to the heater axis (i.e., the x - z plane), was pertinent. The Fourier heat conduction equation, for arbitrary geometries, can be solved through a finite element model (FEM) constructed in computational/CAD environments, for example, COMSOL Multiphysics[®] software, to determine the time-varying temperature along the surface of the device layer in the proximity of the heater and for analyzing the sensitivity of the device layer κ_{ip} . By choosing appropriate free parameters, the calculated temperature profiles could be fit to the measured TR data to elucidate the thermal properties of the device layer.

In accordance with experimental conditions, the following assumptions were used in the model: (1) the heat sink fixes the bottom substrate surface to room temperature, (2) sample surfaces exposed to air were taken to be insulating, that is, $(dT/dx)_{\text{surface}} = (dT/dz)_{\text{surface}} = 0$, (3) the power per unit area dissipated by the heater at the heater/device layer interface is equal to Joule heat generated by the heating current, and (4) the BOX and Si substrate have bulk thermal properties. The sensitivity of the model solution was examined with respect to the cross-plane thermal conductance, heating power, and frequency. It was found that the surface temperature was unperturbed by variation of the device layer κ_{cp} considered in this work, as well as from the introduction of TBRs at the device layer/BOX and BOX/substrate interfaces, with commonly accepted values (Cahill et al., 2003). The thermal epoxy between the substrate bottom and the

sample holder could introduce a resistance in series with the sink, and was estimated to be $\sim 10^{-3} \text{ m}^2\text{K/W}$. However, introduction of the boundary resistance in the model had no effect on the transient component of the surface temperature as the chosen f limits the thermal penetration depth to $\sim 75 \text{ }\mu\text{m}$ in Si, precluding the interaction of the thermal wave with the bottom of the substrate. In addition, variation of the heating power and frequency within the limits of specified machine error, $\sim 0.1\%$, did not yield a change in the surface temperature greater than the obtained precision. An example of the calculated temperature profile along the sample cross section near the peak of the heating cycle is shown in Figure 36.5a.

36.6.3 Experimental Measurements of the Lateral Temperature Variation

A typical TR scan, across the surface of the device layer ($t_{\text{dev}} = 258 \text{ nm}$) indicating the experimental data superposed on calculated temperature profiles with varying κ_{ip} is shown in Figure 36.5b. Each datum represents the maximum amplitude variation of the TR intensity as it varies in time at frequency $2f$, and whose x -coordinate coincides with the center of the probe beam. The error bars due to variation of the TR intensity are also labeled, but are too small to be visible. The magnitude of the TR intensity was scaled such that the values in the limit of $x \rightarrow \infty$, for example, when $x \sim 40 \text{ }\mu\text{m}$ in Figure 36.5b, match the calculated temperature. Although the total TR intensity was an average of the temperature profile convoluted with the spatially distributed intensity of the beam, the resulting measurement was still an accurate indication of the actual temperature at the center of the beam spot (Aubain and Bandaru, 2010). The solid curves represent equivalent, time-variant temperature amplitudes as a function of distance from the heater edge at various modeled κ_{ip} . The goodness-of-fit between the measured results and simulated curves was determined in the range of spatial values where modeled temperature profiles diverged by more than 10%, that is, at $4 \text{ }\mu\text{m} < x < 30 \text{ }\mu\text{m}$. The largest correlation coefficient between the scaled TR measurements and calculated temperature values was $R^2 = 0.9999$, corresponding to the fit with $\kappa_{\text{ip}} \sim 100 \text{ W/mK}$. It was noted that the values of κ_{ip} , matching 60 and 148 W/mK, have R^2 values of 0.9955 and 0.9978, respectively. While correlation remains high for the given calculated temperature profiles, their relative values could be used to indicate the most appropriate κ_{ip} so as to closely match experimental data.

Figure 36.5c shows the measured TR signal phase of the heat wave peak as a function of the distance from the heater, where decreasing phase indicates greater lag with respect to the heating frequency reference. The phase was a direct indication of the difference in sign of dR_{dev}/dT between samples of different thickness (cf., Figure 36.3). More specifically, comparing the signs of the dR_{dev}/dT for $t_{\text{dev}} = 258 \text{ nm}$ to other device thicknesses, that is, $t_{\text{dev}} = 68, 151, \text{ and } 235 \text{ nm}$, the former has a positive value while the latter have negative values. Consequently, the lock-in measurement of two TR signals with identical phase but opposite sign would be manifested in a π phase shift, as indicated in Figure 36.5c. It was also noted that the TR signal phase did not indicate a π phase shift as the beam was rastered from the Au heater line to the device layer, implying that dR_{dev}/dT and dR_{Au}/dT are of the same sign. As it was previously established that dR_{Au}/dT is negative at $\lambda = 633 \text{ nm}$ (Tessier et al., 2001; Raad et al., 2008; Christofferson et al., 2001), the dR_{dev}/dT of these SOI samples at these thickness must be negative as well, further supporting the OCM predictions plotted in Figure 36.3.

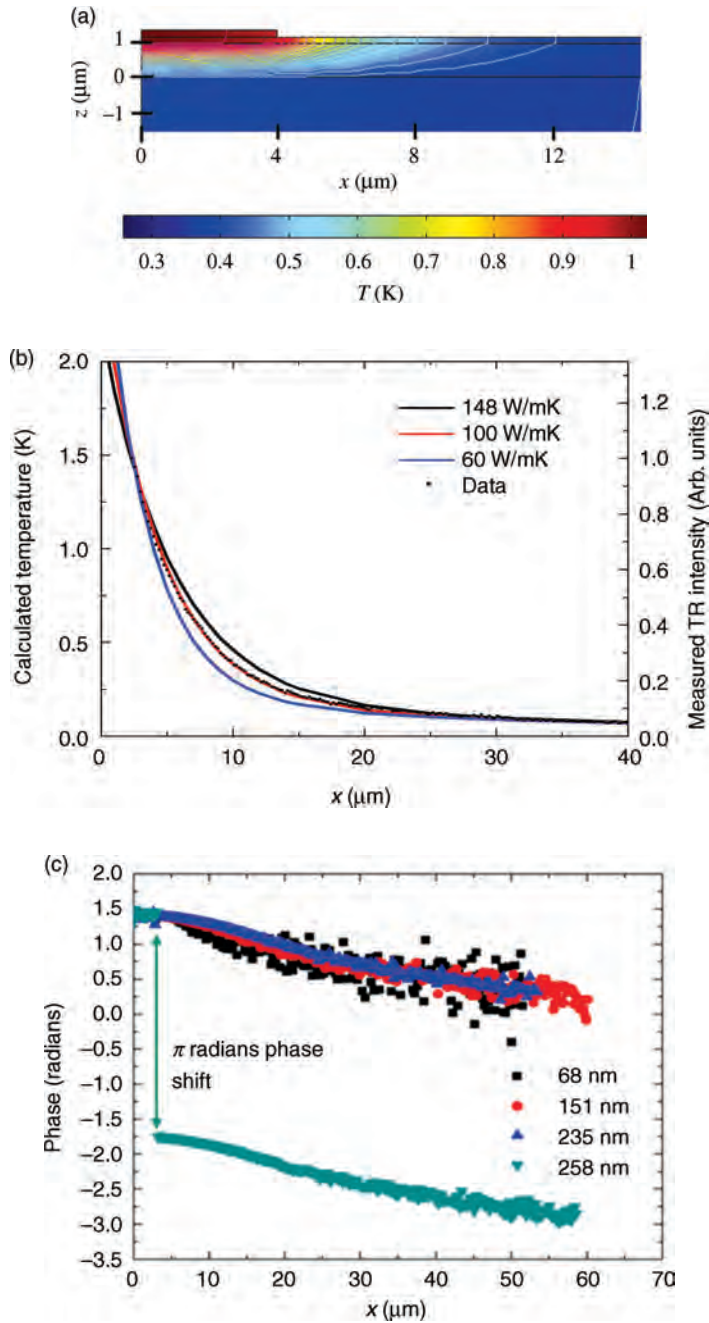


FIGURE 36.5 (a) Modeled variation, through FEM, of the temperature along the sample cross-section, indicated at the peak of a given heating cycle. (b) A typical TR scan across the surface of the device layer ($t_{\text{dev}} \sim 250$ nm) indicating the experimental data superposed on modeled temperature profiles with varying κ_{ip} . The error bars are too small to be visible. (c) The measured TR signal phase of the heat wave peak as a function of the distance from the heater is a direct indication of the difference in sign of TR intensity between samples of different thickness, cf., Figure 36.3b.

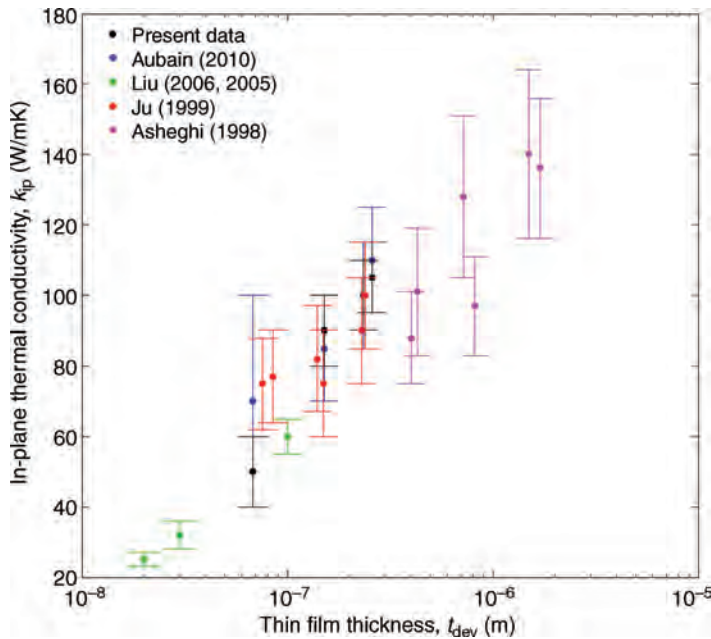


FIGURE 36.6 A comparison of the thermal conductivity values (obtained from the TR intensity fits in the present work) with those obtained previous literature (Aubain et al. 2010; Liu et al. 2005, 2006; Ju et al. 1999; Asheghi et al. 1998).

36.6.4 Estimation of the In-plane Thermal Conductivity of Si Thin Films

The principles outlined above were used to determine the values of the κ_{ip} and compare with the results obtained by other previous measurements (Ju and Goodson, 1999; Aubain and Bandaru, 2010), which have been plotted in Figure 36.6. Generally, a decreasing κ_{ip} is observed with respect to t_{dev} . The high accuracy and precision of the data obtained in this work—as in Figure 36.5b, stems from the accuracy in modeling as well as improved spatial resolution. Most notably, there seems to very good agreement between our measured κ_{ip} and those obtained through measurements with suspended Si structures fabricated from SOI substrates (Liu and Asheghi, 2006). Generally, suspended beam geometry greatly simplifies thermal analysis due to the restriction on heat conduction paths. However, fabrication of such geometries is elaborate and nontrivial involving controlled wet etching and is often not easily implemented for many thin films of interest, for example, those lacking a sacrificial intermediate layer. The comparison of the κ_{ip} previously measured through electrically resistive elements of suspended (Liu and Asheghi, 2006) and supported (Asheghi et al., 1997, 1998; Ju and Goodson, 1999) Si films show that measurements of supported samples seem to be less sensitive to the film properties (Ju and Goodson, 1999) or have high variability (Asheghi et al., 1998).

36.7 SUMMARY AND OUTLOOK

The modulation of reflected light as a function of the surface temperature of the material forms the basis of TR, and has been discussed in this work as a means to understand the

thermal conductivity (κ) of lower dimensional structures such as submicron thickness thin films. The TR response (both the magnitude and the sign) was found to be sensitive to the thickness of the thin film layer and modeled using an optical characteristic matrix formulation to choose appropriate thin film thicknesses. High aspect ratio, on-chip heating elements produced temperature profiles that were measured as a function of distance with respect to the heater edge, and the profiles were fit to a robust finite element model with the in-plane thermal conductivity (κ_{ip}) as the only free parameter. Comparison of the obtained thermal conductivity values with previous measurements confirms the validity of the technique and suggests that complete isolation of the thin film, as in earlier methodologies, from the substrate may not be required.

The advantage of the use of such optical methods to determine the κ is the possible elimination of electrical resistance-based sensing/thermometry, which suffers from thermal boundary resistance-related issues, which vary from device to device due to contact-material interfaces. The proposed TR-based methodology then provides a quick, noncontact-based optical metrology to determine the temperature profiles and thermal energy distributions.

In probing the frontiers of scientific and technological research, it is of interest to explore experimental techniques for the determination of the κ tensor, which could then yield insight, for example, into the relevance of the isotropy assumption prevalent in the literature. It would be reasonable to expect that anisotropy influences would be manifest in lower symmetry crystal structures as well as lower dimensional materials, such as two-dimensional thin films or one-dimensionally oriented nanowires and the methods described here may be useful in such investigations.

ACKNOWLEDGMENTS

We gratefully acknowledge support from the National Science Foundation (Grant ECS 0643761). The assistance of B. Fruhberger and R. Anderson at the Nano3 facility at UC, San Diego and discussions with A. Arriagada, M. Gollner, and R. Mifflin are appreciated.

REFERENCES

- Arriagada A, Yu ET, Bandaru PR. Determination of thermal parameters of one-dimensional nanostructures through a thermal transient method. *Journal of Thermal Analysis* 2009;97:1023–1026.
- Asheghi M, Kurabayashi K, Kasnavi R, Goodson KE. Thermal conduction in doped single-crystal silicon films. *Journal of Applied Physics* 2002;91:5070–5088.
- Asheghi M, Leung YK, Wong SS, Goodson KE. Phonon-boundary scattering in thin silicon layers. *Applied Physics Letters* 1997;71:1798–1800.
- Asheghi M, Touzelbaev MN, Goodson KE, Leung YK, Wong SS. Temperature-dependent thermal conductivity of single-crystal silicon layers in SOI substrates. *Journal of Heat Transfer* 1998;120:30–36.
- Aubain M, Bandaru PR. Determination of diminished thermal conductivity in silicon thin films using scanning thermoreflectance thermometry. *Applied Physics Letters* 2010;97:253102.
- Baba T, Ono A. Improvement of the laser flash method to reduce uncertainty in thermal diffusivity measurements. *Measurement Science & Technology* 2001;12:2046–2057.

- Baillis D, Randrianalisoa J. Prediction of thermal conductivity of nanostructures: Influence of phonon dispersion approximation. *International Journal of Heat and Mass Transfer* 2009;52:2516–2527.
- Bandaru PR. Electrical properties and applications of carbon nanotube structures. *Journal of Nanoscience and Nanotechnology* 2007;7:1239–1267.
- Berglund CN. Temperature-modulated optical absorption in semiconductors. *Journal of Applied Physics* 1966;37:3019.
- Borca-Tasciuc T, Kumar AR, Chen G. Data reduction in 3w method for thin-film thermal conductivity determination. *Review of Scientific Instruments* 2001;72:2139–2147.
- Born M, Wolf E. *Principles of Optics*. New York: Pergamon Press; 1964.
- Boukai AI, Bunimovich Y, Tahir-Kheli J, Yu J-K, Goddard III, WA, Heath JR. Silicon nanowires as efficient thermoelectric materials. *Nature* 2008;451:168–171.
- Brockhouse BN. Lattice vibrations in silicon and germanium. *Physical Review Letters* 1959;2:256–258.
- Cahill DG. Thermal conductivity measurement from 30 to 750 K: the 3w method. *Review of Scientific Instruments* 1990;61:802–808.
- Cahill DG, Goodson KE, Majumdar A. Thermometry and thermal transport in micro/nanoscale solid-state devices and structures. *Journal of Heat Transfer* 2002;124:223–241.
- Cahill DG, Ford WK, Goodson KE, Mahan GD, Majumdar A, Maris HJ, Merlin R, Phillpot SR. Nanoscale thermal transport. *Journal of Applied Physics* 2003;93:793–818.
- Capinski WS, Maris HJ. Improved apparatus for picosecond pump-and-probe optical measurements. *Review of Scientific Instruments* 1996;67:2720–2726.
- Capinski WS, Maris HJ, Ruf T, Cardona M, Ploog K, Katzer DS. Thermal-conductivity measurements of GaAs/AlAs superlattices using a picosecond optical pump-and-probe technique. *Physical Review B* 1999;59:8105–8113.
- Carslaw HS, Jaeger JC. *Conduction of Heat in Solids*. 2nd ed. New York: Oxford University Press; 1986.
- Che J, Cagin T, Goodard WA. Thermal conductivity of carbon nanotubes. *Nanotechnology* 2000;11:65–69.
- Chen G. Thermal conductivity and ballistic phonon transport in the cross-plane direction of superlattices. *Physical Review B* 1998;57:14958–14973.
- Chen G. *Nanoscale Energy Transport and Conversion*. New York (NY): Oxford University Press; 2005.
- Chen G. Particularities of heat conduction in nanostructures. *Journal of Nanoparticle Research* 2000;2:199–204.
- Chen G, Shakouri A. Heat transfer in nanostructures for solid-state energy conversion. *Transactions of the ASME* 2002;124:242–252.
- Chen Y, Li D, Yang J, Wu Y, Lukes JR, Majumdar A. Molecular dynamics study of the lattice thermal conductivity of Kr/Ar superlattice nanowires. *Physica B: Condensed Matter (Amsterdam)* 2004;349:270–280.
- Christofferson J, Maize K, Ezzahri Y, Shabani J, Wang X, Shakouri A. Microscale and nanoscale thermal characterization methods. *Journal of Electronic Packaging* 2008;130:041101.
- Christofferson J, Vashaee D, Shakouri A, Melese P. Real time sub-micron thermal imaging using thermoreflectance. International Mechanical Engineering Congress and Exhibition New York (NY); 2001.
- CRC Handbook of Chemistry and Physics*. 85th edition, D. Lide [Ed.], CRC Press; 2004.
- Dolling G, Cowley RA. The thermodynamic and optical properties of germanium, silicon, diamond and gallium arsenide. *Proceedings of the Physical Society* 1966;88:463–494.

- Fox M. *Optical Properties of Solids*. New York (NY): Oxford University Press; 2001.
- Holland MG. Analysis of lattice thermal conductivity. *Physical Review* 1963;132:2461–2471.
- Ju YS, Goodson KE. Short-time-scale thermal mapping of microdevices using a scanning thermoreflectance technique. *Journal of Heat Transfer* 1998;120:306–313.
- Ju YS, Goodson KE. Phonon scattering in silicon films with thickness of order 100 nm. *Applied Physics Letters* 1999;74:3005–3007.
- Kelly MJ. Acoustic phonon transmission in superlattices. *Journal of Physics C: Solid State Physics* 1985;18:5965–5973.
- Kittel C. *Introduction to Solid State Physics*. New York: John Wiley; 1996.
- Koh YK, Cao Y, Cahill DG, Jena D. Heta-transport mechanisms in superlattices. *Advanced Functional Materials* 2009;19:610–615.
- Lee S-M, Cahill DG. Heat transport in thin dielectric films. *Journal of Applied Physics* 1997;81:2590–2595.
- Lee SM, Cahill DG. Thermal conductivity of Si-Ge superlattices. *Applied Physics Letters* 1997;70:2957–2959.
- Little WA. The transport of heat between dissimilar solids at low temperatures. *Canadian Journal of Physics* 1959;37:334–349.
- Liu W, Asheghi M. Thermal conductivity measurements of ultra-thin single crystal silicon layers. *Journal of Heat Transfer* 2006;128:75–83.
- Maize K, Ezzahri Y, Wang X, Singer S, Majumdar A, Shakouri A. Measurement of thin film isotropic and anisotropic thermal conductivity using 3- ω and thermoreflectance imaging. 24th IEEE Semi-Therm Symposium 2008:185.
- Mayer PM, Luerßen D, Ram RJ, Hudgings J. Theoretical and experimental investigation of the resolution and dynamic range of CCD-based thermoreflectance imaging. *Journal of the Optical Society of America A: Optics and Image Science* 2006;25:1156–1163.
- Mazumder S, Majumdar A. Monte carlo study of phonon transport in solid thin films including dispersion and polarization. *Journal of Heat Transfer* 2001;123:749–759.
- Mills AF. *Basic Heat & Mass Transfer*. 2nd ed. Upper Saddle River: Prentice Hall; 1999.
- Narumanchi SVJ, Murthy JY, Amon CH. Submicron heat transport model in silicon accounting for phonon dispersion and polarization. *Journal of Heat Transfer* 2004;126:946–955.
- Narumanchi SVJ, Murthy JY, Amon CH. Comparison of different phonon transport models for predicting heat conduction in silicon-on-insulator transistors. *Journal of Heat Transfer* 2005;127:713–723.
- Ohring M. *Materials Science of Thin Films*. 2nd ed. San Diego: Academic Press, 2002.
- Ozisik N. *Boundary Value Problems of Heat Conduction*. New York(NY): Dover Publications Inc.; 1968.
- Paddock CA, Eesley GL. Transient thermoreflectance from thin metal films. *Journal of Applied Physics* 1986;60:285–290.
- Parker WJ, Jenkins RJ, Butler CP, Abbott GL. Flash method of determining thermal diffusivity, heat capacity, and thermal conductivity. *Journal of Applied Physics* 1961; 32:1679–1684.
- Peierls RE. *Quantum Theory of Solids*. Oxford (UK): Oxford University Press; 1955.
- Pollack GL. Kapitza resistance. *Reviews of Modern Physics* 1969;41:48–81.
- Pop E. Energy dissipation and transport in nanoscale devices. *Nano Research* 2010;3:147–169.
- Quelin X, Perrin B, Peretti P. Three-dimensional thermal-conductivity-tensor measurement of a polymer crystal by photothermal probe-beam deflection. *Physical Review B* 1993;48:3677–3682.

- Raad PE, Komaraov PL, Burzo MG. Thermal characterization of embedded electronic features by an integrated system of CCD thermography and self-adaptive numerical modeling. *Microelectronics Journal* 2008;39:1008–1015.
- Rosencwaig A, Opsal J, Smith WL, Willenborg DL. Detection of thermal waves through optical reflectance. *Applied Physics Letters* 1985;46:1013–1015.
- Saleh BEA, Teich MC. *Fundamentals of Photonics*. 2nd ed. Hoboken, NJ: John Wiley & Sons; 2007.
- Schmidt AJ, Chen X, Chen G. Pulse accumulation, radial heat conduction, and anisotropic thermal conductivity in pump-probe transient thermoreflectance. *Review of Scientific Instruments* 2008;79: 114902-1-9.
- Schmidt AJ, Cheaito R, Chiesa M. A frequency-domain thermoreflectance method for the characterization of thermal properties. *Review of Scientific Instruments* 2009;80: 094901-1-6.
- Sondheimer EH. The mean free path of electrons in metals. *Advances in Physics* 1952;1:1–42.
- Stoner RJ, Maris HJ. Kapitza conductance and heat flow between solids at temperatures from 50 to 300K. *Physical Review B: Condensed Matter* 1993;48:16373–16387.
- Swartz ET, Pohl RO. Thermal boundary resistance. *Reviews of Modern Physics* 1989;61:605–668.
- Sze SM. *Semiconductor Devices: Physics and Technology*. 2nd ed. Singapore: John Wiley & Sons, Inc.; 2003.
- Sze SM, Ng KK. *Physics of Semiconductor Devices*. Hoboken, New Jersey: Wiley-Interscience; 2006.
- Taur Y, Ning TH. *Fundamentals of Modern VLSI Devices*. New York: Cambridge University Press; 1998.
- Tessier G, Holé S, Fournier D. Quantitative thermal imaging by synchronous thermoreflectance with optimized illumination wavelengths. *Applied Physics Letters* 2001;78:2267–2268.
- Tessier G, Jerosolimski G, Holé S, Fournier D, Filloy C. Measuring and predicting the thermoreflectance sensitivity as a function of wavelength on encapsulated materials. *Review of Scientific Instruments* 2003;74:495.
- Tessier G, Jerosolimski G, Holé S, Fournier D, Filloy C. Measuring and predicting the thermoreflectance sensitivity as a function of wavelength on encapsulated materials. *Review of Scientific Instruments* 2003;74:495–499.
- Tessier G, Polignano M-L, Pavageau S, Filloy C, Fournier D, Cerutti F, Mica I. Thermoreflectance temperature imaging of integrated circuits: calibration technique and quantitative comparison with integrated sensors and simulations. *Journal of Physics D: Applied Physics* 2006;39:4159–4166.
- Volz S, Chen G. Molecular dynamics simulation of thermal conductivity of silicon nanowires. *Applied Physics Letters* 1999;75:2056–2058.

SELECTION OF METALS FOR STRUCTURAL DESIGN

MATTHEW J. DONACHIE

- 37.1 Introduction
 - 37.1.1 Metals and alloys
 - 37.1.2 Purpose
- 37.2 Common alloy systems
- 37.3 What are alloys and what affects their use?
- 37.4 What are the properties of alloys and how are alloys strengthened?
- 37.5 Manufacture of alloy articles
- 37.6 Alloy information
- 37.7 Metals at lower temperatures
 - 37.7.1 General
 - 37.7.2 Mechanical behavior
- 37.8 Metals at high temperatures
 - 37.8.1 General
 - 37.8.2 Mechanical behavior
- 37.9 Melting and casting practices
 - 37.9.1 Melting
 - 37.9.2 Casting to prepare for subsequent processing
 - 37.9.3 Casting practices for producing articles
 - 37.9.4 Casting considerations in alloy selection
- 37.10 Forging, forming, powder metallurgy, and joining of alloys
 - 37.10.1 Forging and forming
 - 37.10.2 Powder metallurgy processing
 - 37.10.3 Forging/working considerations in alloy selection
 - 37.10.4 Joining
 - 37.10.5 Considerations in joining process selection
- 37.11 Surface protection of materials
 - 37.11.1 Intrinsic corrosion resistance
 - 37.11.2 Coatings for protection
 - 37.11.3 Coating selection
- 37.12 Postservice refurbishment and repair
- 37.13 Alloy selection: a look at possibilities

- 37.13.1 General
 - 37.13.2 Impact of materials' data validity on selection
 - 37.14 Level of property data
 - 37.15 Thoughts on alloy systems
 - 37.15.1 General
 - 37.15.2 Iron
 - 37.15.3 Copper
 - 37.15.4 Aluminum
 - 37.15.5 Magnesium
 - 37.15.6 Titanium
 - 37.15.7 Nickel
 - 37.15.8 Superalloys
 - 37.16 Selected alloy information sources
 - 37.16.1 General
 - 37.16.2 Selected websites for alloy selection
- Further readings

37.1 INTRODUCTION

37.1.1 Metals and Alloys

Metals are a unique class of elements that provide special properties not available in naturally occurring materials such as wood, cement, and ceramics or in human-produced non-metallic polymeric materials. (Note: Some metals occur in elemental form in nature but most do not.) Since metals are frequently combined (alloyed) with other metallic and with some nonmetallic elements to produce alloys (metal combinations with certain desired properties), we refer all metallic materials in this chapter as “alloys” with the understanding that often pure or nearly pure metal elements may be used in design.

37.1.2 Purpose

The purpose of this chapter is to create a sufficient understanding of alloys so that selection of them for specific designs will be appropriate. The primary intent will be to cover alloy selection for structural purposes, that is, where the alloy must support its weight and, probably, the additional loads of a design. The chapter will provide sufficient information for the selector to work with designers to create successful components as well as to evaluate the capability of alloy providers and manufacturers to meet the required end uses of the design. To this end, mechanical, physical, and environmental property behavior that can influence alloy selection will be described.

There is no cook book for alloy selection. Proprietary alloys and/or proprietary/restricted processing can create conditions and properties not listed in a handbook or catalog of available alloys. Critical applications may require a selector to work with one or more manufacturers to develop an understanding of what is available and to determine what one can expect from a chosen alloy. At times, it may be necessary to develop a new or adapt different manufacturing processes for the utilization of alloys for specific applications.

37.2 COMMON ALLOY SYSTEMS

While there are many metallic elements, which are used to produce alloys, the more industrially important alloys in structural design are iron and its alloys (called ferrous alloys) or other (nonferrous) alloys based on copper, aluminum, magnesium, nickel, cobalt, titanium, and a handful of other metallic elements. Some special elements (e.g., zirconium and beryllium) also may find structural use in applications such as nuclear reactors or other devices. Precious metals (e.g., gold, platinum, and silver) are used for applications from jewelry to jet engines! Zinc-based alloys are important in die casting of small articles. [Note: The words *article* or *component* may be used interchangeably to indicate a designed item which, in service, will likely be part of a collection of items (e.g., a gas turbine). A gas turbine high-pressure turbine blade would be an article or a component.]

37.3 WHAT ARE ALLOYS AND WHAT AFFECTS THEIR USE?

Alloys, for purposes of this chapter, are solid entities of a given chemical composition, crystal structure, and metallurgical structure at the temperature of use. The range of temperature use can be from near absolute zero to many thousands of degrees. Alloys are made from metallic elements (and some nonmetallic ones added in minor amounts), which are recovered from various ores. In the process of recovery from ores and transfer to desired products, alloys most often are subject to melting (liquification), casting, and/or mechanical deformation processes. This chapter is not concerned with extraction from ores or the subsequent purification of alloy elements before the creation of specific alloys for structural use.

Alloys normally exist as a crystalline arrangement of individual atoms (see Figure 37.1). Aggregates of multiples of the basic crystalline individual units (e.g., face-centered cubic and body-centered cubic) are termed *grains*. These grains have the same arrangement/orientation of unit cells of the given alloy across their dimension. The mechanical properties of alloys are normally affected by chemistry and intrinsic unit-cell capability, grain size, grain shape and grain orientation, temperature, and certain other factors. Typically, properties of interest are tensile properties (Figure 37.2), fatigue properties (Figure 37.3), and high-temperature time-dependent properties such as creep rupture (Figure 37.4). Modulus of elasticity is another mechanical property used in design and additional properties such as toughness or crack propagation characteristics can be incorporated into structural design, though such properties are not always available.

The methods of preparing alloys can significantly affect their resultant properties. The preparation of an alloy normally will result from melting appropriate elements/alloys followed by pouring (casting) the molten alloy into a mold to produce an electrode for further primary melting or an ingot for remelting or for deformation processing. Generic mold shapes may be used to produce ingots if the cast alloy is to be remelted and cast into a specific mold shape (e.g., a blade for use in a gas turbine engine). Specific mold shapes may be used if articles are to be created directly by casting. The electrode, ingot, or component produced will show a “cast” structure, which will be influenced by the casting process and subsequent treatments applied to create the desired alloy properties.

Often, an alloy may be processed by one or more deformation processes, that is, a billet cut from an ingot may be formed to shape or mostly to shape by hot rolling, forging, and so on, to produce a “wrought” structure. There are a variety of possible cast structures

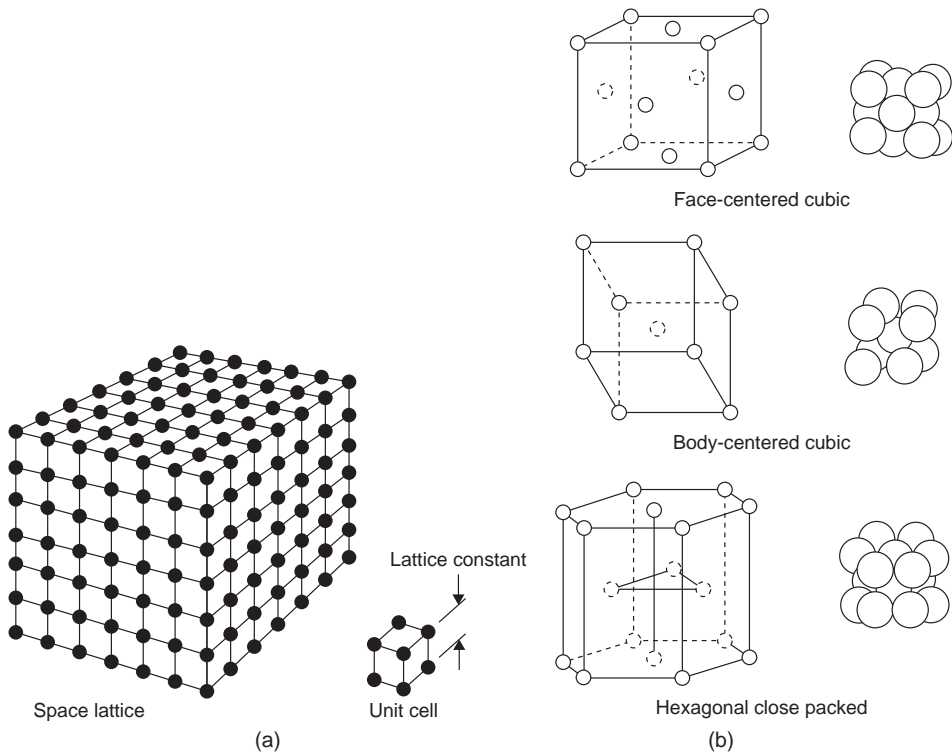


FIGURE 37.1 (a) Simple cubic lattice. Atoms are represented by solid spheres and, in this lattice, atoms are located at the corners of cubes which repeat indefinitely in three-dimensional space. (b) Common lattice structures for most metals/alloys.

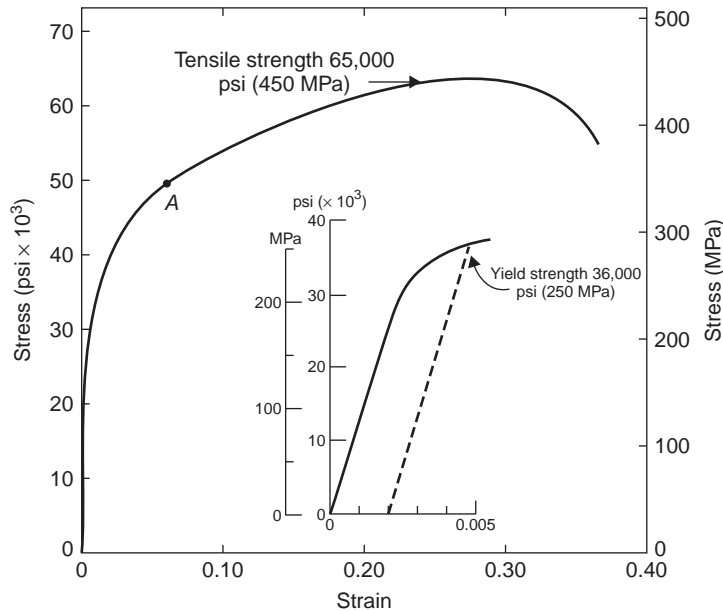


FIGURE 37.2 Ultimate tensile strength, UTS or TS, and defined yield strength at 0.2% offset (inset), TYS or YS. Young's modulus is the slope of the elastic region before plastic deformation begins.

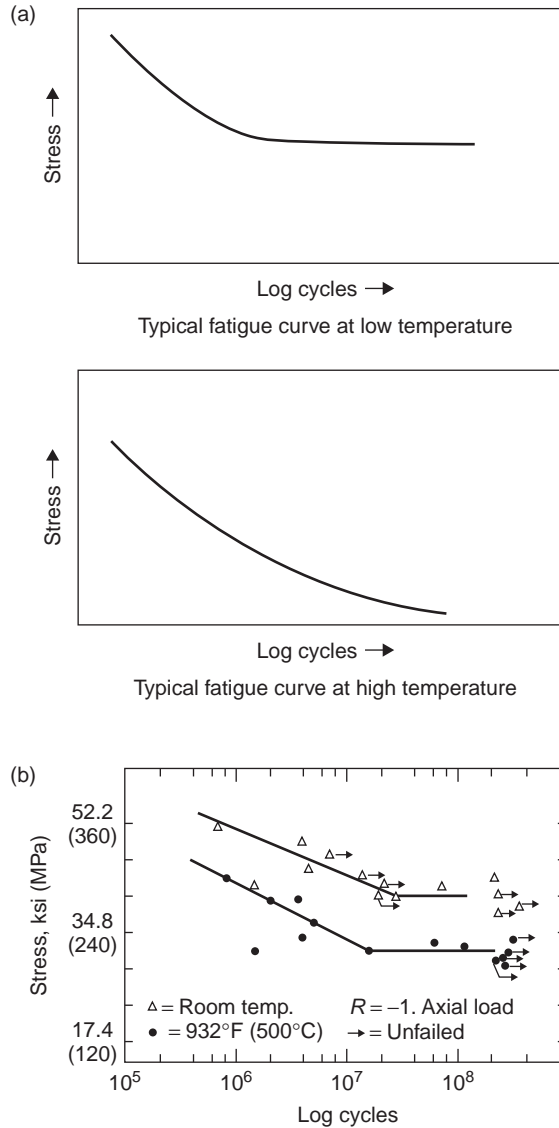


FIGURE 37.3 (a) Typical fatigue response at high temperature versus that at low temperature. (b) Actual fatigue curves at room and one elevated temperature for specific nickel-base superalloy.

as well as a variety of wrought structures. By examining a component under a microscope, microstructure (appearance at $100\times$ magnification and up) can be viewed. Alloys also can be examined at low magnifications for macrostructure (probably $1\times$ to $10\times$ magnification). Subsequent to production of a component, the application of special heating cycles (heat treatment) is often used to convey certain property values or characteristics to alloys. In-process heat treatments may be used during the shaping of the component.

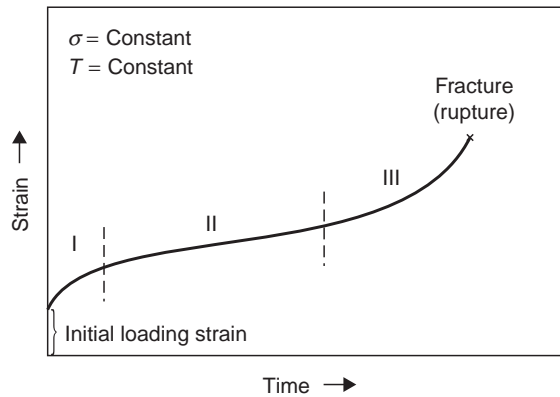


FIGURE 37.4 Time-dependent deformation under constant load at constant high temperature followed by rupture. All loads below short-time yield strength.

37.4 WHAT ARE THE PROPERTIES OF ALLOYS AND HOW ARE ALLOYS STRENGTHENED?

Metals are crystalline as noted above and, in the solid state, the atoms of a metal or alloy may arrange themselves in various crystallographic groupings, often occurring as cubic structures. Some crystal structures tend to be associated with better property characteristics than others. Crystalline structure normally is a characteristic of the major (base) alloy element in an alloy. In addition to basic crystal structure, crystalline aggregates of atoms have orientation relationships in space. As noted, crystalline aggregates are called grains and, in an alloy, there are usually many grains with random orientation directions within each manufactured article. Metal alloy articles with multiple random grain directions are known as polycrystalline (frequently referred to as having equiaxed grain structure), often having roughly equal dimensions in all directions. However, columnar-shaped grains (with one relatively longer axis) are common in polycrystalline cast products.

Properties of alloys may be mechanical, physical, or chemical. Mechanical behavior encompasses the strength properties determined in appropriate tests and tends to be the more important properties for selection of alloys for structural applications. Some mechanical properties of alloys are tensile or compressive strength (e.g., onset of plasticity, yielding strength, and fracture strength under normal stress loading), fatigue strength (i.e., cycles to first crack or fracture or stress for a specific number of failure cycles), modulus of elasticity, and so on, which are measured primarily from lower to intermediate temperatures of operation. In addition, high-temperature mechanical properties such as creep or creep rupture strength, fatigue strength, and thermally induced fatigue capability are of concern but primarily at temperatures above about 0.5 of the absolute melting point of an alloy. At high temperatures, mechanical properties may be influenced by sequences of cyclic loading and time as well as frequency of load application in addition to temperature of testing.

Physical properties include heat capacity, electrical conductivity, thermal conductivity, and magnetic properties. Chemical properties may be of most concern in environmental attack when alloys for structural applications are considered. Intrinsic chemical properties

of alloys are modified (as are mechanical properties) by chemical changes during alloy invention and by heat treatment of alloys during processing. In addition, surface chemical properties are frequently modified by surface changes induced, for example, by coatings.

While the physical properties to a major extent depend on the chemistry of the alloy, the mechanical properties are dependent not only on the chemistry of an alloy but more importantly on the microstructure. Microstructure is assumed to mean grain shape, orientation, grain size, and so on, and the presence, absence, type, and location of new crystal phases in the alloy. The mechanical properties result from the interaction of imperfections such as dislocations with the microstructural components.

The peripheral surface of a grain is called a grain boundary. Aggregates of atoms without grain boundaries are rarely created in nature. However, alloys without internal grain boundaries (i.e., single crystals) or alloys with aligned boundaries (i.e., columnar-grained structures) can be produced in some systems by appropriate manufacturing techniques. The introduction of different atom types and new crystal phases and/or the manipulation of grain boundaries plus the use of stored energy of deformation enable inhibition of the movement of the imperfections (e.g., dislocations) that enable deformation to occur.

It is quite important for the engineer selecting alloys to have a realistic understanding of the strengthening process in alloys as the mechanical properties of alloys can be modified considerably by processing to manipulate the strengthening level achieved. The energy of mechanical deformation (working) can be stored in an alloy (work hardening), producing higher strengths than the unworked (e.g., soft, annealed) alloy. Alloys such as steels (ferrous alloys) can have increased strengths produced by the introduction of fine structure from transformations of crystal lattices. This fine structure usually results from the formation of nonequilibrium martensite phases (in preference to the expected equilibrium phases). The fine structure produced by the martensitic reaction acts to restrict deformation by dislocation movements. This type of hardening often produces very high strengths (as in iron alloys) but much reduced ductility. Subsequent heat treatment (tempering) normally is used to enhance ductility at the expense of strength in transformation-hardened alloys. Industrial use of martensite to strengthen alloys is most prevalent with iron-base alloys (containing carbon) or in titanium alloys.

Dislocation movement at lower temperatures is reduced not only by fine structure but also by smaller grain sizes. Ultimate tensile strength is not particularly affected by grain size changes, although ductility effects are noticed. As operating temperatures increase, finer grain size material usually becomes less strong than coarse-grained material. Increased temperatures and time at high temperatures may cause recrystallization (new grain formation and subsequent loss of stored energy) and/or grain growth. Consequently, work hardening to increase strength has upper limits of application since the high-strength benefits of (cold) work may be lost with time at higher operating temperatures.

Some alloys derive their strength from solid-solution hardeners (elements dissolved in the basis metal) as well as from secondary precipitates that form in the matrix (principal phase and produce precipitation (age) hardening from the dispersed particles. Precipitates may be phases or they may be zones of enrichment in one of the alloy elements present. In lower temperature use alloys, such as aluminum, age hardening may be from zones or phases. In higher temperature use alloys, precipitated phases are the norm. Principal strengthening precipitate phases vary with alloy but in nickel-base and iron-nickel-base super alloys they are ordered particles of γ' , η , and/or γ'' . Phases such as carbides may provide limited direct strengthening (e.g., through dispersion hardening) or, more commonly, indirectly (e.g., by stabilizing grain boundaries against movement at high

temperature). In some instances, ceramic (e.g., silica, yttria) or other phases may be dispersed mechanically instead of metallurgically in a matrix to produce dispersion hardening. Thus, the hardening of alloys can consist of

- solid-solution hardening,
- work hardening,
- dispersion/precipitate hardening (sometimes with ordered precipitates),
- transformation (martensite) hardening,
- grain size/morphology (shape/orientation) hardening.

The results of hardening vary by alloy and by the temperature of testing. Improvements in lower temperature short-time strength may not necessarily be transferred to high-temperature ranges of operation.

Control of grain structure can significantly influence mechanical properties. The extent to which the secondary phases contribute directly to strengthening depends on the alloy and its processing. It should be noted that improper distributions of phases such as carbides and precipitate phases can be detrimental to properties. Work hardening is an important characteristic of alloys since work hardening may be used to strengthen an alloy (Figure 37.5) but also may contribute to the need for intermediate heating (in-process annealing) during forging/forming of an article to prevent cracking of the article or damage to fabrication equipment.

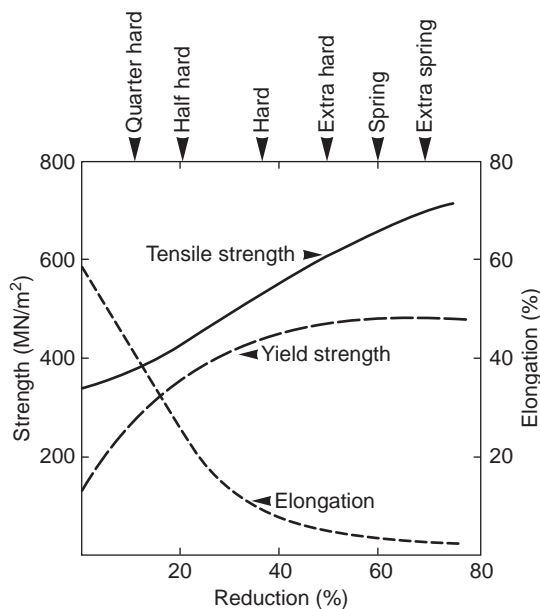


FIGURE 37.5 Terms used in copper industry to indicate the degree of cold working are shown at the top of the plot.

37.5 MANUFACTURE OF ALLOY ARTICLES

Appropriate compositions of most alloy types can be worked, that is, they can be forged, rolled to sheet, or otherwise deformed mechanically into a variety of shapes. More highly alloyed compositions may be processed as castings. Castings may be less costly than equivalent forgings. Large and small castings can be made. Fabricated alloy structures can be built up by combining (joining) separate article pieces, particularly by welding or brazing, but the more highly alloyed the composition, the more difficult it may be to join an alloy.

Many alloys may be available and used in cast or wrought form. In the latter situation, alloys may be available in extruded, forged, or rolled form. Bar stock, strip, sheet, plate, wire, and other wrought forms/sizes may be obtained. Deformation of alloys may be performed hot or cold. Often, higher strength or “exotic” alloys are only able to be produced and used in the cast condition. Powder metallurgy (PM) processing is an accepted method to produce articles of difficult-to-process articles of certain higher strength or demanding alloys in “wrought” conditions. However, simple components of ordinary alloys, particularly of ferrous metals, are often produced by PM for reduced cost compared to conventional cast or wrought processing.

As noted above, cold deformation (work hardening), for example, by cold rolling, may be used to increase short-time strength properties for applications at low to intermediate temperatures. Energy stored by deformation processing is an accepted method for improving alloy mechanical properties below a use temperature of about 0.5 of the absolute melting point. As noted above, heat treatment (heating of components in sometimes multiple and complex steps) is frequently used to obtain desired mechanical property values.

37.6 ALLOY INFORMATION

There is no substitute for consultation with alloy producers or manufacturers about the forms (cast, wrought, and PM), which can be provided and the exact chemistries/properties available in alloys. It should be understood that not all alloys are readily available as off-the-shelf items. Although many thousands of alloy compositions have been evaluated, some for over a century, only a relative handful are routinely produced. Moreover, some alloys are not available for use in all forms and sizes. Sometimes, the highest strength alloys will be useful only as powder metal products or as castings. In many instances, specific alloy compositions may only be available at limited times in a production cycle.

Design data for alloys are not intended to be conveyed here, but Tables 37.1–37.9 offer limited information on some typical properties for a few alloy classes. Design properties should be obtained from internal testing if possible. Data from producers or other validated sources may be substituted if sufficient test data are not available in-house. Typical properties are merely a guide for comparison. Exact alloy chemistry, article section size, heat treatment, and other processing steps must be known to generate adequate property values for design.

The alloy selector needs to be aware that processing treatments such as forging conditions, heat treatment, and coatings for corrosion protection dramatically affect properties of alloys. All data should be reconciled with the actual manufacturing specifications and processing conditions expected. Alloy selectors should work with competent

TABLE 37.1 Elastic Moduli and Tensile Strengths of Selected Representative Metals and Alloys

Material	Elastic Modulus (GPa)		Tensile Strengths (MPa)
	Absolute	Specific	
<i>Pure metals</i>			
Aluminum	70	26	70
Beryllium	295	160	290
Copper	130	14	250
Iron	210	27	420
Magnesium	45	26	180
Molybdenum	270	25	700
Nickel	220	25	340
Titanium	120	26	300
<i>Alloys</i>			
Aluminum	70	26	80–400
Copper	130	14	250–750
Iron	210	27	400–2000
Titanium	120	26	750–1500

metallurgical engineers to establish the validity of data intended for design as well as to specify the processing conditions that will be used for component production.

Application of design data must be taken into consideration the probability of components containing locally inhomogeneous regions under some circumstances. Short-time properties such as tensile strength may be little affected by local inhomogeneities. Time- or cyclic-dependent properties at high temperatures and cyclic properties at low to intermediate temperatures can be detrimentally affected by inhomogeneities. For wrought alloys, the probability of occurrence of such regions (which are highly detrimental to fatigue life) is principally dependent upon the melting method selected to produce the cast ingot for subsequent processing. For alloys, whether used as cast or in the wrought form, the degree of inhomogeneity and the likelihood of defects such as porosity are related to the alloy composition, the casting technique used, and the complexity of the final component. Defects may be detected by metallurgical surface examination, radiography, or sonic inspection, but all methods have limitations on the defect size that can be detected.

For sources of property data other than data from the producers (e.g., melters and forgers) of an alloy or from an alloy selector's own institution, one may refer to organizations such as ASM International, which publish compilations of data that may form a basis for the development of design allowables for many alloys. A list of some trade and professional organizations where alloy information may be obtained is provided at the end of this chapter.

Standards organizations such as the American Society for Testing and Materials (ASTM) publish information about alloys, but that information does not ordinarily contain design data. It is important to note that the same nominal alloy chemistry may have some composition modifications made from one manufacturer or customer to another. Sometimes, this extends from one country to another. Tweaking of the casting or "wroughting" processes or the heat treatment often associated with what seem to be minor

TABLE 37.2 Mechanical Characteristics and Typical Applications for Carbon and Alloy Steels

Hot-Rolled Material					
AISI/SAE or ASTM Number	Tensile Strength [psi × 10 ³ (MPa)]	Yield Strength [psi × 10 ³ (MPa)]	Ductility (% Elongation in 2 in.)	Typical Applications	
Plain low-carbon steels					
1010	47 (325)	26 (180)	28	Automobile panels, nails, and wire	
1020	55 (380)	30 (205)	25	Pipe; structural and sheet steel	
A36	58 (400)	32 (220)	23	Structural (bridges and buildings)	
A516 Grade 70	70 (485)	38 (260)	21	Low-temperature pressure vessels	
High-strength, low-alloy steels					
A440	63 (435)	42 (290)	21	Structures that are bolted or riveted	
A633 Grade E	75 (520)	55 (380)	23	Structures used at low ambient temperatures	
A656 Grade 1	95 (655)	80 (552)	15	Truck frames and railway cars	
Oil-Quenched and Tempered					
Mechanical Property Ranges					
AISI Number	UNS Number	Tensile Strength [psi × 10 ³ (MPa)]	Yield Strength [psi × 10 ³ (MPa)]	Ductility (% Elongation in 2 in.)	Typical Applications
Plain carbon steels					
1040	G10400	88–113 (605–780)	62–85 (430–585)	33–19	Crankshafts, bolts
1080 ^b	G10800	116–190 (800–1310)	70–142 (480–980)	24–13	Chisels, hammers
1095 ^b	G10950	110–186 (760–1280)	74–120 (510–830)	26–10	Knives, hacksaw blades
Alloy steels					
4063	G40630	114–345 (786–2380)	103–257 (710–1770)	24–4	Springs, hand tools
4340	G43400	142–284 (980–1960)	130–228 (895–1570)	21–11	Bushings, aircraft tubing
6150	G61500	118–315 (815–2170)	108–270 (745–1860)	22–7	Shafts, pistons, gears

^aAISI, American Iron and Steel Institute; SAE, Society of Automotive Engineers; ASTM, American Society for Testing and Materials; UNS, Unified Numbering System.

^bClassified as high-carbon steels.

Source: Data for hot-rolled material adapted from *Metals Handbook: Properties and Selection, Irons and Steels*, Vol. 1, 9th ed., Bardes, B. editors. Materials Park, OH: American Society for Metals; Materials Park, OH, 1978. pp. 190, 192, 405, 406.

TABLE 37.3 Designations, Compositions, Mechanical Properties, and Typical Applications for Austenitic, Ferritic, Martensitic, and Precipitation-Hardenable Stainless Steels

AISI Number	UNS Number	Composition (wt%) ^a				Condition ^b	Mechanical Properties			Typical Applications
		C	Cr	Ni	Other		Tensile Strength [psi × 10 ³ (MPa)]	Yield Strength [psi × 10 ³ (MPa)]	Ductility (% Elongation in 2 in.)	
<i>Ferritic</i>										
409	S40900	0.08	11		1.0Mn, 0.75Ti	Annealed	65 (448)	35 (240)	25	Automotive exhaust
446	S44600	0.20	25		1.5Mn	Annealed	80 (552)	50 (345)	20	Valves (high temperature), glass molds
<i>Austenitic</i>										
304	S30400	0.08	19	9	2.0Mn	Annealed	85 (586)	35 (240)	55	Food processing
316L	S31603	0.03	17	12	2.0Mn, 2.5Mo	Annealed	80 (552)	35 (240)	50	Welding construction
<i>Martensitic</i>										
410	S41000	0.15	12.5		1.0Mn	Annealed	70 (483)	40 (275)	30	Rifle barrels, cutlery
440A	S44002	0.70	17	1.0Mn, 0.75Mo	Q and T	140 (965)	100 (690)	23	Cutlery, surgical tools	
					Annealed	105 (724)	60 (414)	20		
							260 (1790)	240 (1655)	5	
<i>Precipitation hardenable</i>										
17-7PH	S17700	0.09	17	7	1.0Mn, 1.0Al	Solution treated	130 (897)	40 (275)	35	Knives, springs
						Precipitation hardened	215 (1480)	195 (1345)	9	

^aThe balance of the composition is iron.
^bQ and T[™] denotes quenched and tempered.
Source: Adapted from *Metal Progress 1982 Materials and Processing Databook*. Copyright © 1982 American Society for Metals.

TABLE 37.4 Designations, Minimum Mechanical Properties, Approximate Compositions, and Typical Applications for Various Gray, Nodular, and Malleable Cast Irons

Grade	UNS Number	Composition (wt%) ^a			Mechanical Properties				
		C	Si	Other	Matrix Structure	Tensile Strength	Yield Strength	Ductility	
						[psi × 10 ³ (MPa)]	[psi × 10 ³ (MPa)]		
<i>Gray iron</i>									
SAEG2500	F10005	3.3	2.2	0.7Mn	Pearlite + ferrite	25 (173)	–	–	Engine blocks, brake drums
SAEG4000	F10008	3.2	2.0	0.8Mn	Pearlite + ferrite	40 (276)	–	–	Engine cylinders and pistons
<i>Ductile (nodular) iron</i>									
ASTM A536 60-40-18	F32800	3.5–3.8	2.0–2.8	0.05Mg, <0.20Ni, <0.10Mo	Ferrite	60 (414)	40 (276)	18	Valve and pump bodies
	F34800					100 (690)	70 (483)	3	High-strength gears
	120-90-02					F36200	120 (828)	90 (621)	2
<i>Malleable iron</i>									
32510	F22200	2.3–2.7	1.0–1.75	<0.55Mn	Ferrite	50 (345)	32 (224)	10	General engineering service at room and elevated temperatures
45006	–	2.4–2.7	1.25–1.55	<0.55Mn	Ferrite + pearlite	65 (448)	45 (310)	6	

^aThe balance of the composition is iron.

Source: Adapted from *Metals Handbook: Properties and Selection: Irons and Steels*, Vol. I, 9th ed., Bardes, B. editors. Materials Park, OH: American Society for Metals; 1978.

TABLE 37.5 Compositions, Mechanical Properties, and Typical Applications for Eight Copper Alloys

Alloy Name	UNS Number	Composition (wt%)				Mechanical Properties			Typical Applications	
		Cu	Zn	Sn	Other	Condition	Tensile Strength	Yield Strength		Ductility
							[psi × 10 ³ (MPa)]	[psi × 10 ³ (MPa)]		(% Elongation in 2 in.)
<i>Wrought alloys</i>										
Electrolytic tough pitch	C11000	99.9	–	–	0.04O	Annealed	32 (220)	10 (69)	55	Roofing, rivets, radiators
Beryllium– copper	C17200	97.9	–	–	1.9Be, 0.2Co	Annealed	68 (470)	25 (172)	48	Springs, diaphragms
Cartridge brass	C26000	70	30	–	–	Precipitation hardened Annealed	165 (1140)	145 (1000)	7	
Phosphor bronze, 5% A	C51000	95	–	5	Trace P	Annealed	44 (303)	11 (76)	66	Ammunition components
Copper– nickel, 30%	C71500	70	–	–	30Ni	Annealed	47 (324)	19 (131)	64	Bellows, welding rods
							54 (372)	20 (138)	45	Saltwater piping
<i>Cast alloys</i>										
Leaded yellow brass	C85400	67	29	1	3Pb	As cast	34 (234)	12 (83)	35	Battery clamps, fittings
Tin bronze	C90500	88	2	10	–	As cast	45 (310)	22 (152)	25	Bearings, bushings
Aluminum bronze	C95400	85	–	–	4Fe, 11Al	As cast	85 (586)	35 (241)	18	Gears, valve seats

Source: *Metal Progress 1980 Databook*. Copyright © 1980 American Society for Metals.

TABLE 37.6 Compositions, Mechanical Properties, and Typical Applications for Eight Common Aluminum Alloys

Aluminum Association Number	UNS Number	Composition (wt%) ^a				Mechanical Properties				
		Cu	Mg	Mn	Other	Condition	Tensile Strength	Yield Strength	Ductility (%)	
							[psi × 10 ³] (MPa)]	[psi × 10 ³] (MPa)]	Elongation in 2 in.)	
Wrought, non-heat-treatable alloys										
1100	A91100	0.12	–	–	–	Annealed	13 (90)	5 (34)	35	Sheet metal work
3003	A93003	0.12	–	1.2	–	Annealed	16 (110)	6 (42)	30	Cooking utensils
5052	A95052	–	2.5	–	0.25Cr	Annealed	28 (195)	13 (90)	25	Bus, truck uses
Wrought, heat-treatable alloys										
2014	A92014	4.4	0.5	0.8	0.8Si	Heat treated	70 (485)	60 (415)	13	General structures
6061	A96061	0.3	1.0	–	0.6Si, 0.2Cr	Heat treated	45 (310)	40 (275)	12	Trucks, towers, furniture
7075	A97075	1.6	2.5	–	5.6Zn, 0.23Cr	Heat treated	83 (570)	73 (505)	11	Aircraft structural parts
Cast, heat-treatable alloys										
295.0	A02950	4.5	–	–	1.1Si	Heat treated	36 (250)	24 (165)	5	Crankcases, aircraft wheels
356.0	A03560	–	0.3	–	7.0Si	Heat treated	33 (230)	24 (165)	4	Water-cooled cylinder blocks

^aThe balance of the composition is aluminum.

Source: Adapted from *Metals Handbook: Properties and Selection: Nonferrous Alloys and Pure Metals*, Vol. 2, 9th ed., Baker, H. Managing Editor. Materials Park, OH: American Society for Metals; 1979.

TABLE 37.7 Compositions, Mechanical Properties, and Typical Applications for Six Common Magnesium Alloys

ASTM Number	UNS Number	Composition (wt%) ^a				Mechanical Properties				Typical Applications
		Al	Mn	Zn	Other	Condition	Tensile Strength [psi × 10 ³ (MPa)]	Yield Strength [psi × 10 ³ (MPa)]	Ductility (% Elongation in 2 in.)	
Wrought alloys										
AZ80A	M11800	8.5	0.12	0.5	—	As extruded	49 (340)	36 (250)	11	Highly stressed extrusions
HM31A	M13312	—	1.20	—	3.0Th	Artificially aged	37 (255)	26 (179)	4	Missile and aircraft use to 425 °C
ZK60A	M16600	—	—	5.5	0.45Zr	Artificially aged	51 (350)	41 (285)	11	Forgings of maximum strength for aircraft
Cast alloys										
AZ92A	M11920	9.0	0.10	2.0	—	As cast	25 (170)	14 (97)	2	Pressure-tight castings
EZ33A	M12330	—	—	2.6	3.2 Rare earths, 0.7Zr	Artificially aged	23 (160)	16 (110)	3	Pressure-tight castings for use between 175 and 250 °C
AZ91A	M11910	9.0	0.13	0.7	—	As cast	33 (230)	24 (165)	3	Parts for cars, lawnmowers, luggage

^aThe balance of the composition is magnesium.
Source: Adapted from *Metals Handbook: Properties and Selection: Nonferrous Alloys and Pure Metals*, Vol. 2, 9th ed., Baker, H. Managing Editor. Materials Park, OH: American Society for Metals; 1979.

TABLE 37.8 Compositions, Mechanical Properties, and Typical Applications for Four Common Titanium Alloys

Alloy Type	UNS Number	Composition (wt%)	Condition	Mechanical Properties			Typical Applications
				Tensile Strength [psi × 10 ³ (MPa)]	Yield Strength [psi × 10 ³ (MPa)]	Ductility (% Elongation in 2 in.)	
Commercially pure	R50550	99.1Ti	Annealed	75 (517)	65 (448)	25	Chemical, marine, aircraft parts
α	R54521	5Al, 2.5Sn, balance Ti	Annealed	125 (862)	117(807)	16	Aircraft engine compressor blades
$\alpha-\beta$	R56401	6Al, 4V, balance Ti	Annealed	144 (993)	134 (924)	14	Rocket motor cases
β	R58010	13V, 11Cr, 3Al, balance Ti	Precipitation hardened	177 (1220)	170 (1172)	8	High-strength fasteners

Source: Adapted from *Metal Progress 1978 Databook*. Copyright © 1978 American Society for Metals.

TABLE 37.9 Characteristics of Selected Superalloys

Designation	Alloy Type	Form	Condition	Ultimate TS, 1000 °F (540 °C)	0.2% YS, 1000 °F (540 °C)	10 ³ h Rupture Life, 1400 °F (760 °C)	10 ³ h Rupture Life, 1600 °F (870 °C)	Applications
Haynes 188	Cobalt base	Sheet	Prob. annealed	107 (740)	44 (305)	24 (165)	10 (70)	GT burners
Hastelloy X	Nickel base	Sheet	Prob. annealed	94 (650)	42 (290)	15 (105)	6 (40)	GT burners
Inconel X-750	Nickel base	Bar	Age hardened	152 (1050)	105 (725)	—	7 (50)	GT/ST general
A 286	Iron base	Bar	Age hardened	131 (905)	88 (605)	15 (105)	—	GT/ST disks
Waspaloy	Nickel base	Bar	Age hardened	170 (1170)	105 (725)	42 (290)	16 (110)	GT disks, hardware
Udimet 700	Nickel base	Bar	Age hardened	185 (1275)	130 (895)	62 (425)	29 (200)	GT blades
Astrolloy	Nickel base	Bar	Age hardened	180 (1240)	140 (965)	62 (425)	25 (170)	GT disks
Inconel 718	Nickel base	Bar	Age hardened	185 (1275)	154 (1065)	28 (195)	—	GT disks
Rene 95	Nickel base	Bar	Age hardened	224 (1550)	182 (1255)	—	—	GT disks
Inconel 718	Nickel base	Cast	Age hardened	—	—	—	—	GT cases
IN 713 C	Nickel base	Cast	Age hardened	125 (860)	102 (705)	44 (305) ^a	31 (215)	GT blades
IN 792	Nickel base	Cast	Age hardened	—	—	55 (380) ^a	38 (260)	GT blades
Rene 80	Nickel base	Cast	Age hardened	—	—	—	35 (240)	GT blades
PWA 1480	Nickel base	Cast (SC)	Age hardened	164 (1130)	131 (905)	—	—	GT blades
CMSX-2	Nickel base	Cast (SC)	Age hardened	188 (1295)	181 (1245)	—	50 (345)	GT blades
X-40	Cobalt base	Cast	As cast	80 (550)	40 (275)	—	15 (105)	GT vanes
W152	Cobalt base	Cast	As cast	108 (745)	64 (440)	—	22 (150)	GT vanes
Mar-M 509	Cobalt base	Cast	Prob. as cast	83 (570)	58 (400)	—	20 (140)	GT vanes

Note: TS = tensile strength, YS = yield strength, SC = single crystal, GT = gas turbine, ST = steam turbine, ksi (mPa) = strength units.

^a1500°F (815°C).

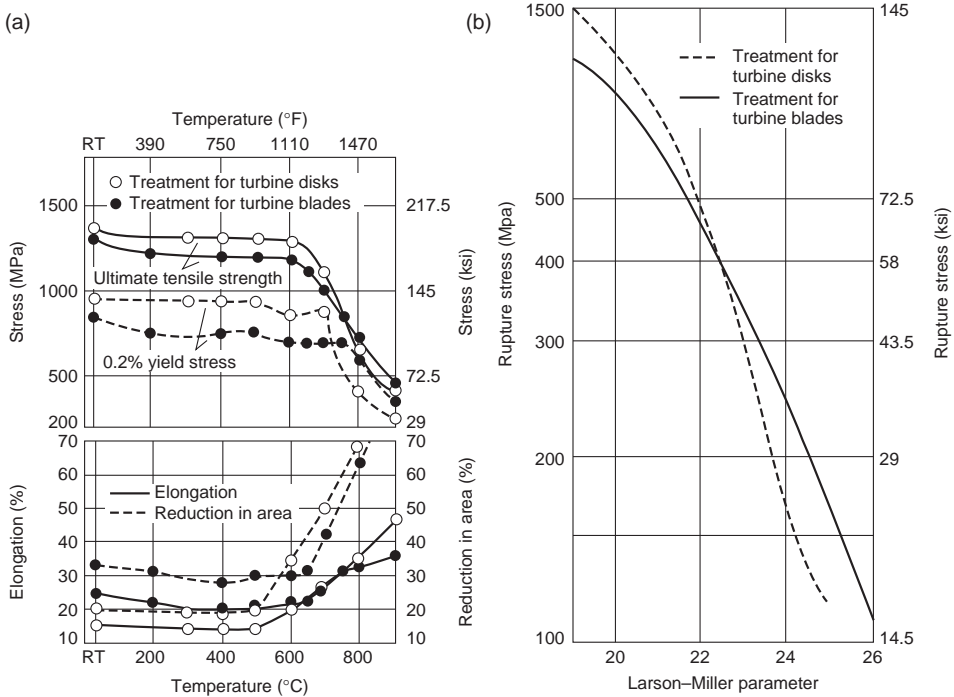


FIGURE 37.6 Heat treatment for two different applications showing how (a) tensile properties and (b) rupture properties differ; $P_{LM} = T(C + \log t)$, where $C = 20$ is Larson–Miller constant, T is absolute temperature (K), and t is time in hours. RT = room temperature.

composition changes can cause significant variations in properties. All facets of chemistry and processing need to be considered when selecting an alloy for an application.

There are instances when different heat treatments may be applied to the same nominal alloy composition with resultant differences in properties. An illustration of this is shown in Figure 37.6 where a wrought superalloy (Waspaloy) was produced for two different gas turbine engine parts, a disk and a turbine blade. Note the changes in properties. It is also not unusual for the heat treatments applied to cast parts of the same alloy composition to differ from the heat treatments for wrought products of the same alloy.

37.7 METALS AT LOWER TEMPERATURES

37.7.1 General

Alloy strengths at lower temperatures (usually considerably below $0.5 T_m$) are not a function of time under normal conditions. Environmental attack (general overall corrosion, selective corrosion, etc.) can be significant and will be a function of time, temperature, and environment. Presence of notches and/or cyclic conditions may cause premature failure of a component. Consequently, in addition to evaluation of tensile (e.g., ultimate, yield strengths) properties, possible environmental considerations should be checked and

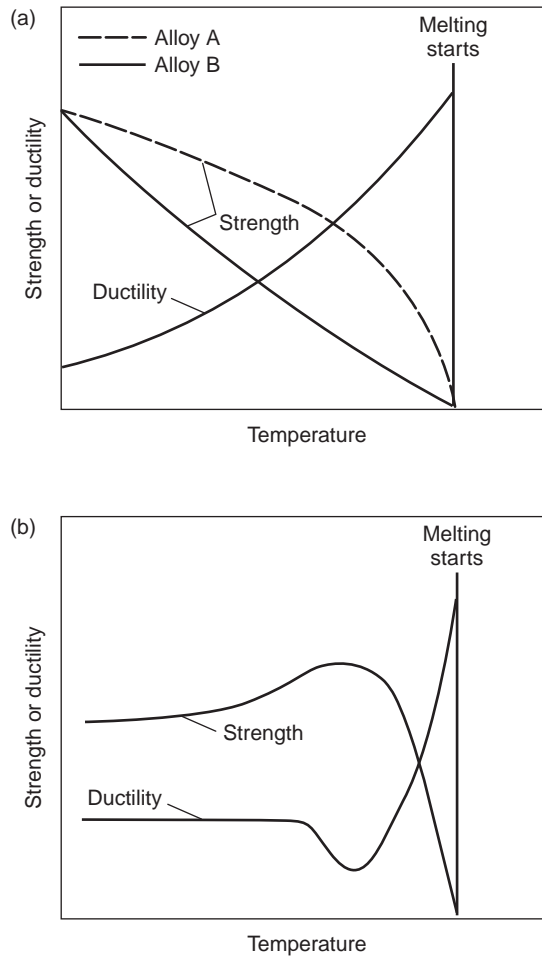


FIGURE 37.7 (a) Most metals and alloys show a decrease in strength and increase in ductility as temperatures increase. (b) Superalloys precipitation hardened by γ' particles often show peak in strength and minimum in ductility as temperatures increase.

the fatigue (low cycle/high cycle, smooth/notched) strengths, crack propagation characteristics, and toughness behavior should be reviewed.

37.7.2 Mechanical Behavior

Temperature or time of load application may influence the property behavior, including failure mode, of an article in mechanical testing. Failure may be defined as initiation of a crack or of fracture of an article into two or more pieces. In case of short-time tensile properties (yield strength, ultimate strength), strengths usually are reduced as temperatures increase from absolute zero to higher temperatures near about 0.5 of the absolute melting temperature. Figure 37.7 illustrates the trends in tensile strength and ductility of alloys as temperatures increase.

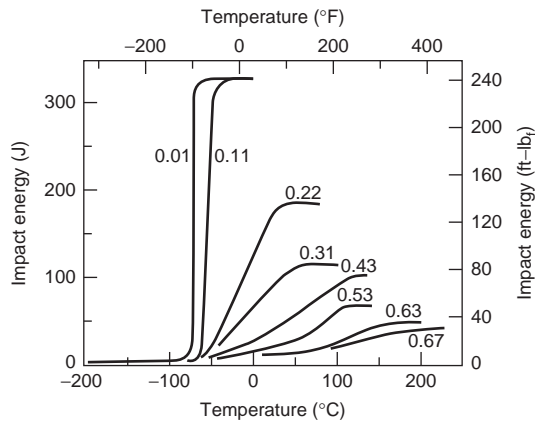


FIGURE 37.8 Charpy V-notch energy versus temperature behavior for steel.

Ductility of an alloy, as measured by elongation at fracture or reduction of area in cross section after fracture, tends to increase with temperature. Sometimes, ductility or toughness of an alloy is measured by energy absorbed during impact tests. Strength properties tend to appear in design calculations while ductility values do not. Ductility is considered important in actual alloy use since limited ductility may cause an alloy article to fail with no immediate prior indication of likelihood of failure. The “forgiveness” of higher ductility alloys makes designers more comfortable with application of some alloys. Attempts to build “ductility” considerations into alloy selection and design resulted in various types of fracture toughness property concepts being developed and applied to alloy selection. Fracture toughness (critical fracture toughness) is one concept for which design data may be available for alloys. Higher fracture toughness is most desirable, as is higher ductility.

Some alloys may exhibit a “ductile-to-brittle” transition. This concept occurs when an alloy undergoes a rapid transition to much lower failure ductility over a very small temperature range on cooling of the alloy. Steels, for example, may show a ductile-to-brittle transition, particularly at temperatures below room temperature. Figure 37.8 illustrates how toughness, as measured by ductility, for steels drops as temperature decreases (and carbon content increases).

Hardness properties (usually measured by means of standard indentation tests) may be used to judge an alloy’s strength capability at lower temperatures and confirm the results of heat treatment but play no role in design.

37.8 METALS AT HIGH TEMPERATURES

37.8.1 General

While material strengths at low temperatures are usually not a function of time, the time of load application, at high temperatures, becomes very significant for mechanical properties. Concurrently, environmental attack by oxygen and/or other elements at high temperatures accelerates the conversion of some of the metal atoms to oxides or other compounds. Environmental attack proceeds much more rapidly at high temperatures than at room or lower temperatures. Environmental considerations can vary significantly with different types of environment and different temperatures, in some instances, showing a

greater rate of attack at one temperature than the attack of the same alloy at a higher temperature. As in the case of lower temperatures, presence of notches and/or cyclic conditions may cause premature failure of a component. Thus, in addition to evaluation of short-time tensile (e.g., ultimate, yield strengths) properties, creep rupture properties, fatigue, and thermomechanical fatigue (smooth/notched) strengths, crack propagation characteristics, toughness behavior, and possible environmental considerations should be checked.

37.8.2 Mechanical Behavior

In case of short-time tensile properties (yield strength, ultimate strength), the mechanical behavior of alloys at higher temperatures is similar to that at room temperature, but with alloys becoming weaker as the temperature increases. However, when steady loads below the normal yield or ultimate strength (determined in short-time tests) are applied for prolonged times at higher temperatures, the situation is different. Figure 37.4 illustrates the way in which most alloys respond to steady extended-time loads at high temperatures.

Because of the higher temperature, a time-dependent extension (creep) is noticed under load. If the alloy is exposed for a long time, the alloy eventually fractures (ruptures). The degradation process is called creep or, in the event of failure, creep rupture (sometimes stress rupture) and alloys for elevated temperature use are selected on their ability to resist creep and creep rupture failure. Data for superalloys frequently are provided as the stress, which can be sustained for a fixed time (e.g., 100-h rupture) versus the temperature. Figure 37.9 shows such a plot with ranges of expected performance for various superalloy families.

One of the contributory aspects to elevated temperature failure is that alloys tend to come apart at the grain boundaries when tested for long times above about 0.5 of their absolute melting temperature. Thus, fine-grained alloys that are usually favored for lower temperature applications may not be the best materials for creep-rupture-limited applications at high temperatures. Elimination or reorientation/alignment of grain boundaries is sometimes a key factor in maximizing the higher temperature life of an alloy.

A factor not often recognized in mechanical behavior is that the static modulus (e.g., Young's modulus E) is affected by increases in temperature. This should be obvious from the above discussion about creep. The dynamic modulus of alloys shows a decrease with increasing temperatures. However, dependent on rate of load application, as test temperatures increase, static moduli determined by measurements from a (short time) tensile test tend to gradually fall below moduli determined by dynamic means. Because moduli are used in design and may affect predictions of life and durability, every effort should be made to determine the dynamic, not just static, moduli of alloys for high-temperature applications. Table 37.10 shows dynamic moduli of cast superalloys for illustration.

Moduli of cast alloys also may be affected dramatically by the orientation of grains or by use of single crystals. Thus, a columnar grain nickel-base super alloy will show a dynamic modulus parallel to the growth direction of the columnar grains that is around 18×10^6 psi (124.1 GPa) at 70°F (21°C). This is much lower than the polycrystalline (random-orientation) value of around 30×10^6 psi (206.8 GPa). A more accurate method of depicting moduli would be to use the appropriate single-crystal elastic constants (three needed) of an alloy, but these are rarely available.

Cyclically applied loads that cause failure (fatigue) at lower temperatures also cause failures in shorter times (lesser cycles) at high temperatures. For example, Figure 37.3

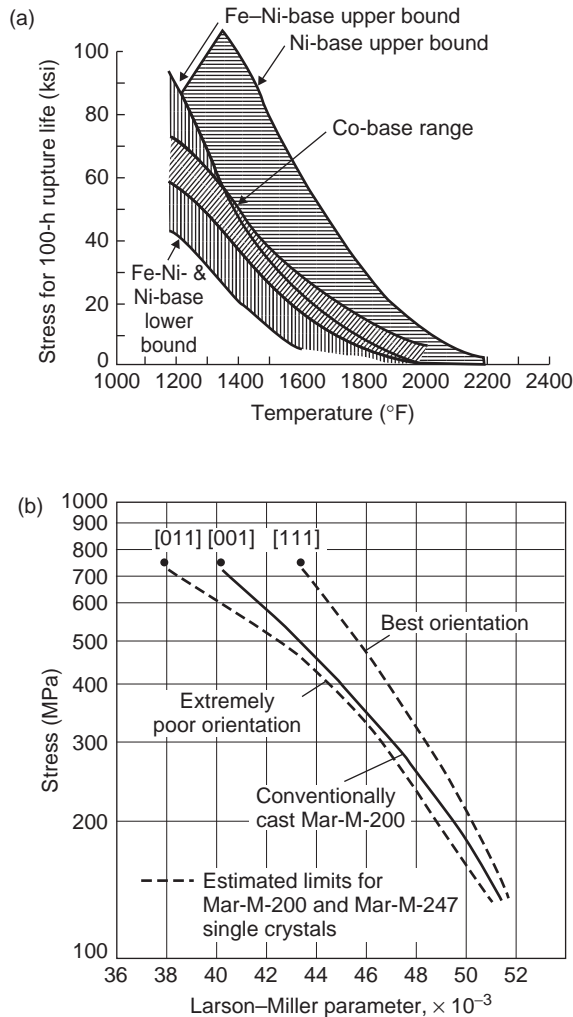


FIGURE 37.9 (a) Stress for 100-h rupture life for three classes of superalloys. (b) Log failure stress versus Larson–Miller parameter for single-crystal nickel-base superalloys in three possible test orientations ($P_{LM} = \text{actual } P_{LM} \times 1.8$). $P_{LM} = T(C + \log t)$, where $C = 20$ is Larson–Miller constant, T is absolute temperature (K), and t is time in hours.

shows schematically how the cyclic resistance is degraded at high temperatures when the locus of failure is plotted as stress versus applied cycles (S – N) of load. From the S – N curves shown, it should be clear that there is not necessarily an endurance limit for metals and alloys at high temperatures.

Cyclic loads can be induced not only by mechanical loads in a structure but also by thermal changes. The combination of thermally induced and mechanically induced loads leads to failure in thermomechanical fatigue (TMF). TMF failures occur in a relatively low number of cycles. Thus, TMF is a low-cycle fatigue (LCF) process (less than about

TABLE 37.10 Dynamic Modulus of Elasticity for Selected Cast Superalloys

Alloy	Dynamic Modulus of Elasticity					
	At 21 °C (70° F)		At 538 °C (1000 °F)		At 1093° C (2000° F)	
	GPa	10 ⁶ psi	GPa	10 ⁶ psi	GPa	10 ⁶ psi
Nickel base						
IN-713 C	206	29.9	179	26.2		
IN-713 LC	197	28.6	172	25.0		
B-1900	214	31.0	183	27.0		
IN-100	215	31.2	187	27.1		
IN-162	197	28.5	172	24.9		
IN-738	201	29.2	175	25.4		
MAR-M-200	218	31.6	184	26.7		
MAR-M-246	205	29.8	178	25.8	145	21.1
MAR-M-247	—	—	—	—	—	—
MAR-M-421	203	29.4	—	—	141	20.4
Rene 80	208	30.2	—	—		
Cobalt base						
Haynes 1002	210	30.4	173	25.1		
MAR-M-509	225	32.7				

10⁵ cycles). LCF can be induced by high repeated mechanical loads while lower stress repeated mechanical loads lead to fatigue failure in a high number of cycles (HCF greater than 10⁶ cycles). Dependent on application, LCF failures in structures can be either mechanically induced or TMF type. In airfoils for the hot section of gas turbines, TMF is a major concern. In highly mechanically loaded parts such as gas turbine disks, mechanically induced LCF is the major concern. HCF normally is not a problem with alloys unless a design error occurs and a component is subjected to a high-frequency vibration that forces rapid accumulation of fatigue cycles. Although $S-N$ plots are common, the designer should be aware that data for LCF and TMF behavior frequently are gathered as plastic strain (ϵ_p or $\Delta\epsilon_p$) versus applied cycles. The final component application will determine the preferred method of depicting fatigue behavior.

While life under cyclic load ($S-N$ behavior) is a common criterion for design, resistance to crack propagation is an increasingly desired property. Thus, the crack growth rate (da/dn) versus a fracture toughness parameter is required. The toughness parameter, in this instance, is the stress intensity factor K range ΔK over an incremental distance, which a crack has grown. A plot of the resultant type (da/dn vs. ΔK) is shown schematically in Figure 37.10.

37.9 MELTING AND CASTING PRACTICES

37.9.1 Melting

To produce most alloys, the appropriate chemistries are made liquid; mixing occurs (if more than one element or master alloy is present) and then the liquid is poured into containers (molds) to produce the desired shape. Most alloys have melting points sufficiently

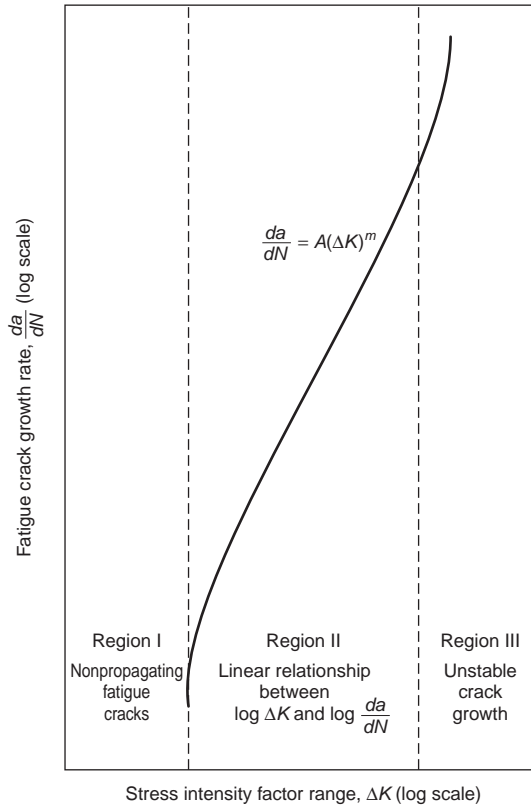


FIGURE 37.10 Three regions of crack growth response.

low that they can be melted and cast to shapes economically and with acceptable purity and quality. Melting consists of providing a furnace with an appropriate (usually ceramic-lined) crucible to contain the molten-metal charge and a power source (usually gas or electricity) to produce the appropriate temperatures for melting. For common industrial alloys (white metals, magnesium, aluminum, copper, steels, and superalloys), the melting range needed probably varies from around 800°F to about 2500°F. Melting temperatures for some alloys such as titanium alloys range higher than this.

As indicated earlier, there are occasions when a primary alloy to eventually be used in a final article shape is cast into an electrode or an ingot one or more times before an article is produced. Electrodes are the precursors to final forged or deformation processed articles. They are intended to be remelted by electrical energy to refine chemistry, microstructure, and so on, prior to being cast as ingots to be delivered to the manufacturer who may cut the ingots to billets from which to make the (raw shape at least) final article by deformation processing. Cast ingots may be remelted without additional processing for use in the direct casting of articles. Alloys produced using the variety of available furnaces in ambient air environments satisfy most common needs. Vacuum melting and similar processes produce higher quality alloys and usually are the production melting methods of choice for applications such as those in aerospace.

For metals such as titanium alloys and superalloys, melting is done under vacuum to retain and/or enhance an alloy's properties. Ordinarily, most alloys to be cast into articles or ingots (which are then processed into billets for wrought processing) are melted open to the atmosphere (air) except for the use of slags on the molten-alloy surface to minimize pickup of gases or loss of elements from the molten alloy or to introduce elements to the alloy from the slag. Furnaces may be resistance wire heated, gas fire heated, arc heated, or induction heated. For more sophisticated melting, especially where vacuum is required, alloys may be heated by induction, arc, electron beam, or plasma. The latter three processes can produce higher local heat input and so may be used to melt alloys such as titanium. The more sophisticated the melting techniques, the higher the cost but, generally, the better the quality.

Vacuum induction melting (VIM) is used for super alloys and, in some instances, stainless steels. The VIM furnace consists of a ceramic-lined crucible built up around water-cooled induction coils. The crucible is mounted in a vacuum chamber. The vacuum chamber may have several vacuum ports (of various sizes) built into it so that, without breaking vacuum,

- charge material may be introduced into the crucible,
- molds may be introduced into the chamber, and
- systems for removing slag from the pour stream (tundish) may be introduced.

Sampling of the molten metal is required for chemistry control in primary alloy production and when the VIM unit is producing ingot for subsequent processing. VIM is more costly than electric arc furnace or other routine processes for melting alloys. A major advantage of VIM is excellent control of chemistry and reduction of impurities such as gases (oxygen and nitrogen) and detrimental tramp elements. Sometimes, in lieu of VIM, an alloy (e.g., of steel) may be vacuum degassed (VD) after melting but before pouring into molds. Improved properties result from VIM or VD processing. The essence of selecting a melting process to create the desired composition and electrode/ingot/cast article structure hinges on the correct choice and adaptation of melting principles and attention to detail in processing.

Consumable vacuum arc remelting (CVAR, most often referred to as VAR) is used to primary melt and to remelt titanium and to remelt superalloys. High heat inputs can be achieved and the alloys are cast into water-cooled molds. No slag is used but rather the ultrahigh vacuum acts to protect the alloy. An electrode is used to strike an arc between the consumable electrode of the alloy and a striker plate in the water-cooled mold. The molten metal produced from the electrode by the arc drips into a pool and solidifies in characteristic ways related to alloy composition, melt rate, mold size, and so on. Generally, the mold is larger than the electrode. Molds for VAR processing are circular. Sometimes, multiple VAR is applied to an alloy.

Another melting process is electroslag remelting (ESR), which uses an arc within the slag to produce heat that melts the face of an electrode. In ESR, current is applied to an electrode situated inside a water-cooled crucible containing a molten-slag charge. The intended circuit of current is from the electrode, through the molten slag, through the solidifying ingot, through the water-cooled stool, and back to the electrode. The molten droplets from the electrode face fall through the slag to a molten pool and then solidify. Changes in chemistry can be effected in ESR to a

greater degree than in VAR processing. There are subtle aspects of the processing for both VAR and ESR, which may be discussed with a knowledgeable alloy producer. For superalloy primary melting, where larger ingot sizes of highly alloyed materials are needed, VAR is used in preference to ESR.

VAR and ESR are secondary melting processes (following primary VIM melting to produce electrodes). Several producers of critical rotating super alloy components in the gas turbine industry have adopted the use of a hybrid secondary melt process. Primary melting by VIM to ensure production of a low-oxygen, precise chemistry, initial electrode is followed by ESR. The ESR electrode will be clean and sound. The clean, sound electrode is then remelted by VAR. The improved cleanliness and soundness of the electrode facilitate VAR control. This product is referred to as “triple melt” and has a much reduced frequency of defect occurrence compared to double-melt (VIM + VAR) superalloy product. The cost of the triple-melt process is clearly higher than double melt (VIM + VAR or VIM + ESR).

As noted, alloy-melting practices may be classified as either primary (the initial melt of elemental materials and/or scrap which sets the composition) or secondary (remelt of a primary melt for the purpose of making larger electrodes, controlling the solidification structure, etc.). The melt type or combination of melt types selected depends on both the alloy composition and intended application (mill form and size desired, properties desired, sensitivity of the final component to localized inhomogeneity in the alloy, economics, etc.).

Although not widely used owing to economic considerations, the use of electron beams or plasma playing on the surface of an electrode and on the molten pool provides an alternate source for clean melting of some specialty alloys.

37.9.2 Casting to Prepare for Subsequent Processing

When alloys are cast, the resulting object may be (a) an electrode for remelting, (b) an ingot to be melted (in whole or in part) later and then cast into articles, (c) a cast article of commerce, or (d) an ingot for processing into billet. This section discusses the casting of ingot for later use, not directly into articles of commerce.

From the billet, wrought-shaped articles may be formed. Also from billet stock, mill products may be produced. Mill products would be bar, plate, sheet, wire, and so on, which might later be deformation processed into a shape or used directly. For example, bar stock might be machined into fasteners while wire might be sold and used with no further processing. Alloys usually not only are available in standard mill forms (plate, sheet, and bar, from which components may be machined) but also may be produced as specific component shapes by the use of forging (or casting). It is critically necessary that the property requirements of a component be fully understood to maximize properties and minimize costs. For instance, a component selected to resist corrosive attack may not need to be melted by a process that assures maximum fatigue resistance.

The most critical feature of melt process selection is the size of the component to be manufactured from an ingot. Larger components require larger ingot sizes. The higher the melting point of an alloy and the more highly alloyed the chemistry of the alloy, the greater may be the restrictions of maximum ingot size. For forged components, consideration must be made not only of the final component size but also of the capability of the ingot to be sufficiently deformed so as to develop the properties characteristic of a forged structure. Internal ingot cleanliness is also a consideration in selecting the melting route.

Many alloys will have no special cyclic requirements and routine melting and casting to ingots is the rule. However, for more sensitive applications (e.g., aircraft structures), more stringent requirements will exist. Generally, for fatigue-sensitive components, a melt practice is selected which will guarantee the best structure. The needs of sensitive applications may require the generation of specific structure in an ingot cast and machined to billet. For example, a specific grain size may be required by the wrought processor to obtain the ultimate customer's specified properties after a forging process and subsequent heat treatment. Specific grain sizes and/or ingot surface finish may be required to maximize the effectiveness of ultrasonic or etch inspections. The processes for producing ingot for sensitive wrought-product applications will require that the alloy selector work closely with the melter and forger on the melt-to-ingot-to-billet-to-article sequence. Those selecting materials for less critical articles may be able to rely on standard "mill" practices to generate adequate product.

37.9.3 Casting Practices for Producing Articles

Generally, cast articles are made by remelting ingots of the desired alloy. Articles may range from grams (ounces) in weight for jewelry (e.g., rings) to tens or hundreds of pounds (machinery, vehicle applications, power turbine components, and aircraft gas turbine components). The casting invariably is done in a foundry. The alloy generally is melted in smaller versions of the furnaces used to do primary/secondary melting to produce ingots. Melting in the furnace is done in crucibles, which are almost always ceramic-lined containers. To cast an article after it has been designed, a copy of the design must be replicated as a negative cavity in a mold. Obviously, the mold must be resistant to attack by the molten alloy to be poured into it.

Multiple molds generally are prepared and dependent on the weight/size of the intended cast article. The molds must be connected in such a way that transfer of liquid alloy to them is easy. Design of molds is critical to successful casting. The cavity in the mold must be slightly larger than the intended article to provide for the shrinkage that occurs in the solidification of alloys. There will be a basin or tundish into which the molten metal is poured and from which the metal is dispersed to the individual molds. To transfer the metal and properly feed the cavities, various channels (sprues, risers, runners, etc.) are needed. The final mold assembly must have sufficient rigidity to be stored and to be moved into place for casting at the appropriate time.

Different media are used for molds. Sand is a common and an inexpensive molding material and plaster and other similar compositions may be used. Surface finish and dimensions are affected by the molding material as is the cost of the casting process. In some instances, molten metal can be melted and injected into water-cooled metal dies producing fine definition with die-cast articles. The alloys used for such processing are lower melting alloys but work has been done to cast high-temperature alloys this way. Cost will be a factor and die casting is more commonly a process for large runs of small-to moderate-size articles such as medals cast in moderate-melting-range alloys.

Complex shapes can be produced by casting. Often, the article shape is such that cores must be added to the negative cavity to produce desired passages and holes in the finished casting. Most routine casting will not require extra thin article walls or extra high tolerances. However, for aircraft or power gas turbine parts, in particular, very high tolerances are required and wall thicknesses (nickel-base superalloys) may drop to 0.040 in. (0.101 cm) or so.

The principal casting practice for high-precision casting is investment casting (also known as the lost-wax process). A reverse-cast model of the desired component is made and wax is solidified in the resultant die. Then, a series of these wax models are joined to a central wax-pouring stem. The assembly is coated (invested) with appropriate ceramic, processed to remove wax, and fired to strengthen the invested ceramic mold. A small percentage of a (VIM) alloy ingot is remelted and cast into the mold. Upon solidification, a series of components in the desired form are created, attached to the central pouring stem. These objects, frequently gas turbine hot-section airfoil components, are removed and then machined and processed to desired dimensions. Superalloy investment-cast cases now are available up to about 40 in. (101.6 cm) in diameter. Titanium alloys similarly can be cast in large sizes using investment-casting techniques. Turbine airfoils can now be cast not only in smaller sizes for aircraft gas turbines but also with airfoil lengths of several feet for land-based power turbines.

Most investment-cast articles are of conventional, polycrystalline structure with more or less randomly oriented grains. Mechanical properties are nominally random but may show some directionality. Increased high-temperature property strength levels have been achieved in nickel-base superalloys by columnar grain directional solidification (CGDS), which removes grain boundaries that are perpendicular to the applied principal load in turbine airfoils. The ultimate solution for optimum high-temperature strength is single-crystal directional solidification (SCDS) to produce a superalloy with no grain boundaries. Maximum creep rupture strength in nickel-base superalloys now is achieved with SCDS alloys. SCDS has been applied not only to aircraft gas turbine engines but also to large-frame land-based gas turbines. Most alloy systems used commercially have not shown a need for increases in the high-temperature strength, which might be produced by single-crystal casting. Such processing would considerably increase cost and property improvements with SCDS depend on alloy chemistry. Certain nickel-base superalloys are the only alloys where acceptable cost-benefit trades are possible.

37.9.4 Casting Considerations in Alloy Selection

Alloy selection for cast articles generally will be on the basis of alloy type needed (e.g., copper and aluminum) and then on specific alloys within the selected type so as to provide the desired properties. Not all alloys are equally good for casting. In many alloys systems, compositions of casting alloys differ from those of alloys used for wrought applications. Depending on the application, there may be more or less tolerance for lesser quality castings. Need for pressure tightness, for example, might dictate a casting process different from that for another application. Casting defects occur and vary with alloy composition, article shape, and dimensions, as well as casting practice. Oxide (dross) carried over from the alloy-melting production process may cause surface and internal inclusion problems in some alloy systems. Porosity is another major concern, especially with large castings. Porosity may be controlled by mold design and pouring modifications. In some alloy systems where articles for critical applications are required, non-surface-connected porosity may be closed by hot isostatic pressing (HIP) of the cast articles. Surface porosity in large castings may be repaired by welding. Other casting concerns may include surface finish, intergranular attack (IGA) caused during chemical removal of molding materials, selectively located coarse grains in PC materials, and so on. Sometimes, alloys may be modified in chemistry to optimize product yield but with a possible compromise in properties. Product yield is an important determinant of final component cost.

37.10 FORGING, FORMING, POWDER METALLURGY, AND JOINING OF ALLOYS

37.10.1 Forging and Forming

Forging is one of the most common methods of producing modest production lots of wrought articles for alloy applications. Mill products such as bar stock and wire generally are produced in greater quantities than forgings but, for the most part, are used to make non-complex shapes. The most demanding applications use wrought forged ingot metallurgy components. One or more intermediate shape stages usually are involved when conventional forging is practiced. Special processing by isothermal (usually superplastic) forging may enable the forging to go directly from billet to final stage in one step using a closed die.

A billet is the precursor to a forging in large-sized articles. Billets are transformed (forged) to shapes approximating the final desired shapes with the approximation dimensions being dependent on the alloy system and known issues in the forging process. Multiple reductions may be needed and blocker dies (open dies) may be needed before closed die forging to final dimension/shape. Forging may be accomplished with one or a group of dies and in open or closed dies. While closed dies better define the finish form of a forging, the cost of using such dies exceeds that of open dies. Some dies may be just flat platens that are used to transfer the forging load. The more complex the dies, the more difficult it becomes to extract heat. Significant costs can be incurred for die design and procurement.

Forging is usually a process carried out at elevated temperatures with the temperature being determined by alloy composition, amount of reduction needed, and so on. Forging always requires significant deformation of the alloy as it is transformed from billet to article. Sometimes, the forging deformation is retained to increase strength; sometimes forging is used to modify the microstructure of the article when forging is combined with appropriate heat treatments. In the case of high-performance superalloys, the objectives of a forging cycle may be:

- uniform grain refinement
- control of second-phase morphology
- controlled grain flow
- structurally sound component

Hot forging is a multistep process with several to many cycles of heating to forge temperature, then forging, followed again by heating, forging, and so on.

Forming, while containing elements of deformation, generally restricts itself mostly to sheet that, for example, may be bent or formed to a mandrel or other forming die. Forming of automotive hoods from sheet metal would be an example. Deformation induced in forming may not be extreme and forming is not used to control microstructure and properties of alloys. Basically, forming produces relatively two-dimensional shapes while forging produces three-dimensional shapes with distinct properties. Roll forming is often used to produce small shapes from bar stock or rod.

There are two major types of forging: open die and closed die. Open-die forgings are less expensive but in general require more stock removal by machining to get to the finish dimensions. Closed-die forgings are more costly but reduce the machining needed to achieve finish dimensions in the article. In forging, alloy billets are squeezed between

two platens or dies using an appropriately sized press. Several to many forging steps may be needed to bring the billet to the approximate article shape and dimensions. A great deal of skill is necessary not only in die design but by the operators of the forge presses in order to achieve optimum properties and cost balances. As alloy strengths increase, it becomes increasingly difficult to move an alloy around during forging. When higher forging pressures are required, defects become more probable.

Superplastic forming became available in the 1960s and was first used to produce aluminum alloy door panels, for example, for high-end automotive vehicles. Superplastic forging was introduced for super alloys when the strengths of alloys were bumping up against metallurgical limits and forging was becoming too difficult. A process called superplastic forming/diffusion bonding was developed in the air frame business to produce large articles with better properties than existing technology could produce. This is a high-technology process without applicability to processing routine materials.

37.10.2 Powder Metallurgy Processing

PM has been in use for many decades, dating back over 50 years. The principal intent of PM was/is to produce alloy articles at less cost. However, since the latter part of the twentieth century, PM also has meant a way to create billets of the highest strength alloys (without casting) and subsequent forging of billets to final parts. Different powder production manufacturing processes exist for alloys. A common practice is gas atomization where a prealloyed ingot is remelted and then blasted by a jet of air/argon/helium to produce powder in a chamber. The powder is recovered and classified for size, for example. Powder production generally is a batch process.

Molds (dies) for making PM parts are made to the shape of the planned article and then filled with powder, tapped, pressed, and sintered (high-temperature heated) or otherwise metallurgically bonded (by heat/pressure) to increase density and properties. The process often can be automated. Dimensions of dies will need to reflect the expansion coefficient of the die material and the alloy powder. There are different process variables in PM, including multiple methods of powder production. Compaction consists of processes such as cold pressing or vacuum hot pressing or HIP, among others. Some high-strength superalloys are made only by PM methods. Because powder articles made by other than HIP are not 100% dense, there will be porosity in the form of very small voids in the powder article compact. For the most part, a small number of defects such as holes do not significantly affect the properties of alloys, at least in static tests. For fatigue-limited applications, higher quality production processes would need to be used.

A relatively new aspect of the use of PM is in the superalloy field where alloys, which can be cast in relatively small parts such as turbine blades, are not workable by cogging, forging, and so on, to produce a billet for disk forging. By using powder of the desired alloy, a compact of the desired article can be made by HIP or extrusion and then the compact can subsequently be worked (usually forged) to final shape and machined to produce a higher strength wrought article than can be produced from conventional wrought-alloy (“ingot metallurgy”) procedures.

37.10.3 Forging/Working Considerations in Alloy Selection

The stronger the alloy to be utilized in design, the more difficult it will be to manipulate by mechanical working forces. In addition to working difficulties, problems with ordinary

forgings can arise from many sources. Poor grain size control, grain size-banded areas, poor second-phase morphology/distribution, internal cracking, and surface cracking are among the sources for rejection of forged parts. Limits exist on the capability of an alloy to be worked without cracking, encountering other defects, or stalling a forge or extrusion press. Some shaping equipment such as rotary forging devices may be better able to change the shape of the stronger alloys than other equipment. PM offers an alternate processing route for high-strength alloys such as superalloys used for rotating disks, shafts, hubs, and so on, in aircraft and power generation turbines. Powder production reflects an art that is not always directly transferable from one producer or one alloy to another. Powder components are best made by producing an ingot from powder and then forging the ingot to component shape. Extensive work with an alloy melt shop and an associated powder producer may be necessary to create a satisfactory metal powder of a high-strength/high-performance alloy selected for design.

37.10.4 Joining

Welding, soldering, and brazing are used to manufacture articles in many instances. Soldering is a lower temperature, less than 800°F (427°C), process of letting molten metal be drawn into a very tight crevice between the components to be joined and then allowing it to be solidified. Brazing is accomplished in the same general way, but the braze alloys melt above about 800°F (427°C). Braze alloys produce stronger bonds than solder alloys. Solder finds extensive use in the electronics industry. Brazing is more often found in structures that require relatively high strength. Such structures might be aircraft gas turbine stator assembly parts, for example.

Welding produces a metallurgical bond between components of an article. There are many welding processes. The simplest definition is solid-state bonding versus fusion welding. Relative to fusion welding, resistance welding techniques are low-cost methods to join parts such as tabs on sheet metal. Arc welding is perhaps the most common fusion welding technique with higher quality associated with gas–tungsten arc or gas–metal arc welding. There are other techniques used for massive structures and also specialized techniques that produce maximum-quality welds. Those techniques use electron beams, laser beams, or plasmas to heat the weld metal. Relative to solid-state welding, a number of techniques exist, including friction or inertia bonding, diffusion bonding, and ultrasonic bonding. The success of these techniques is dependent on clean surfaces and optimum bonding areas.

Many complex articles in devices such as gas or steam turbines were/are combinations of components joined to produce a larger article. In some applications, when the joined article is not too large or complex, efforts have been made to replace such articles with cast components. However, surface-connected defects often are found in large superalloy castings. In addition to large superalloy castings, many articles produced in industry can have production defects. If defects can be repaired, articles may be acceptable for service. Consequently, repair welding is used with great frequency, dependent on the dollar value of the article and other economic and availability aspects.

37.10.5 Considerations in Joining Process Selection

An alloy selector will need to consider the weldability or joinability of an alloy before suggesting initial production use of joining techniques and/or their use in repair. Many

nickel-base superalloys are not weldable. High-performance alloys such as titanium alloys and sheet metal of cobalt-base, solid-solution-strengthened nickel-base and lower hardener content nickel-base alloys can be joined as can iron-nickel-base alloys. Aluminum alloys can be welded with high-heat input welding techniques. Some techniques such as electron beam welding (EBW) may be best for high-performance materials owing to the vacuum environments of the processes and the very narrow fusion zone that is created in the process. High-vacuum EBW and similar joining processes may not find wide use in production of large quantities of consumer goods owing to the cost of the process. On the other hand, the use of nonvacuum electron beam welding (NVEBW) has been shown to be cost effective for steels used in auto frames.

37.11 SURFACE PROTECTION OF MATERIALS

37.11.1 Intrinsic Corrosion Resistance

Most alloys are not intrinsically resistant to surface attack by corrosion. Precious metal alloys containing gold or platinum may be chemically stable. Most alloys consist of metallic elements that want to revert to compounds (e.g., oxides) in chemical states such as those in which the elements are found naturally. Corrosion is the interaction of elements in an alloy with the environment to produce chemical compounds. Surface attack occurs most frequently as the surface is the primary interface with the surroundings of an article in use. Most typically, articles “see” oxygen from the air and a water-based environment owing to humidity in the air. The processes of uniform corrosion in aqueous environments and with selected other environments are fairly well understood. Corrosion also can take place in nonuniform ways through the action of or development of pits, crevices, or cracks. Under some circumstances, individual alloy elements may be selectively removed from an alloy.

Using an understanding of the nature of the corrosion process, metallurgists have developed alloys that have improved resistance to corrosion in a variety of environments. In fact, many alloys containing certain elements such as chromium or aluminum demonstrate natural resistance to corrosion in oxidizing environments. However, alloys that are corrosion resistant in oxidizing environments often are not resistant in reducing environments, especially when attacked in the presence of halide elements such as chlorine. Some alloys/elements are more noble with respect to other alloys/elements. One representation of the corrosion sensitivity of alloys is the “galvanic series of alloys/elements.” Alloys/metals that do not corrode or are more corrosion resistant in aqueous environments are called noble or cathodic while more corrosion prone alloys/elements are called active or anodic. Table 37.11 shows a galvanic series for some commercial alloys and metals in seawater. Note that some more common alloys other than those of the precious elements are near the top. Specifically, titanium and its alloys are very resistant to aqueous corrosion. Similarly, titanium alloys display excellent corrosion resistance in body fluids. Consequently, titanium alloys are often choices for applications such as prostheses for bone implants or in dental work. Commercially, pure titanium finds extensive use in the chemical processing industry.

Although cast iron and steel are near the bottom of the series and so are active (steels rust in most instances of environmental exposure), chromium additions to iron can produce a passive layer that inhibits oxidation. Hence, alloys of iron with sufficient chromium additions became “stainless.” Stainless steels are near the top (cathodic end) of the

galvanic series and perform very well in a variety of commercial applications. Some construction steels are alloyed to produce a rust-resistant finish and are used for uncoated/unpainted steel work on bridges, guard rails, and so on.

Different alloys may display varying behavior dependent on the environment. In some instances, alloys that display good aqueous corrosion resistance (general corrosion attack resistance) show tendencies to selective attack such as that which occurs in stress corrosion. Connections between anodic and cathodic alloys in service invariably lead to accelerated attack of the anodic element/alloy.

Some alloys have moderate resistance to corrosion. That is, they form discolored surface layers but do not have surface recession at a great rate. A case in point would be copper or copper alloy flashing often seen on roofs at chimney lines. The alloys degrade owing to formation of a surface patina, but there is no significant surface recession and the copper retains its cross section and hence strength integrity for a long time.

Some alloys show adequate corrosion resistance so long as the surface of the article is not degraded. However, surface erosion or inadvertent scratching or similar degradation can cause corrosion to accelerate. Pitting and similar corrosion owe their existence to such tendencies.

TABLE 37.11 Galvanic Series of Selected Commercial Metals and Alloys in Seawater^a

Noble or Cathodic	Platinum
	Gold
	Graphite
	Titanium
	Silver
	Stainless steel, austenitic, P (18%Cr, 8%Ni, low C)
	Stainless steel, ferritic, P (10–30% Cr, high C)
	Nickel–chromium–iron alloy, P (80% Ni, 13% Cr, 7% Fe)
	Nickel, P
	Silver solder
	Nickel–copper alloy (70% Ni, 30% Cu)
	Copper–nickel alloys (60–90% Cu)
	Copper–tin bronzes
	Copper–zinc brasses
	Nickel–chromium–iron alloy, A (80% Ni, 13% Cr, 7% Fe)
	Nickel, A
	Tin
	Lead
	Lead–tin solders
	Stainless steel, austenitic, A (18% Cr, 8% Ni, low C)
	Stainless steel, ferritic, A (10–30% Cr, high C)
	Cast iron
	Steel
	Aluminum alloys, precipitation hardened
	Cadmium
	Aluminum, commercially pure
	Zinc
Active, or anodic	Magnesium and magnesium alloys

^aP indicates passive condition; A indicates active condition.

Some alloy elements confer better corrosion resistance in certain temperature/environment regimes than in others. Nickel-base superalloys oxidize at elevated temperatures. Nickel was alloyed with chromium in the early twentieth century to produce oxidation-resistant materials used for electrical resistance wires. In the mid-twentieth century, the nickel–chromium alloys were adapted for use as superalloys for gas turbine engines. In the early years of use, these chromium-containing alloys showed excellent oxidation resistance. Later, as temperatures went higher, alloys with higher amounts of aluminum (and reduced chromium) than in the early superalloys showed improved high-temperature oxidation resistance.

Generally, as temperatures increase, corrosion attack increases. However, for some forms of environmental attack, the kinetics of the attack process may be such that a specific type of attack begins at a certain temperature and increases with increased temperature but then drops again to lower rates at still higher temperatures. Nickel-base superalloys exposed to sulfur and/or halide-containing environments may show lower temperature hot-corrosion rates that are greater than the hot-corrosion attack seen at higher temperatures.

In addition to general corrosion, grain boundary oxidation/corrosion or the selective attack of some secondary phases in an alloy can create notches. Coatings will help protect against such attack. Other types of chemical attack are more subtle. For example, brass (copper–zinc) alloys can be susceptible to stress corrosion cracking. In this instance, the attack may be on/in the grains and cause cracking owing to a residual or applied stress on an article made of the alloy. In the early days of manufacture of cartridge brass (70 Cu and 30 Zn), cases would crack or be embrittled by stress corrosion cracking and would burst on firing. Stress corrosion cracking can be found in titanium alloys and in some stainless steels among other materials. Reduced residual stresses help to alleviate the problem for brass.

37.11.2 Coatings for Protection

From early use of steel, the concept of rust was prevalent. At first, no alloys were made that were rust resistant, but it was discovered that, if one could protect the iron or low-carbon steels from the environment, then there was no attack. In the automotive industry, this concept was applied to bumpers by electroplating coatings onto the bumpers. First, a nickel plate was applied, then a high-reflectivity chromium was plated over the nickel. So long as the nickel did not develop a pore or scratch or suffer a similar breakdown, the steel bumper did not rust. Similarly, before the widespread use of polymer trash cans, steel cans were coated with a sacrificial layer of zinc, an element anodic to iron. The trash cans did not rust because the zinc corroded preferentially to the steel. These two illustrations show the use of impervious coatings or anodic coatings to protect materials.

Corrosion at lower temperatures is not the same as corrosion at high temperatures where the interaction of an alloy with oxygen can cause an oxide to form with consequent reduction in the cross-sectional area of an alloy. If the oxide itself is essentially protective, for example, chromium oxide at temperatures below about 1500–1600°F (186–871°C) or aluminum oxide above about 1800°F (982°C), then the alloy is protected from oxygen attack for some time. It was reasoned that the production of a deliberately introduced chromium-oxide-forming or aluminum-oxide-forming coating on alloys for high-temperature use would increase their resistance to oxidation.

Such coatings were introduced and applied by a number of methods. The early coatings were diffusion coatings on cobalt- or nickel-base superalloys and were created by pack aluminizing or slurry application. The chemistry of the coating was determined by

the chemistry of the alloy. Some high-temperature protective aluminide coatings were applied by painting or by spray processes. Later, (second-generation) coatings were produced by overlaying a specific chemistry of a protective nature on the surface of the component using physical vapor deposition (evaporation of elements of an alloy by using an electron beam and redeposition of the elements on the target alloy's surface). Overlay coatings are generally more expensive than diffusion coatings. However, vapor-transported diffusion-type coatings can coat internal (non-line-of-sight) surfaces while overlay coatings can only coat external line-of-sight surfaces that can be seen by the coating apparatus. Some commercial diffusion coating processes are available, but most overlay coating processes are proprietary, having been developed by users such as aircraft gas turbine manufacturers.

Of course, for lower temperature use, spray or electrodeposited coatings may provide adequate protection for many alloys and overlay coatings might be overkill.

Despite the coatings on high-temperature superalloys, oxidation (or other corrosion) continues but at a markedly lower rate. Since the coating is being used up over time, eventually the coating will not be sufficiently protective and the surface of the alloy must be recoated. This concept of eventual coating degradation will apply at both high temperatures in oxidation and lower temperature corrosion (e.g., aqueous corrosion). Depending on the alloy's structural application, the cost to replace the article that is coated, and the cost of recoating, a decision will need to be made about continued use of the article in its application. Generally, replating or recoating of articles is more apt to occur if the cost of the article's replacement is very high. So, gas turbine airfoils which may cost in the \$100–\$1000 range to make as castings are more apt to be recoated and reused than a car bumper that has rusted.

37.11.3 Coating Selection

Coating selection is based on knowledge of oxidation/corrosion behavior in laboratory, pilot-plant, and field tests. Attributes that are required for successful coating selection include:

- high resistance to general oxidation/corrosion,
- ductility sufficient to provide adequate resistance to thermal mechanical fatigue if the article sees high temperatures since coatings are not particularly ductile and coating cracking/spalling could occur,
- compatibility with the base alloy to be coated (important anywhere but especially for high-temperature operation),
- low rate of interdiffusion with the base alloy (if used at high temperatures),
- ease of application and low cost relative to improvement in component life,
- ability to be stripped and reapplied without significant reduction of base-metal dimensions or degradation of base-metal properties.

37.12 POSTSERVICE REFURBISHMENT AND REPAIR

One important aspect of modern alloy use is the concern for maximizing service life of articles. Manufacturers do not wish articles or devices to fail in less than the warranted life. Selection of alloys for design should include concern for potential early removal and repair or refurbishment. Surface degradation and mechanical property loss are major

economic factors in applications of articles. These factors have become of greater interest as the base cost of materials and subsequent components has risen dramatically over the past 40 years. (Note: As this chapter is being written, for example, the price of nickel, a common base material or alloy element, has risen 10-fold in 3 years!)

Although initial cost has usually prevailed in alloy selection for consumer devices and articles, as the cost of machinery (nonelectronic) has increased owing to various factors, more attention in alloy selection has been given not only to initial ability to produce a viable article but also with regard to future life and refurbishment/repair concerns. In practice, when public safety consideration limits or product integrity design limits are reached, costly components may be withdrawn from use. Some components may appear to be unaffected by service time. For economic reasons, there may be incentives to return these components to service.

Other components after service may have visible changes in appearance. For example, a gas turbine high-pressure turbine blade may be missing a coating in the hottest regions of the airfoil or a crack may develop in a vane airfoil. Seals may be worn. It is highly desirable that damage can be repaired so that the costly parts can be returned to service.

If possible, alloys would be selected on the basis of restoration of capability after initial capability has (apparently) been reduced by service exposure of the component. However, most applications do not permit alloy selection on that basis. Many applications have limited lifetimes to reduce initial cost (and, possibly, to improve the likelihood that new products will be purchased). The best alloy from a property and initial economic viewpoint is usually the choice. However, it is a common practice with certain applications to refurbish or repair many components, which have visible external changes.

Stripping and recoating of turbine airfoils is one example of refurbishment and repair practice. Oxidation- and corrosion-resistant coatings and thermal barrier coatings may be reapplied (after appropriate surface-cleaning treatments) to restore resistance of the surface to heat and gaseous environments. When a high-performance article made of a superalloy is to be refurbished by recoating, all traces of the original coating should be removed before recoating is attempted. In case of missing, eroded, cracked, and/or routed material, welding traditionally has been used to fill the gaps for some alloys and components. Care must be taken to assure that no additional alloy degradation occurs owing to the refurbishment and repair practice.

The restoration of mechanical properties degraded by creep and/or fatigue is not clear-cut. In the laboratory, reheat treatment has been shown to restore the mechanical properties of some superalloys after service exposure. The degree of restoration is a function of the mechanical history of the component. Results of reheat treatment of service-exposed parts are variable. Most postservice procedures for high-cost flight safety articles do not provide for mechanical property restoration.

It is important to recognize that refurbishment or repair may not result in cost-effective performance.

37.13 ALLOY SELECTION: A LOOK AT POSSIBILITIES

37.13.1 General

Selection of alloys for design may require a comprehensive review of the design and potential materials or, as noted below, may rely on prior use to dictate alloy selection. On

the assumption that a design will start from scratch with no preconceived notions of alloy usage, several questions need to be answered:

- What is the expected temperature of use?
- What is the expected environment (gas, liquid, moving, static, etc.)?
- What strength levels are required?
- Is there a cyclic component to any loads?
- Are elastic properties likely to be a criterion?
- Will loading be uniform or will some directions within the article see different loads?
- Are there special characteristics at issue, for example, heat capacity, electrical, magnetic?
- For how long must the article to be designed and manufactured last?
- Will successful operation depend on special cooling or other conditions?
- What weight or dimensional levels must not be exceeded?
- Is there an aesthetic component to the application of the desired article?

Assuming that the above questions can be answered in a definitive way, referral to general alloy characteristics will narrow the choices available. For example, operating temperatures above about 1000°F (538°C) will eliminate aluminum, magnesium, and zinc alloys as alloys of those metals will be molten at such operating temperatures. Elastic modulus concerns might further limit alloys. For example, if it would appear that materials for a given article would need Young's modulus of near 30×10^6 psi (206.8 GPa), then titanium alloys would be eliminated as their moduli are less than 20×10^6 psi (137.9 GPa). Would the application be in oxidizing gases above 1000°F (538°C)? Then iron would be ruled out owing to its lower oxidation resistance compared to other alloys at those temperatures. Stainless steels might be possible but creep rupture strength might be too low, thus moving the likely candidates to superalloys, those of cobalt, iron–nickel, or nickel base.

Now, attention would need to be directed to the manufacturing ability of alloys to be considered. If the alloy is to be mechanically deformed to reach its near final shape, ductility/workability will need to be considered. Is the article complex, with internal cooling passages and thin walls? Then investment casting may be the best production method! Consideration must also be given to the fact that some processing methods produce or permit improved property levels over other processing methods. For example, investment-cast superalloys are superior in creep rupture strength to wrought superalloys and SCDS investment-cast superalloys are superior to polycrystal investment-cast superalloys.

At this point, alloy families will have been reduced considerably from the total matrix of alloys to one or two alloy types, for example, cobalt-base or nickel-base superalloys that can be investment cast. Design now will hinge on strength (static and/or dynamic-cyclic). A review of strength requirements might show creep rupture capability requirements on the order of 24-h life at 1850°F/36 ksi (1010°C/248.2 MPa). Cobalt alloys cannot reach those levels so nickel-base superalloys become the choice. Now the selection may hinge on specific alloy chemistry, past experience with an alloy, and investment-casting experience with the alloys from which selection needs to be made. Economics of the basic alloy chemistry, economics of the investment-casting route (polycrystalline,

columnar grain, and single crystal) to be used, ease of applying surface coatings (if required), minimum alloy properties which are needed, and perhaps other factors need to be considered.

The final selection will need to be verified by actual manufacturing and device testing in simulated or (preferably) actual service. Refurbishment procedures may need to be considered if the article is particularly costly and can have its life extended by suitable and timely refurbishment.

It may be seen that the alloy selection in this instance is not exactly a case of working from “ground zero” to scientifically evaluate the likely candidates. Rather, the selection has depended on real engineering data and judgment and would have involved potential alloy suppliers and manufacturing specialists as soon as possible. A great deal of “alloy selection” that appears in the literature has been done in retrospect, that is after some particular article has been manufactured (e.g., a skateboard), scientific approaches to the evaluation of the actual article show the requirements which were met to produce it. There is nothing in principle wrong with scientific alloy selection. However, the general lack of data (not enough, maybe virtually no, property data), no or limited manufacturing experience, insufficient awareness of the “shop” climate (what furnaces might be available for heat treating the articles and how would the treatment impact the shop schedule), and so on, can cause unacceptable delays or even inability to manufacture an acceptable article.

Assuming that minimum alloy property data will be available, the key to alloy selection is to find knowledgeable engineers in-house and engage them and other knowledgeable persons from the manufacturing pipeline in conversation early on in the selection process.

37.13.2 Impact of Materials’ Data Validity on Selection

There ought to be one basic rule for alloy selection and that is: “Do not believe all you see or hear” (from your fellow employees or from persons purveying new alloys, coatings, or processes). The most common type of statement from alloy/process developers with whom one talks (the “sellers”) at the start of new alloy (or process) evaluation is: “This is our average product/property.” Translated, this statement should probably be treated by the alloy selectors (the “buyers” in this instance) as meaning: “It’s the best we have gotten!” One needs to be skeptical and inquisitive about statements from the seller!

One should be especially wary of selective use of data. There was a situation where an alloy (never used in production) was claimed to have property levels up to 150°F (94°C) better than any other similar alloy then existent. Unfortunately, when a laboratory program was run at a different laboratory (lab 2) than the laboratory (lab 1) of alloy invention, results were troublingly low. On inquiry to the “inventor of the alloy,” the question was raised as to how he got such good results. The answer? The inventor only cut out good material to test.

At that point, it was reported to the manager of lab 2 that, with careful work in the laboratory, one might get a “100°F (62.5°C)” improvement over the best existing alloy but never 150°F (94°C). And, it was said that, in a production shop, the best improvement might be no better than 50°F (31°C). The latter projection was proven true when maximum alloy improvement turned out in production to be only 47°F (29.4°C).

Engineers have been known to discard data that are not as good as what they expect (and occasionally get). There need to be valid reasons for excluding data. When much data (properties, results of processing, etc.) get discarded, the discarded data often come

back to haunt the user of the alloy/production process. Dig into the database for anything you use in your alloy selection. Assure yourself that there are no hidden items that may appear again in production.

In an instance of mechanical property tests of a new wrought nickel-base superalloy to be used for a gas turbine disk, low creep rupture test results were routinely discarded. The alloy (as a disk) went into production and most of the first production run began to fail the creep rupture acceptance tests in the shop, with test values almost identical to those which had been routinely discarded during alloy development. A modification of heat treatment was necessary to bring the alloy creep rupture life back to the range promised in development and much time and money were spent in the recovery operation.

37.14 LEVEL OF PROPERTY DATA

Different organizations have different ways to treat property data. Generally minimum and typical properties are desired by design engineers. Minima can vary in definition (-2σ , -3σ deviations are used). One of the cardinal rules of alloy property development is that enough data will never be generated to truly determine a statistically valid property level over all possible property space. Data cost money! Estimates are made and may often be used in design. If these estimates are conservative and can be justified by alloy data on the alloy of interest and similar alloys, satisfactory designs with the selected alloy usually result.

Not all properties require minimum values. Some property values used for design are typical values. While minima may be used in design, estimates of typical behavior of a component made to track its operation may require typical property values.

37.15 THOUGHTS ON ALLOY SYSTEMS

37.15.1 General

Melting temperatures for homogenous alloys range near but generally below the melting temperature for the major (base) element. Natural segregation of alloy elements during solidification can lead to even lower melting temperatures. The incipient melting temperatures (lowest melting point of an alloy after solidification) can be many tens of degrees less than an alloy's nominal melting temperature. While some homogenization may occur with heat treatment or with wrought processing and heat treatment, it is possible for alloys, particularly cast alloys, to still show incipient melting. This factor needs to be kept in mind during processing and use of a given alloy. Some alloy elements cause significant depression of the melting temperatures of an alloy system while others show relatively little effect. For example, boron, carbon, zirconium, and hafnium depress the melting temperatures of nickel-base superalloys. As these elements were added primarily to enhance grain boundary strength/ductility in cast versions of many alloys, the introduction of single-crystal nickel-base superalloys enabled most or all of those elements to be removed from the alloy chemistry. The resultant melting temperature increase enabled enhanced heat treatments and optimized strength properties in these alloys.

Another aspect of alloys is the effect of alloy elements on the density of the base metal of an alloy. When much lighter elements than the base metal are introduced, significant density reductions occur. Such additions may or may not be beneficial. For example, early

nickel-base superalloys tended to have densities of about 0.3 lb/in.³ (8.3 g/cm³). Some inventors added vanadium, a relatively light metal, to an alloy and reduced the density of the alloy to about 0.28 lb/in.³ (7.75 g/cm³). When first introduced, the vanadium-containing alloy seemed destined to become an excellent turbine blade airfoil performer. However, the introduction of vanadium made the alloy sensitive to hot-corrosion degradation, and it was dropped from consideration. Major changes in alloy density for an application may often be achieved but often only by selecting a different base alloy.

Alloy costs today are much more variable than they have been historically. Alloy selection needs to consider the possibility of alloy and processing cost increases during the lifetime of a component design. Alloy costs can be significantly impacted by local conditions in areas of the world where the base metal ore is found. Dramatic increases in prices have occurred when wars, embargoes, or significant new manufacturing restrictions are imposed. Sometimes, such cost increases have led to changes in alloy selection for a component or to chemistry modifications to reduce the amount of costly or strategic element used in the alloy. In the mid-1970s, the price of cobalt shot up by a factor of 10 (the price eventually dropped). Cobalt was a principal ingredient in many superalloys used for gas turbines. Corporate managements were appalled by this state of affairs and decreed that cobalt be removed as soon as possible from the manufacturing process. As a result, a particular iron-nickel-base superalloy, IN 718 (alloy with no cobalt), became the world standard for applications which previously used cobalt-containing alloys. It behooves the alloy selector to consider strategic and economic conditions when recommending an alloy for design.

As mentioned earlier, there are limited numbers of alloy systems available for structural applications. Iron is the most common and least expensive. Aluminum, copper, magnesium, and titanium follow iron in volume of structural alloy application. Nickel and superalloys are widely used as well. Alloys from all of the systems mentioned can be welded, although some alloys or systems may require special attention. High-strength, high-temperature nickel-base superalloys generally are not weldable owing to a tendency to cracking, particularly in the heat-affected zone.

37.15.2 Iron

Iron alloys consist of cast iron, wrought iron, low-carbon steel, alloy steel, stainless steel, precipitation-hardening steels, and so on. Iron has been around for several millennia. Steel, the name commonly applied to many iron alloys, is the most widely used metallic material. The range of property levels achievable in iron alloys is quite broad and the general availability of iron makes it a relatively inexpensive alloy base for structural design. Steels are hardenable by solid-solution strengthening, grain size hardening, work hardening, precipitation hardening, and transformation hardening (martensite or bainite formation). Very high short-time strengths, up to over 200 ksi (1379 MPa), can be produced in steels. Stainless steels provide significant corrosion protection in many environments. Many steels are forgeable and formable with much steel sheet going into vehicle production. Automobiles can contain nearly 60% of their weight in steel. Steels are ubiquitous, being found in tall buildings, aircraft, cars, trucks, small and large internal combustion engines, hand tools and large machine tools, garden implements, kitchen utensils, pots, pans, and so on. Iron-base super alloys are used in gas and steam turbines. Much iron wire was used to fence the western United States in the late 1800s and iron boilers, rails, and rail cars were standard on trains. Iron has a density of 0.285 lb/in.³ (7.89 g/cm³) but

alloys can range as high as 0.3 lb/in.^3 (8.3 g/cm^3) or above. Steels have elastic moduli in the range of about $30 \times 10^6 \text{ psi}$ (207 GPa). Iron melts at 1220°F (660°C) but steels melt at lower temperatures. Steels can be plated, painted, or coated to enhance corrosion protection. The alloys can be joined with relative ease depending on alloy composition, and economics favors iron alloys as materials of choice in many instances.

Other than for certain stainless grades, steel normally is considered an alloy of iron and carbon plus other elements. Now some modern steels incorporate very important changes, for example, interstitial-free steels where carbon is considered an impurity and vacuum-degassed steels with extra low oxygen content. The application of carbon and alloy steels is related largely to the ability to get the desired mechanical properties ("hardness") where we want it. The selection of steels will need to consider the hardenability (maximum thickness at which one can achieve the desired hardness/strength) of the selected alloy with respect to the requirements for strength in the article being designed. Combinations of alloy content and processing can almost always enable the strength (hardness) to be achieved. There are more steels from which to choose than any other alloy class. Strength in steels is frequently referred to in terms of hardness (e.g., on a Rockwell C scale or some other scale). There are sophisticated tools available to determine the likelihood of success in getting an alloy with the correct hardenability.

The properties of steels are modified not only by chemistry changes but also by heating for various times at various temperatures. This heat treatment of steels is effective because iron has a phase transition as it heats up and the lower (room temperature) body-centered-cubic ferrite phase changes to the face-centered-cubic austenite phase. Of course, carbides continue to exist in steels as appropriate to the temperatures and times at which the steel has been held. On cooling, the steel can change back to the preferred lower temperature structure in many ways (not all ways are possible with all steels). Essentially the austenite, which absorbs a lot of carbon at its high temperature of formation, rejects the carbon and tries to transform back to ferrite. Sometimes the austenite forms a nonequilibrium body-centered-tetragonal phase, martensite. Martensite then exists along with carbides. Sometimes austenite transforms to bainite and carbides and sometimes austenite transforms to pearlite (a special form of ferrite and carbide aggregate). The result of these various possibilities is the ability of steels to possess a range of possible properties. Selection of a particular steel is an exercise in estimating the strength needs for a particular property and finding one steel from the multitude of available steels that will get the properties where they are wanted, along with any other desired properties such as corrosion resistance, and at the lowest cost.

Plain carbon steels are used for routine and lower temperature applications. For elevated-temperature applications (e.g., tubing/piping in a steam turbine), higher alloy contents are needed and carbon content is only a part of the selection process. Chromium is often added in small amounts to low-alloy steels used at moderate temperatures. With sufficient chromium, scaling of steels can be prevented. When corrosion is meant to be almost entirely prevented, chromium is added above about 12 wt%, sometimes going up above 20 wt% for certain stainless steels.

Stainless steels are intended to promote integrity of the surface while retaining reasonable strength at high temperatures. The chromium oxide film that forms on stainless steels conveys a passivity and provides corrosion resistance to the alloy. Stainless steels not only see high-temperature service but also find much use at lower temperatures in aqueous or other nongaseous environments. To an extent, stainless steels for elevated-temperature service form a continuum with the iron-nickel- and nickel-base super alloys as well as

the heat-resistant and nickel-base corrosion-resistant alloys. The so-called precipitation-hardened steels are chromium-containing steels with special additives to bring out precipitates for hardening, much like aluminum alloys or nickel-base super alloys. It is important to note that the stainless nature of stainless steels is promoted by chemically oxidizing environments. Reducing environments, particularly environments containing halogens, can have devastating effects on stainless steels.

37.15.3 Copper

Copper is a system which also has been around for millennia. The bronze age was the age of tin alloyed with copper to produce strong (by the standards of the day) alloys. The bronze age was followed by the iron age. There are many benefits to copper compared to iron-base alloys and other systems. The copper alloy systems do not compete solely on a strength basis since copper alloys normally are weaker than iron alloys. However, there are areas where copper alloys show sufficient strength and unique other properties (corrosion resistance, nonsparking characteristics when struck, etc.) to compete with steels. A major property of pure or relatively pure copper alloys is their high electrical and thermal conductivity. Specially strengthened copper alloys find use as electrodes and similar current-carrying articles where high temperatures and surface wear resistance are significant. Copper has a moderately high melting point, 1981°F (1083°C); this is less than the melting points of iron alloys but much greater than the melting points of aluminum and magnesium alloys.

Copper is strengthened by solid-solution hardening, grain size hardening, work hardening, and precipitation hardening. Copper is moderately abundant and copper alloys such as the brasses offer reasonable strength with good formability and corrosion resistance. The modulus of copper is 17×10^6 psi (115 GPa); this is less than the modulus of steels and nickel alloys, higher than the modulus of aluminum, and comparable to the moduli of titanium alloys. Most wrought alloys are available in work-hardened conditions and typically find use in moderate to small parts such as springs, fasteners, hardware, gears, and cams. Copper is very formable and forgeable though forming is more apt to be used than forging to manufacture articles of copper alloys. Casting can be used to create copper articles as well. A major use is in conductivity devices (e.g., wire). Copper alloys find wide use in plumbing. Copper, brasses, bronzes, and cupronickels are used for pipes, valves, and fittings in transporting water. The density of copper alloys is around 0.32 lb/in.³ (8.86 g/cm³) for many compositions.

37.15.4 Aluminum

The aluminum industry is relatively new and got its start in the late 1800s. Aluminum alloys became available in the early part of the twentieth century. Aluminum melts at 1220°F (660°C) and owes its popularity to corrosion-resistant, moderately high strength alloys with a substantially lower density, 0.098 lb/in.³ (2.77 g/cm³), than most other structural alloys. Strong aluminum alloys can be produced for operation up to about several hundred degrees Fahrenheit. Aluminum is strengthened by solid solution and work hardening but the highest strength aluminum alloys are precipitation (age) hardened. Cast aluminum is often used. Aluminum alloys have become synonymous with lightweight articles of acceptable strength when age-hardened aluminum is used. The density is considerably less than the density of steel. Figure 37.11 shows the effect of density on the

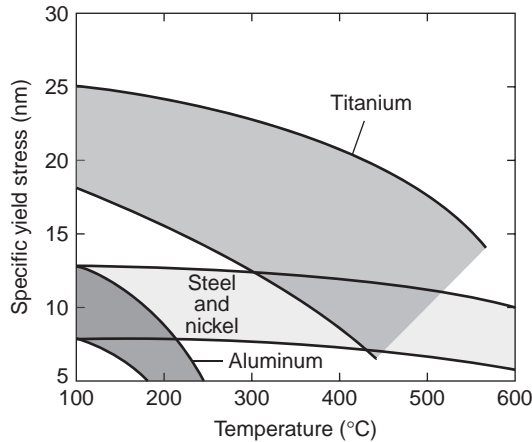


FIGURE 37.11 In certain circumstances, lower densities can make somewhat less strong alloys the materials of choice.

specific strength of several alloy systems. Note that, despite the much greater strength of iron and nickel alloys, at room temperature the specific strength of aluminum alloys makes them competitive with iron alloys.

Aluminum is thermodynamically a reactive metal, as is magnesium. However, it has excellent resistance to corrosion in water (including salt water), atmospheric environments, oils, and many chemicals. Aluminum owes its excellent corrosion resistance to the aluminum oxide films that form naturally but can be enhanced by artificial means. When the excellent corrosion resistance of aluminum alloys in many environments plus the excellent formability/forgability of aluminum alloys is considered, their high specific strength can make them alloys of choice. In weight-sensitive applications at low to moderate temperatures, aluminum alloys perform very well. The modulus of aluminum alloys is about 10×10^6 psi (69 GPa) and so the alloys are not as stiff as iron, copper, or nickel alloys. They find use in aircraft structures, however, and aluminum beams have been used in construction. Aluminum is nontoxic and is found as pots/pans, utensils, and beverage cans and in other uses.

Although aluminum alloys generally are corrosion resistant in atmospheric environments, some alloys are corroded in special localized environments. Exfoliation, intergranular corrosion, stress corrosion cracking, and similar problems can exist. Some alloy types within the aluminum alloy family are less corrosion resistant than other aluminum alloys. However, alloy choice, the availability of anodizing and similar finishing processes, plus the process of cladding aluminum alloys with pure aluminum can provide good corrosion resistance for aluminum alloy systems. Alclad aluminum alloys generally are sheet and tube. Because of the nature of artificially formed aluminum oxide coatings on aluminum alloys, the anodized alloys can be colored. In addition to household goods, aluminum siding provides colorful protection for many homes while resisting denting. Aluminum finds use as ladders and in many outdoor activities where its intrinsic corrosion-resistant oxide surface skin protects it better than iron alloys. Aluminum alloys are extensively used in aircraft structures.

37.15.5 Magnesium

Magnesium alloys became available in the early twentieth century. Magnesium alloy applications are fueled by its density of 0.063 lb/in.³ (1.74 g/cm³), which is the lowest of any commercially used metal. The melting point of magnesium of 1202°F (650°C) is competitive with that of aluminum. The modulus of magnesium is about 6.5×10^6 psi (45 GPa), which is considerably less than aluminum or steel. However, owing to its low density, magnesium can have better specific rigidity than the latter alloys. For rectangular steel, aluminum, and magnesium sections of equal rigidity, calculations show that the magnesium article will weigh only about 71% as much as the aluminum article and 41% as much as the steel article. Magnesium alloys have good strength and very good strength when their density is factored into the selection process. Magnesium alloys generally are forged/formed at elevated temperatures, and forgings can be produced in the same general variety of shapes and sizes as forgings of other metals. However, most magnesium articles are produced as castings. Commercial applications have included luggage, ladders, materials handling equipment, as well as aerospace applications.

Magnesium is the most active element in the galvanic series of elements/alloys and so normally should not be coupled with any more cathodic alloy in an application. (Magnesium is used in hot-water tanks to provide cathodic protection.) Under extreme conditions of galvanic corrosion, rapid corrosion results. However, magnesium alloys generally are used with limited corrosion problems. Magnesium rapidly forms an oxide layer when exposed to corrosion conditions such as atmospheric moisture. The oxide is not too protective but neither is alloy section loss excessive. Use of high-purity magnesium reduces problems and anodizing, chromate coatings, and polymeric coatings have been developed to increase magnesium alloy protection from the environment.

37.15.6 Titanium

Titanium became commercially available at the start of the last half of the twentieth century. Titanium owes its industrial use to two significant factors: it has exceptional room temperature resistance to a variety of corrosive media and a relatively low density of 0.163 lb/in.³ (4.51 g/cm³) and can be strengthened to achieve outstanding properties when compared with competitive materials on a strength-to-density basis. Limited numbers of titanium alloys have been invented and marketed, perhaps owing to the greater cost of the product compared to the metal systems covered thus far. Titanium is produced by a batch process and must be further processed by expensive vacuum-melting techniques if its use is desired for aerospace applications.

There are three main classes of titanium alloys, α , $\alpha-\beta$, and β alloys. Property levels and processing capability vary with class of alloy in addition to specific composition. Titanium has a melting point of 3035°F (1668°C). The high melting point of titanium induced inventors to assume that high-strength alloys comparable to those of iron-, nickel-, and cobalt-base alloys would be capable of being produced. Unfortunately, titanium alloys are not able to operate at very high temperatures. Thus, while titanium and its alloys have melting points higher than those of steels, their maximum upper useful temperatures for structural applications generally range from as low as 800°F (427°C) to the region of about 1000–1100°F (538–595°C), depending on composition. Normally, the alloys are used well below about 1000°F (538°C). However, titanium alloys are capable of being strengthened to high levels and, combined with their reduced densities relative to

iron alloys and excellent corrosion resistance, they were able to replace iron alloys extensively in aircraft gas turbines.

The modulus of titanium is 15.0×10^6 psi (103 GPa). However, titanium undergoes an allotropic transformation on heating and changes from a low-temperature hexagonal-close-packed crystal structure (α) to one that is body-centered cubic (β). On cooling back to lower temperatures, a martensitic reaction can occur or the β can transform to α with varying degrees of fine structure. If a titanium alloy has a β structure at room temperature, the modulus of the alloy will differ from the modulus of an α alloy. The modulus will vary with chemistry and processing. Texturing in rolling processes can further change the effective modulus depending on grain alignment and the amount of β or α present. In addition to the variation of modulus with chemistry and processing, titanium alloy density will vary with the amount and type of alloy elements in a given composition. Many β -forming elements are significantly heavier than titanium while the α -forming element, aluminum, is lighter.

One alloy, Ti-6Al-4V, has been the workhorse of the titanium industry and, for many years, claimed over 50% of the titanium market. This alloy can be used up to the order of 500°F (260°C) to 600°F (316°C). Ti-6Al-4V is an α - β alloy with good all-around properties. Generally, α - β alloys are the most used titanium alloys. However, titanium α alloys have better high-temperature properties and β alloys have superior tensile properties and improved fabricability. Titanium alloys have found use in the chemical processing industry, as golf clubs and eyeglass frames, in dental applications, and as structural implants (e.g., hip replacements) in the human body, in addition to their widespread aerospace use. The alloys have greater high-temperature capability than aluminum alloys and are better in specific strengths than iron alloys, facts which ensure their continued use. A large amount of titanium production is used in aerospace applications. Large fan disks in high-bypass-ratio gas turbines and wing spars for aircraft are made of these alloys.

37.15.7 Nickel

Nickel has been available as an alloy element since the late-nineteenth century. It has exceptional value as an alloy element in steels and other elements as well as being the basis for the high-performance superalloys and heat-resistant alloys used in the aerospace and nuclear industries. Superalloys are covered below. Nickel alloys other than superalloys tend to be more or less relatively pure nickels used in food processing equipment, chemical processing, or some aerospace components. There is another class of nickel alloys which have applications other than those just mentioned—specifically, nickel-copper alloys (trade-names begin with Monel), which are characterized by high strength, good weldability, and excellent corrosion resistance at room temperatures in aqueous conditions. Some of these alloys are age hardened. Monels have applications in pumps, valves, storage tanks, heat exchangers, waveguides, screw machine products, fasteners for nuclear applications, and other uses.

37.15.8 Superalloys

Superalloys came into early use about 1940, with use motivated in large part by military concerns. Stainless steels and cobalt alloys were modified to provide improved creep rupture capability at higher temperatures such as those found in piston engine superchargers. Concurrently, a few nickel-base superalloys such as Inconel X were created and made

available. The terminology “superalloys” did not take hold until the late 1940s when the more widespread use of gas turbine engines prompted a search for improved alloys. Eventually, the addition of more precipitation-creating elements such as aluminum into the basic Nichrome (80 nickel–20 chromium) chemistries plus continued work on modified stainless steels with titanium and other additions led to a class of alloys generally called superalloys. Superalloys are iron-, iron-nickel-, cobalt-, or nickel-based alloys, which have exceptional high-temperature strength properties from about 1000°F (538°C) to about 2000°F (1093°C),

Cast nickel-base superalloys have the best strength in creep rupture at high temperatures above about 1300°F (704°C) while wrought iron–nickel and nickel-base superalloys have the best combination of properties below about 1300°F (704°C). Iron-base superalloys have acceptable properties below about 1000°F (538°C) and are the least expensive superalloys. Cobalt-base superalloys are used for long-time, high-temperature, and somewhat lower stress applications. For highest tensile properties, the nickel- or iron–nickel-base alloys stand out. For best forgeability, iron–nickel-base superalloys such as IN718 are outstanding.

Superalloy articles up to 48 in. (122 cm.) in length are cast as blades for gas turbines by the investment casting process, sometimes using SCDS technology. CGDS blades are made as well. Both types are stronger than polycrystalline castings but decidedly more expensive. Large relatively thin-walled case castings up to nearly 48 in. (122 cm) in diameter are made of IN718. Gas turbine disks up to nearly 36 in. (~1 m) in diameter are forged or PM processed and forged from IN718 and nickel-base alloys. Oxidation/corrosion-resistant coatings are used to protect surfaces of the hottest parts (turbine airfoils) from environmental attack. Superalloys are normally melted with various vacuum-melting techniques using induction or other high-technology furnaces. They may have as many as a dozen or more alloy elements controlled in the alloy and are subject (as is titanium) to extensive surveillance during the melting, casting, and forging operations.

Parts as small as thumb size or as large as gas turbine diffuser cases can be cast in superalloys. Investment-casting techniques produce exceptionally high quality alloy articles with maximum property capability. Most alloys are available to potential customers through normal channels. However, certain manufacturers of gas turbine engine hardware have invented proprietary alloys, and these most likely would not be available for public purchase in the foreseeable future.

37.16 SELECTED ALLOY INFORMATION SOURCES

37.16.1 General

In the selection of alloys for structural applications, selectors may have access to prior-related designs from company records or data from public records that may be applicable to the planned design. Consequently, at least a class of alloys and, possibly, a specific alloy or two may be readily apparent as alloy selection starts. However, more information on alloy types, properties, economics, and availability will most certainly be required eventually. It is beyond the scope of this chapter to provide information on all alloy classes, let alone specific compositions, which might be used for alloy selection within the alloy classes. However, many organizations, technical and sales, exist to promote specific elements and alloys of such elements as well as to promote processes for the manufacture of specific alloys into useful articles.

The number of organizations available to provide alloys, alloy property data, produce articles, and so on, is very large. The following list of websites is provided as a starting point to help alloy selectors determine possible alloy types, properties, economics, and availability. This list almost certainly will change with time and should only be used as a guide to locate potential information providers. Hopefully, the resources at these and similar sites will enable a selector and the designer to come up with an appropriate alloy for a planned design.

It is vital to remember in alloy selection that many alloys are not off-the-shelf consumer items and that properties may vary with material specifications for the alloy. Alloys often are made to user specifications that may vary from user to user and time to time. While a selector may be able to use common readily available alloys, often he or she may not be able to do so. When working with any specialty alloys, diligence in working with the producers will pay dividends in obtaining optimum properties and reducing difficulties for users of alloys. This concept is more costly than buying off the shelf but invariably leads to best quality and more reliable articles in design.

37.16.2 Selected Websites for Alloy Selection

The following is only a partial list. More information as to websites, locations, and so on, may be available. Not all companies or institutions active in any given technology are represented. No recommendation is made or implied by this list.

37.16.2.1 General Information ASM International: www.asminternational.org

The Minerals, Metals & Materials Society of the American Institute of Mining, Metallurgical and Petroleum Engineers: www.tms.org

Institute of Materials, Minerals and Mining: www.iom3.org

37.16.2.2 Specifications/Handbooks/Property Data American Society for Testing and Materials: www.astm.org ASM International: www.asminternational.org

Military property handbooks: <http://projects.battelle.org/mmpds/>

37.16.2.3 Fundamental Approaches to Alloy Selection Granta Design: www.grantadesign.com

37.16.2.4 Specific Elements/Alloys Nickel (Nickel Development Institute): www.nickel-institute.org

Cobalt (Cobalt Development Institute): www.thecdi.com

Chromium (International Chromium Development Institute): www.chromium-assoc.com

Superalloys/special steels (Specialty Steel Industry of North America): www.ssina.com

Copper (Copper Development Association): www.copper.org

Copper (International Copper Association): www.copperinfo.com

Titanium (Titanium Development Association): www.titanium.org

Titanium (Titanium Information Group): www.titaniuminfogroup.co.uk

Aluminum (Aluminum Association): www.aluminum.org

Magnesium (Magnesium Association): www.intlmag.org

Iron/Steel (American Iron & Steel Institute): www.steel.org

Gold (*Gold Bulletin*): www.goldbulletin.org
 Platinum (*Platinum Metals Review*): www.platinum-metalsreview.com
 Silver (Silver Institute): www.silverinstitute.org
 Zinc (American Zinc Association): www.zinc.org
 Tungsten (International Tungsten Institute): www.itia.org.uk
 Rare earth metals (Metall): www.metall.com.cn
 Molybdenum (International Molybdenum Organization): www.imoa.org.uk
 Molybdenum (Climax Molybdenum Corp.): www.climaxmolybdenum.com
 Tantalum/niobium (Tantalum-Niobium International Study Center): www.tanb.org

37.16.2.5 Manufacturing Industry Organizations Casting (American Foundry Society): www.afsinc.org

Forging (Forging Industry Association): www.forging.org
 Forming (Institute for Metal Forming): www.lehigh.edu/~inimf
 Joining (American Welding Society): www.aws.org
 Heat Treating (Heat Treating Society): www.asminternational.org

FURTHER READINGS

Ashby MF, Jones DRH. *Engineering Materials*. Vol. 1, 3rd ed. Oxford: Butterworth Heinemann; 2007.
 Ashby MF, Johnson K. *Materials and Design, the Art and Science of Material Selection in Product Design*. Oxford: Butterworth Heinemann; 2002.
 Ashby MF. *Materials Selection in Mechanical Design*. 3rd ed. Oxford: Butterworth Heinemann; 2005.
 Ashby MF, Shercliff H, Cebon D. *Materials: Engineering, Science, Processing and Design*. Oxford: Butterworth Heinemann; 2007.

Data Related

ASM Metals Handbooks, 10th ed. 20 vols. Covering properties, chemistry, surface effects, metals manufacturing, etc., Materials Park, OH: ASM International, continuously updated, latest updated volumes; 2006.
 Budinsky KG, Budinsky MK. *Engineering Materials, Properties and Selection*. 6th ed. Englewood Cliffs, NJ: Prentice-Hall; 1999.
 Charles JA, Crane FAA, Furness JAG. *Selection and Use of Engineering Materials*. 3rd ed. Oxford: Butterworth Heinemann; 1997.
 Dieter GE. *Engineering Design, a Materials and Processing Approach*. 2nd ed. New York: McGraw-Hill; 1991.
 Farag MM. *Selection of Materials and Manufacturing Processes for Engineering Design*. Englewood Cliffs, NJ: Prentice-Hall; 1989.
 Kutz M. (Ed), *Handbook of Materials Selection*. New York: Wiley; 2002.
 Lewis G. *Selection of Engineering Materials*. Englewood Cliffs, NJ: Prentice-Hall; 1990.

38

MECHANICAL PROPERTIES OF POLYMERS

DANIEL LIU, JACKIE REHKOPF, AND MAUREEN REITMAN

- 38.1 Microstructure and morphology of polymers—amorphous versus crystalline
- 38.2 General stress–strain behavior
- 38.3 Viscoelasticity
- 38.4 Mechanical models of viscoelasticity
- 38.5 Time–temperature dependence
- 38.6 Deformation mechanisms
 - 38.6.1 Microscopic deformations
 - 38.6.2 Shear yielding
- 38.7 Crazing
- 38.8 Fracture
 - 38.8.1 Griffith theory
 - 38.8.2 Fracture mechanics
- 38.9 Modifying mechanical properties
 - 38.9.1 Toughening of polymers
 - 38.9.2 Reinforcement of polymers
- 38.10 Load-bearing applications: creep, fatigue resistance, and high strain rate behavior
 - 38.10.1 Creep and creep rupture
 - 38.10.2 Fatigue resistance
 - 38.10.3 Creep-fatigue interaction
 - 38.10.4 Fatigue resistance of polymers
 - 38.10.5 High strain rate behavior

References

Polymers are widely used in every day consumer products and engineering applications. They can provide design and manufacturing flexibility, as well as desirable combinations of properties. However, successful implementation of any polymer requires a basic understanding of its mechanical behavior and what factors affect that behavior.

Mechanical properties depend not only on the chemical nature of the macromolecules, but also on a number of variables, which include molecular weight, crosslinking, branching, crystallinity, plasticizers, fillers and other additives, orientation, and processing effects. This is similar to how the mechanical properties of some metallic alloys (i.e., magnesium alloys) can depend on variables such as alloying constituents, grain size, and phase distributions. However, the effects of these variables tend to be much greater in polymers than the analogous effects in metallic alloys.

Also, in comparison to metals, the mechanical properties of polymers are dependent on the application temperature and the timeframe over which the polymer is exposed to load. This dependence on temperature and time is due to the viscoelastic nature of polymers, which means polymers behave in manner that has characteristics of both a viscous liquid and an elastic solid.

This overview of the basic principles of polymer mechanical behavior includes a discussion of the microstructure and morphology of polymers, stress–strain behavior, time–temperature relationships, deformation mechanisms, fracture mechanics, polymer toughening, and reinforced polymers (and composites). The behavior of polymers in load-bearing applications is also described, with introductions to creep, fatigue, and high strain rate performance. References to more detailed information on particular topics are provided.

38.1 MICROSTRUCTURE AND MORPHOLOGY OF POLYMERS—AMORPHOUS VERSUS CRYSTALLINE

Figure 38.1 illustrates the relationship between polymer volume and temperature, from the melt temperature through the use temperature. When cooling from the melt temperature, molecular motion slows. If the polymer is branched, crosslinked, or otherwise

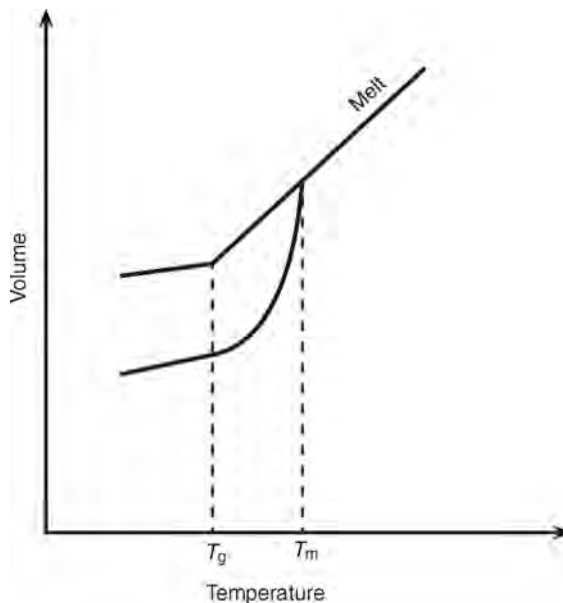


FIGURE 38.1 Specific volume–temperature relationship for polymers. Upper line represents behavior of amorphous material; lower line represents semi-crystalline material behavior.

hindered during this process, it will follow the upper curve; the polymer will experience an abrupt volume change as it hits a sharp thermal transition, called the glass transition temperature (T_g), and will solidify as an amorphous material. In contrast, if the polymer is sufficiently linear, it may pack regularly and follow the lower curve; the polymer will experience an abrupt volume change at the crystallization temperature. This type of polymer will crystallize (and melt) over a temperature range due to incomplete crystallization. The increase in elasticity and solid-like physical properties as a polymer melt cools is associated with the crystallization temperature in semicrystalline materials and with the T_g in amorphous materials.

In semicrystalline materials, the melting range can be very broad. Low molecular weight molecules melt (and crystallize) over a very narrow temperature range, whereas high molecular weight molecules melt over a broad range of temperatures and do not crystallize completely. Most crystallizable polymers contain significant amounts of amorphous material when solidified, and are hence called semicrystalline. Slower cooling leads to higher degrees of crystallinity. Physical properties of these materials are dominated by the crystalline fraction.

The T_g is characteristic of amorphous regions in the polymer, whether or not other regions are crystalline. Like the melting temperature (T_m), the measured value of T_g is dependent on testing conditions. T_g sets an upper limit for the use of amorphous polymers like polymethyl methacrylate (PMMA) or polystyrene (PS) and a lower limit for the rubbery behavior of elastomers. Polymers below T_g are glassy and elastic.

The softening point, which is lower than T_m , is measured as a temperature at which a polymer deforms a given amount under a given load. Semicrystalline polymers are most often used in applications at temperatures between T_g and the softening temperature, and in this region their behavior is described as tough and leathery. Below the T_g , the behavior is described as rigid. These regions are depicted schematically in Figure 38.2 for both amorphous and semicrystalline polymers.

When small molecules crystallize, the crystals have few defects, exhibit well-defined crystal faces and cleavage planes, and often have a single nucleus of growth. In contrast, when large molecules crystallize, a conglomerate of disordered material and clusters of crystallites develop from the growth of many nuclei, the crystal faces are nondistinct, and “tie molecules” traverse more than one crystallite. As with other materials, crystallizable plastics can exhibit more than one crystal form. The form, perfection, orientation, connectivity, and extent of crystals in a plastic affect the observed physical properties.

The morphology of semicrystalline polymers describes the forms that result from crystallization and aggregation of crystallites. The basic units are the crystalline “lamellae,” which consist of folded chain, ribbon-like crystallites (Figure 38.3). The lamellae are typically organized into larger structural features such as spherulites (Figures 38.4 and 38.5). A fibrillar morphology can be created by stretching a crystallizable polymer under temperatures between T_g and T_m , or by stretching spherulites (Figure 38.6). The microstructure and morphology of the polymer control the deformation mechanisms and mechanical properties exhibited in use.

38.2 GENERAL STRESS–STRAIN BEHAVIOR

Mechanical properties refer to the characteristics of a material that describe how it will respond to stress and strain. Stress is the force normalized by the area over which it is

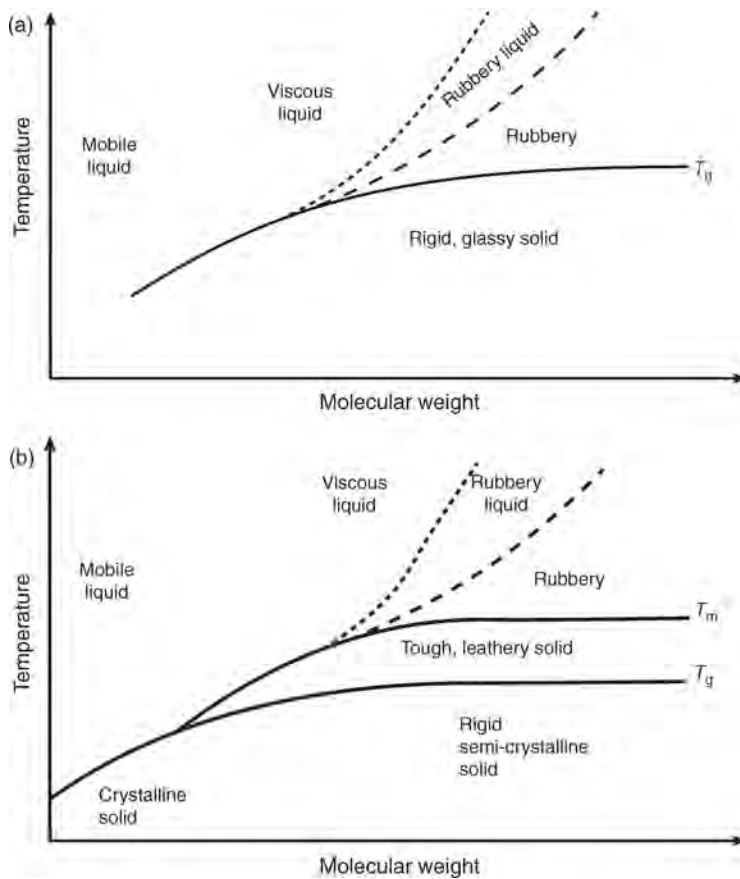


FIGURE 38.2 Relationships between temperature, molecular weight, and physical state.

applied, strain is the deformation normalized by the original dimension of the material. The modulus is the ratio between stress and the corresponding strain. It is easiest to illustrate these concepts by considering uniaxial tensile loading. Figure 38.7 shows monotonic tensile stress–strain curves for a synthetic fiber, a rigid and brittle plastic, a tough plastic, and an elastomer. Of the various forms of plastics, synthetic fibers generally have the highest initial modulus and moderate levels of elongation (strain) to break. Rigid, brittle plastics generally have moderate initial moduli and low elongation to break, whereas tough plastics can exhibit moderate to high initial moduli, a fairly evident yield point, and a larger amount of ductility or post-yield elongation. Elastomers exhibit the greatest amount of elongation to break, which is usually at least several hundred percent and may be more than 1000% as well. Elastomers also exhibit the lowest initial moduli, but the modulus increases rapidly as the elongation increases toward break. These differences arise from the polymer structure and morphology.

When subjected to uniaxial tension, some polymers exhibit (engineering) stress–strain response that appears similar to that of steel, where there is a local maximum yield point,

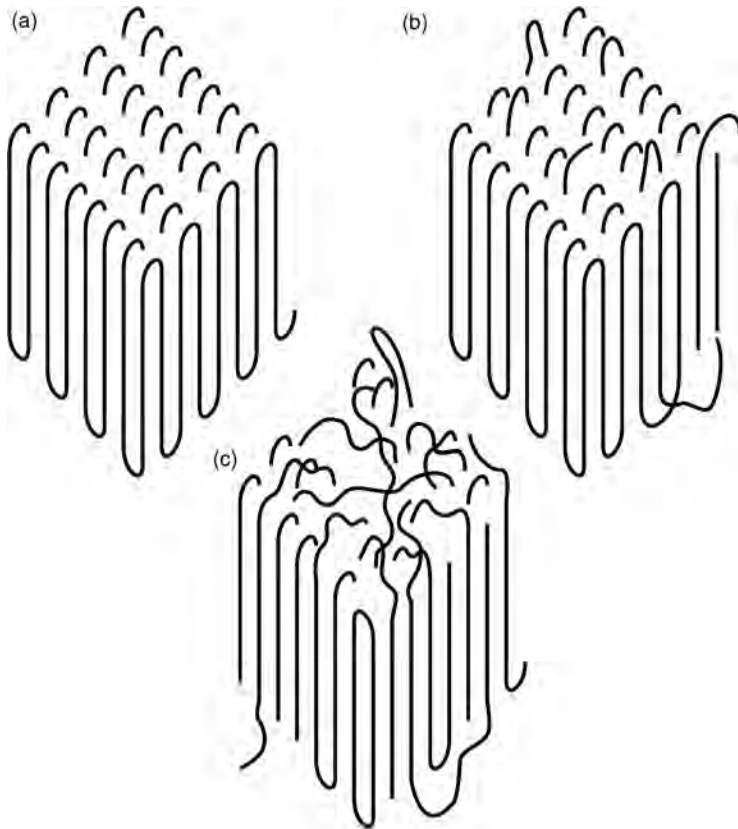


FIGURE 38.3 Crystal lamellae—Possible conformations of polymer chains at the surfaces of chain-folded single crystals. (a) Adjacent re-entry model with smooth, regular chain folds, (b) adjacent re-entry model with rough fold surface, and (c) random re-entry (switchboard) model.

after which the material necks or flows throughout the specimen gauge section at a lower (engineering) stress until the entire gauge section has necked down; the failure (engineering) stress may or may not be lower than the yield stress (Figure 38.7c).

Both semicrystalline and amorphous polymers can yield. Because of the time–temperature dependency, the yield point can become more pronounced at lower temperatures and higher strain rates, and less pronounced at higher temperature and lower strain rates. A uniaxial applied stress is comprised of both shear stress and dilatational stress, and this yielding behavior seems to be predominantly due to the shear component of the applied stress. Polymers often exhibit higher yield stress in compression than in tension, which indicates the behavior has a dependence on the hydrostatic stress (Dowling, 1999).

If the polymer is not loaded (strained) to failure, the unloading stress–strain behaviors also show significant distinguishing traits. For example, both fibers and plastics exhibit some amount of permanent deformation and some amount of delayed viscous recovery,

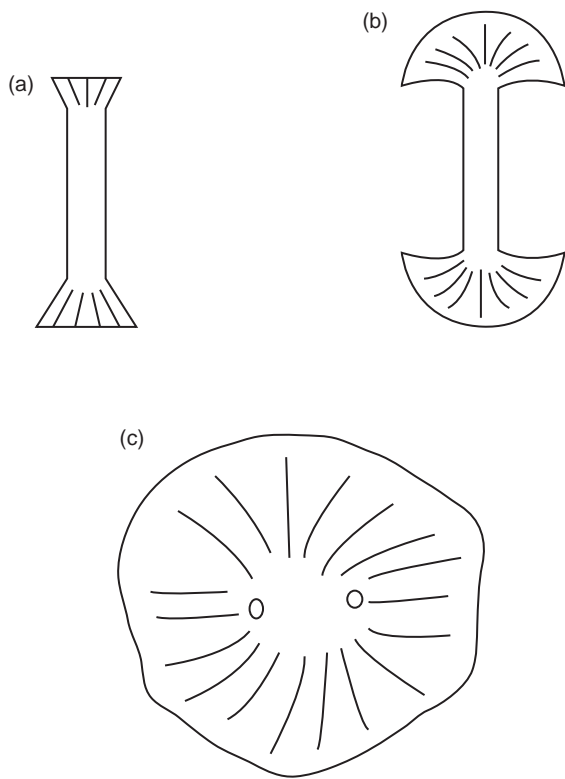


FIGURE 38.4 Successive stages in the development of a spherulite by fanning growth from a nucleus (Rudin, 1999).

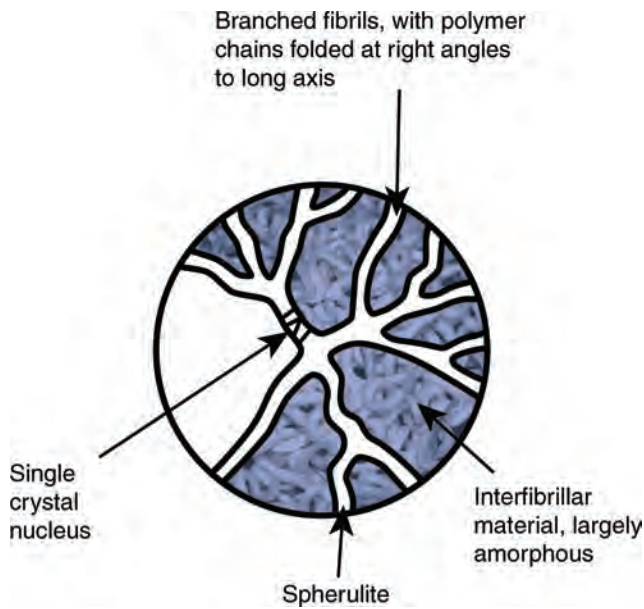


FIGURE 38.5 Schematic representation of basic structure of a polymer spherulite.

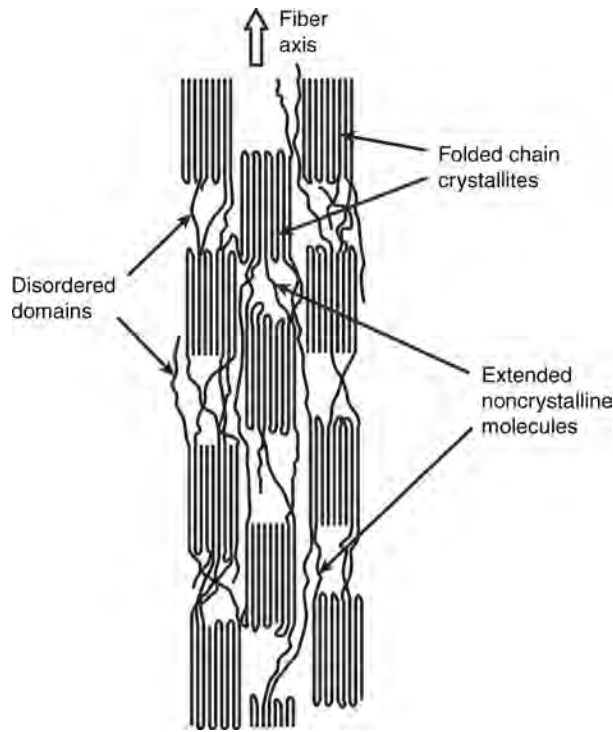


FIGURE 38.6 Schematics of polymer fibril morphology (Rudin, 1999).

but fibers usually exhibit more elastic strain recovery than plastics do. Elastomers, on the other hand, can recover completely and quickly from very high amounts of elongation.

The mechanical behavior of polymers, such as the tensile stress–strain behavior, is dependent on the temperature at which the material is evaluated. Figure 38.8 shows schematically how the tensile modulus changes with temperature for both amorphous and semicrystalline polymers. Thus, a material may be glassy at low temperatures and rubbery at higher temperatures. In the intermediate temperature range, where the modulus changes rapidly (for amorphous and lightly crystalline polymers), the behavior is described as leathery.

In the glassy region, there is no large-scale molecular motion; the material will generally respond elastically to an applied load and fracture under small strains, due to the inability of molecules to flow past each other. As the temperature increases to the T_g , the increased thermal energy permits more coordinated motion, ultimately leading to viscous flow (Table 38.1).

Because the glassy, leathery, and rubbery types of behavior are due to the amount of molecular motions occurring, it is apparent that anything that alters the molecular motions also alters the regions of behavior. Thus, molecular weight, crystallinity, crosslinking and plasticizers, and other factors can affect the temperatures at which these behaviors occur and the magnitude of change in modulus (Rudin, 1999; Hertzberg, 1983). Figure 38.9 illustrates the effects of increasing molecular weight and crosslinking density.

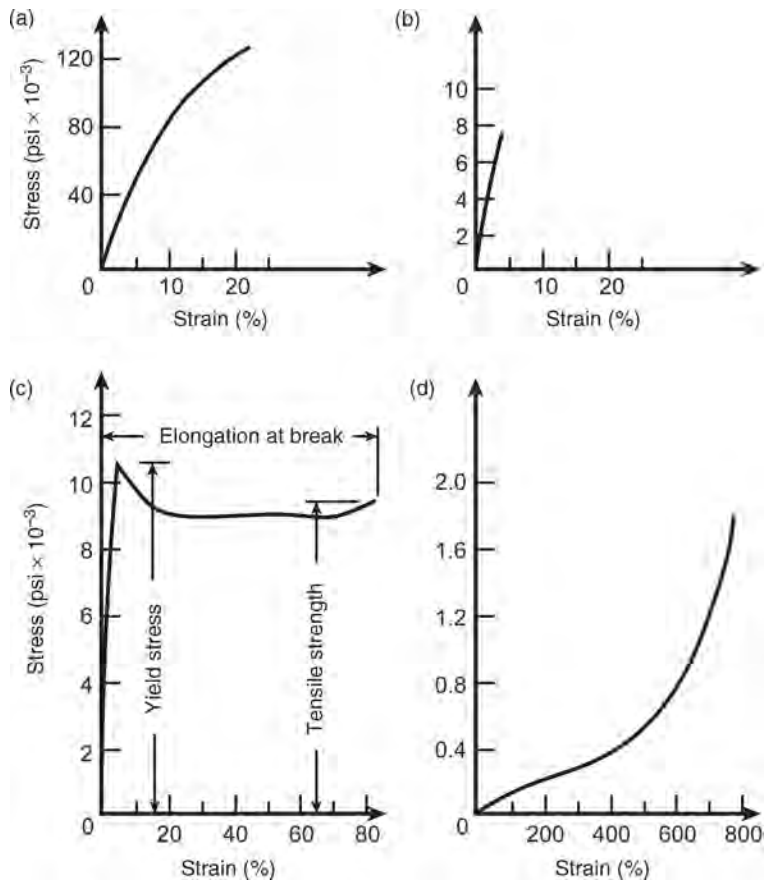


FIGURE 38.7 Monotonic tensile stress-strain curves for (a) a synthetic fiber, (b) a rigid and brittle plastic, (c) a tough plastic, and (d) an elastomer (Rudin, 1999).

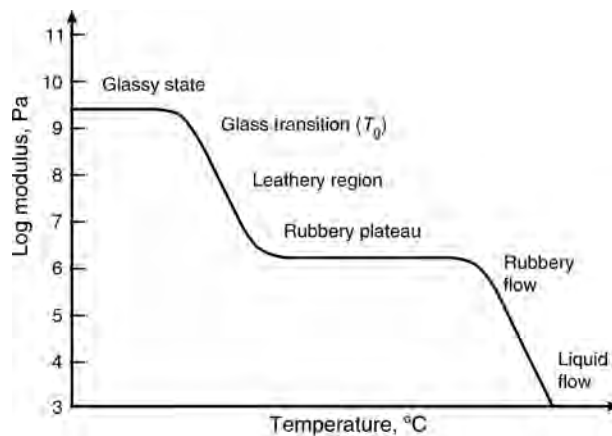


FIGURE 38.8 Modulus-temperature relationship for amorphous and semicrystalline polymers.

TABLE 38.1 Molecular Motions at Various Temperatures (Hertzberg, 1983)

Temperature	Molecular Source of Mechanical Behavior
$T \ll T_g$	Believed to be movements of small groups (few atoms only)
$T < T_g$	Believed to be movement of two to three consecutive repeat units
T_g	Believed to be coordinated movements of 10–20 repeat units
$T > T_g$	Large-scale molecular motions

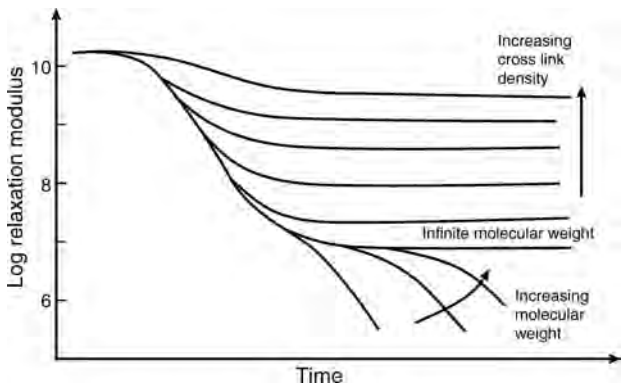


FIGURE 38.9 Effects of molecular weight and crosslink density on relaxation modulus.

38.3 VISCOELASTICITY

Polymeric solids behave in a manner that has characteristics of both Hookean solids and Newtonian fluids; this is called viscoelastic behavior. In an ideal elastic (Hookean) material, there are no time-dependent effects. This means that when stress is applied, the material deforms instantly to a strain that is directly proportional to the amount of stress. When the stress is removed, the deformation is recovered instantly and completely. In contrast, in an ideal Newtonian fluid, no stress can be supported without flow. When shear stress is applied, the fluid instantly responds in viscous flow. For these materials, the viscosity of the fluid is defined as the ratio of shear stress to shear rate.

Figure 38.10 schematically illustrates generic viscoelastic behavior of a crosslinked elastomer, a fiber and an amorphous plastic. Viscoelasticity is evaluated experimentally in creep, stress relaxation, and dynamic mechanical tests. In creep tests, a constant stress is applied and the deformation, which increases over time, is monitored. In stress relaxation tests, a constant deformation is applied and the reaction stress, which decreases over time, is monitored. In dynamic mechanical tests, the response of the material to periodically varying stress (or deformation) is monitored over time (Figure 38.11). In creep and stress relaxation tests, there are difficulties experimentally in measuring the very short-term (elastic) response, and time issues in waiting for the long-term response to occur. Dynamic mechanical tests (also called dynamic mechanical analysis or DMA) were developed to alleviate short- and long-term time

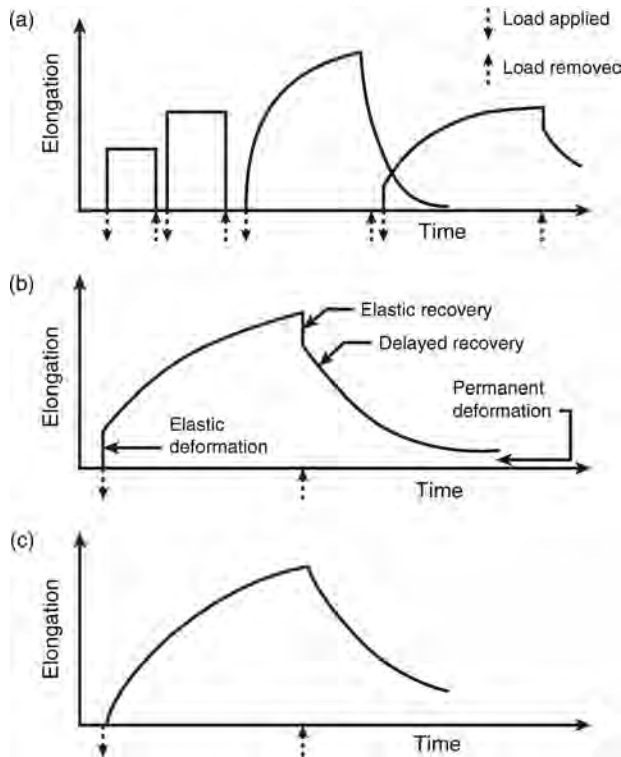


FIGURE 38.10 Elongation-time relationships upon loading/unloading for (a) crosslinked ideal elastomer, (b) fiber, and (c) amorphous plastic.

issues, but are limited to very small amplitude stresses and strains such that linear viscoelasticity methods can be applied. However, a polymer's useful range of stress and strain typically extends beyond the linear region and well into the nonlinear region. Thus, DMA can provide characteristic properties of a polymer, but the response acquired in DMA cannot be directly extended to large strain behavior.

The viscoelastic nature of polymers means that the strain (or stress) response is out-of-phase with the applied stress (or strain). Characteristics of the polymer behavior can be generated from the in-phase and out-of-phase components using vector analysis. The "storage" and "loss" components refer to in-phase and out-of-phase responses, and are linked to recoverable and dissipated energy, which is the result of molecular action. The loss tangent is the ratio of the two components at a particular frequency or temperature, and is often used to identify transitions in the material.

38.4 MECHANICAL MODELS OF VISCOELASTICITY

The viscoelastic behavior of polymers can be thought of as the sum of elastic and viscous behaviors. Elastic behavior of materials is classically modeled with a mechanical linear spring, where the amount of load required to compress the spring is linearly proportional to the amount of compression via a spring constant, k . Viscous behavior of fluids is

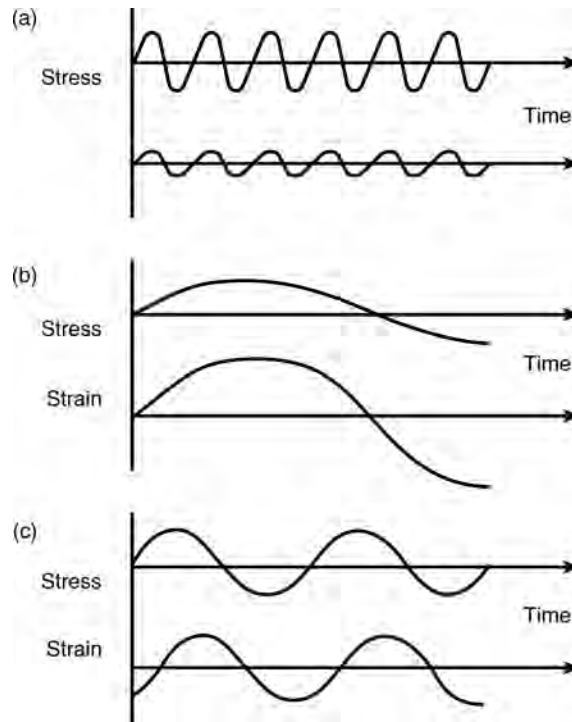


FIGURE 38.11 Effect of frequency on dynamic response of an amorphous, lightly crosslinked polymer: (a) elastic behavior at high frequency with stress and strain in phase, (b) liquidlike behavior at low frequency with stress and strain in phase, and (c) general case with stress and strain out of phase.

classically modeled with a dashpot (in which a piston moves through a vessel of viscous fluid), where the amount of shear stress required to move the piston is directly proportional to the shear rate via the fluid viscosity.

When the spring and dashpot are combined in series, the mechanical model is called a Maxwell element, and the resulting behavior is appropriate for representing stress relaxation. With this model, when load is removed, the model will instantly recover, but will not completely recover its deformation. When the spring and dashpot are combined in parallel, the mechanical model is called a Voigt or Kelvin element, and the resulting behavior is appropriate for representing creep behavior. With this model, when load is applied, the model will deform gradually to an asymptotic value.

The Maxwell and Voigt models are the two simplest mechanical models that are used to describe viscoelastic behavior. Many other, more complex models have been developed to represent certain physical behavior. Rudin (1999), Dowling (1999), and Hertzberg (1983) provide more information on various mechanical models of viscoelastic behavior. Dowling presents the simplest model that is roughly representative of real-world polymer creep, in that it combines elastic, steady-state creep, and transient creep elements; this model consists of a Maxwell element in series with a Voigt or Kelvin element. It is important to remember that the model is simply a tool to mathematically represent a physical behavior.

38.5 TIME-TEMPERATURE DEPENDENCE

There is a time-temperature similarity in the behavior of polymers, which is often called a time-temperature equivalence. The viscoelastic behavior at reduced temperatures is similar to that at reduced time periods, and the behavior at elevated temperatures is similar to that at extended time periods. This similarity or equivalence is related to the mobility of the chains, molecules, side groups, and so on, at a given temperature or time period.

From the work of researchers Williams, Landau, and Ferry, the WLF equation was developed to empirically relate this behavior equivalence in time and temperature. The WLF equation defines a shift factor (a_T) as a function of the T_g and temperature of interest. This shift factor applies to most amorphous polymers in the temperature region between T_g and $T_g + 100^\circ\text{C}$, and is calculated according to the equation below.

$$\log a_T = -\frac{C_1(T - T_g)}{C_2 + (T - T_g)}$$

The constants C_1 and C_2 have generally accepted values of 17.4 and 51.6. A more detailed discussion of this relationship and its link to other descriptions of polymer behavior can be found in Rudin (Rudin, 1999).

Long-term behavior can be determined from short-term behavior at higher temperatures. Master curves for creep and stress relaxation can be created using this equation if the user is cognizant not to apply it at temperatures at which the deformation mechanisms in the material change or at which crystallization can occur.

38.6 DEFORMATION MECHANISMS

38.6.1 Microscopic Deformations

Semicrystalline polymers consist of two distinctive regions: crystalline structure and amorphous structure. Crystalline structure is typically presented as lamellae, which are numerous chain-folded ribbons. Lamellar structures are separated by amorphous materials, which are randomly coiled polymer chains. The term “rigid amorphous phase” is sometimes used to describe the structure intermediate between crystalline and amorphous phases.

When a tensile stress is applied to a polymer, elastic deformation will take place; in that, the polymer chains will elongate along the direction of the stress by stretching of the covalent bonds within the molecular chain. In addition, displacement of adjacent molecules may also be possible, which involves the destruction of Van der Waals bonds between chains.

After the tensile stress reaches the yielding point, permanent, irrecoverable deformation will occur. The deformation mechanisms in amorphous polymers are relatively simple, with the material deforming under load by way of shear yielding or normal yielding (crazing, see Figure 38.12). In contrast, semicrystalline polymers undergo a more complex process of deformation due to their complex morphology.

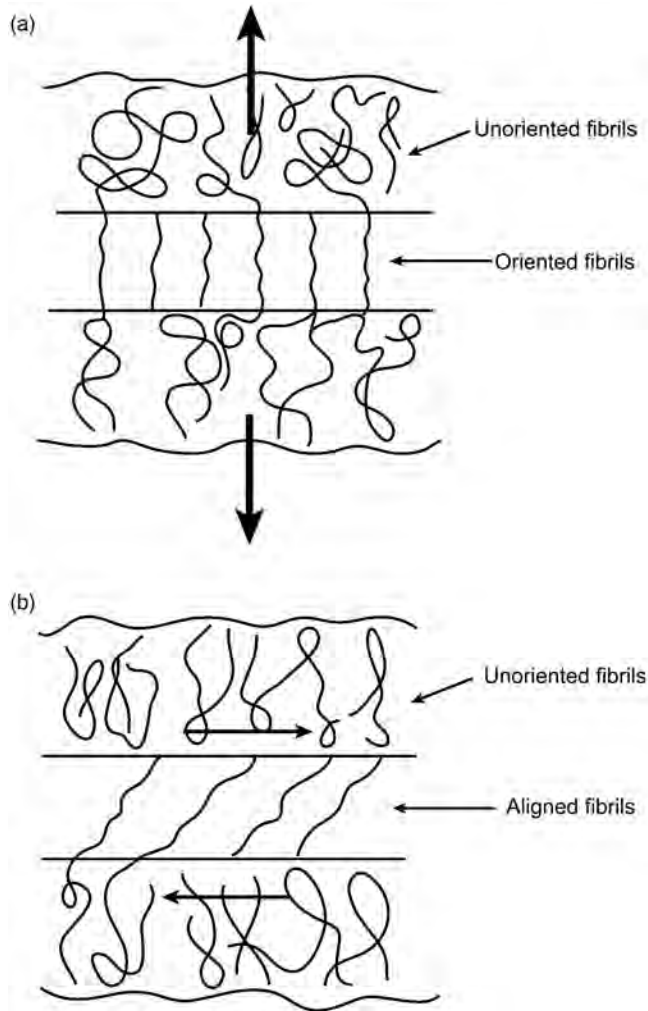


FIGURE 38.12 Deformation mechanisms in amorphous polymers: (a) normal yielding (crazing) and (b) shear yielding (Hertzberg, 1983).

For a semicrystalline polymer, the plastic deformation is usually a multistep process, which is schematically illustrated in Figure 38.13. In the initial stage of deformation, the tie molecules in the amorphous regions slip past each other and become extended and aligned in the tensile direction, while the lamellae regions maintain their structures as blocks of folded ribbons. In the second stage, as the deformation continues, the lamellae blocks continue to tilt and the chain folds become aligned in the tensile direction. In the next stage, crystalline segments separate from the lamellae and remain attached to each other by tie molecules. Finally, the segments and tie molecules become orientated in the direction of the tensile axis and the polymer produces a fairly large amount of deformation and orientation.

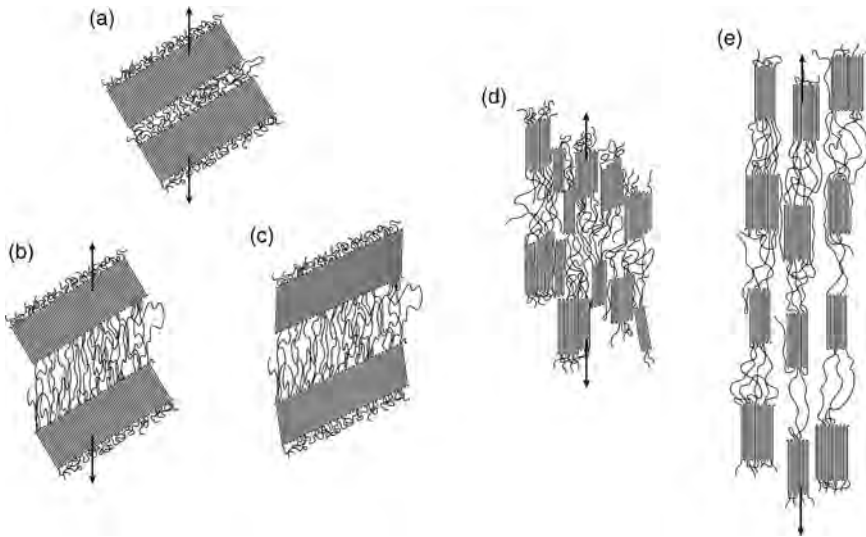


FIGURE 38.13 Multistep process in the plastic deformation of a semicrystalline polymer (Callister, 2003). Sequentially, this entails extension and alignment in the amorphous regions, followed by lamellar disruption and crystalline orientation.

38.6.2 Shear Yielding

One of the important mechanisms leading to plastic deformation in polymers is shear yielding. This process consists of a change in specimen shape at constant volume and is more common in semicrystalline polymers than in glassy polymers. In semicrystalline polymers, shear yielding occurs in forms of slip, twinning, and martensitic-like transformations. In glassy polymers, shear yielding takes place through the formation of shear bands. Example micrographs are shown in Figure 38.14.

Shear yielding is closely related to the process and mechanism of polymer fracture. Tresca and von Mises criteria are the two simplest criteria for yielding. The Tresca criterion states that yield will occur when the maximum shear stress on any plane reaches a critical value. It is expressed as

$$|\sigma_1 - \sigma_3| = 2\tau_y$$

where σ_1 , σ_2 , and σ_3 are principal stresses and $\sigma_1 > \sigma_2 > \sigma_3$. τ_y is the yield stress of the material in pure shear, and is correlated with σ_y , the uniaxial tensile yield stress by

$$2\tau_y = \sigma_y$$

The von Mises criterion states that yield will occur when the elastic shear strain–energy density reaches a critical value. It is expressed as

$$(\sigma_1 - \sigma_2)^2 + (\sigma_2 - \sigma_3)^2 + (\sigma_3 - \sigma_1)^2 = 6\tau_y^2$$

It is noted that both Tresca and von Mises criteria were originally developed for metals, and may not describe the shear yielding of polymers adequately in certain aspects. For example, based on the above criteria, the yield stress measured in uniaxial tension will be equal to that in uniaxial compression; this may not be true for polymers.

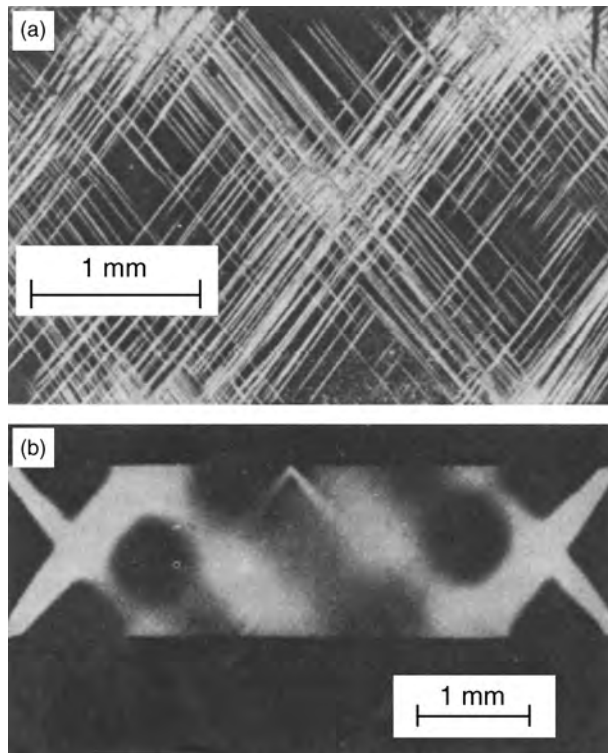


FIGURE 38.14 Optical micrographs images (under crossed polarized light) of the shear bands in glass polymers under a plain-strain compression: (a) polystyrene and (b) poly(methyl methacrylate) (Kinloch and Young, 1983).

Shear yielding plays an important role in the initiation and propagation of cracks in polymers. In glassy polymers, highly localized shear bands may lead to the initiation of cracks due to the formation of microvoids at the intersection of shear bands. Upon crack growth, if the plastic deformation is still highly localized at the crack tip, the polymer will exhibit a brittle fracture with a low toughness value. However, if the shear yielding can be homogeneously extended throughout the material, a ductile failure will generally occur and the material will exhibit high toughness, because a significant amount of fracture energy can be absorbed by the extensive plastic deformation. In such a case, shear yielding is an effective toughening mechanism for polymers.

38.7 CRAZING

Another important energy dissipation mechanism in polymer deformation is crazing. A craze can be considered a microcrack bridged by multiple, highly-oriented polymer fibrils. The main difference between shear yielding and crazing is that shear yielding occurs at constant volume, while crazing involves an increase in volume. In other words, shear yielding requires a deviatoric component of the stress tensor, while crazing requires a dilatational component to the stress tensor.

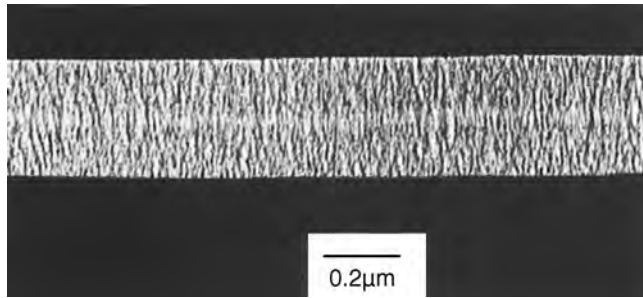


FIGURE 38.15 TEM micrograph of a craze formed in PS (Brostow, 2001).

A transmission electron microscopy (TEM) micrograph of a craze in polystyrene is shown in Figure 38.15. Crazes are typically initiated at sites with defects or molecular inhomogeneities. First, microvoids nucleate at the defects under a tensile stress. Then, the microvoids grow in a plane perpendicular to the maximum principal stress, which is a cavitation process. Next, instead of coalescing and forming a crack, these microvoids are stabilized by the surrounding highly-oriented polymer fibrils spanning the craze. A craze growth mechanism of “meniscus instability” is illustrated in Figure 38.16, showing that the craze tip is broken into a series of “fingers,” and the fibrils are formed behind the craze tip as the fingers pinch together.

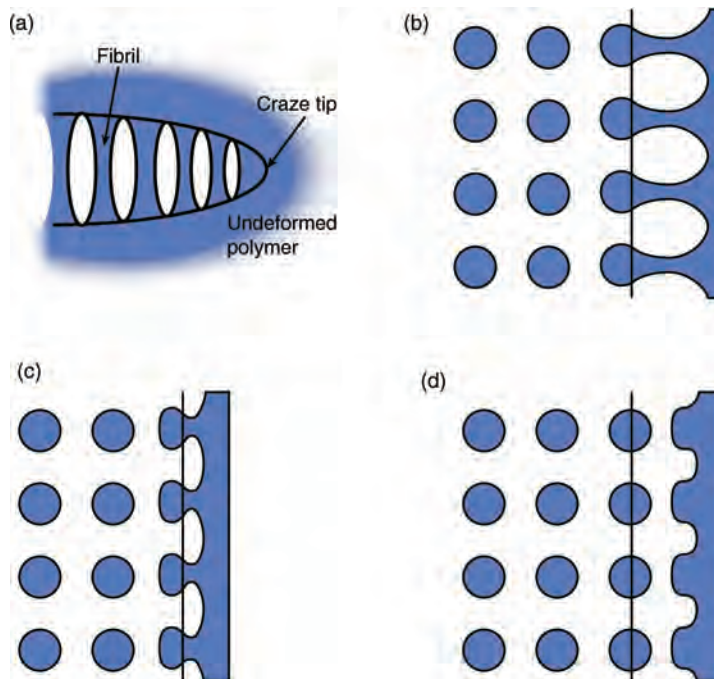


FIGURE 38.16 Illustration of craze growth by the “meniscus instability” mechanism.

Crazing has a close relationship with the fracture of glassy polymers, because the breakdown of crazes usually leads to a brittle fracture through microvoid coalescence and crack extension. Like shear yielding, crazing also involves considerable plastic deformation and energy dissipation, and thus can result in improved fracture resistance. In certain polymers, especially multiphase polymers like high impact polystyrene (HIPS) and acrylonitrile butadiene styrene copolymer (ABS), multiple crazing can be an effective mechanism in toughening.

38.8 FRACTURE

38.8.1 Griffith Theory

One important approach to explain the fracture of polymers is the energy balance theory developed by Griffith to model the brittle fracture of a glassy polymer with a pre-existing flaw. The flaw can be a scratch, a notch, or a sharp crack and serves as a stress concentration. The presence of a flaw in a body can be accounted for as an elliptical crack in an infinite plate loaded with a uniform stress, σ_0 , as shown in Figure 38.17. The stress at the tip of the crack, σ_t , can be written as

$$\sigma_t = \sigma_0 \left(1 + 2\sqrt{\frac{a}{\rho}} \right)$$

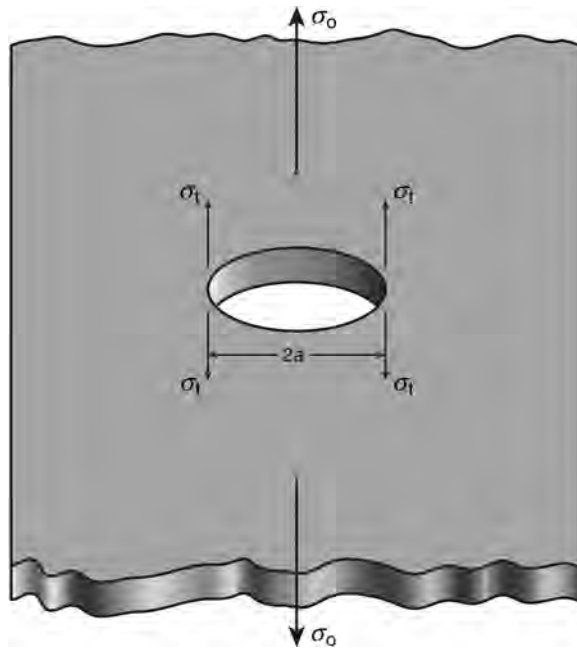


FIGURE 38.17 Model of an elliptical crack of length $2a$ in an infinite plate loaded with a uniform stress, σ_0 .

where $2a$ is the length of the crack and ρ is the radius of curvature of the tip. The above equation shows that the presence of a sharp crack will cause a large concentration of stress that has its maximum at the tip of the crack.

Griffith calculated the energy released in the fracture process by putting a sharp crack into a plate and related this to the energy required to create new surface. In Griffith's calculation, the critical stress of fracture, σ_c , can be expressed as

$$\sigma_c = \sqrt{\frac{2E\gamma}{\pi a}} \quad (\text{plane stress})$$

or

$$\sigma_c = \sqrt{\frac{2E\gamma}{\pi(1-\nu^2)a}} \quad (\text{plane strain})$$

where E is the Young's modulus, γ is the specific surface energy, and ν is Poisson's ratio. This calculation considers only the energy associated with surfaces and not the energy associated with deformation mechanisms that are often active in polymeric materials. The magnitude of the fracture stress depends strongly on the geometric constraints of the system. In plane strain, one of the three principal strains is equal to zero and the deformation is at constant volume. This is often obtained in the deformation of thick plates or constrained conditions around crack tips. Deformation of thin sheets results in plane stress conditions, in which the two principal stresses that are parallel to the free surfaces are finite and the third principal stress that is normal to the surfaces is zero. Thus, experimentally, thin sheet geometries will result in a higher fracture stress than for the same material with thicker, constrained geometries.

38.8.2 Fracture Mechanics

38.8.2.1 Linear Elastic Fracture Mechanics In the case of glassy polymers undergoing brittle fractures, linear elastic fracture mechanics (LEFM) can be applied. In this case, the material is linearly elastic and any yielding is restricted to a small region around the crack tip. In the LEFM analysis, critical stress intensity factor (K_c) and critical strain energy release rate (G_c) are the two most important parameters.

K_c is defined as

$$K_c = \sigma_c \sqrt{\pi a}$$

in the case of a wide plate sample containing a small crack. As K_c characterizes a material's resistance to brittle fracture, it is called fracture toughness.

The energy term, 2γ , can be replaced by G_c , allowing a correlation between K_c and G_c of

$$G_c = \frac{K_c^2}{E} \quad (\text{plane stress})$$

and

$$G_c = \frac{K_c^2(1-\nu^2)}{E} \quad (\text{plane strain})$$

There are three different modes of crack displacement. Mode I is an opening (or tensile) mode, where the fracture surfaces separate symmetrically with respect to the crack plane; Mode II is a sliding (or in-plane shearing) mode, where the fracture surfaces slide symmetrically with respect to normal, but asymmetrically with respect to the crack plane; Mode III is a tearing (or antiplane shearing) mode, where the fracture surfaces slide asymmetrically with respect to both the crack plane and its normal. Although the fracture of a material may involve either one or a combination of the three crack displacement modes, Mode I is the most commonly encountered and thus the fracture toughness and fracture energy for Mode I are cited for most situations.

For relatively thick samples, K_c is independent of sample thickness; however, for thin plates where sample thickness is comparable to the crack dimensions, K_c becomes dependent on the thickness, as shown in Figure 38.18. The stress condition of the crack in a relatively thick sample is generally plane strain state, so the K_c value at this condition is known as plane strain fracture toughness, K_{Ic} . The subscript “I” denotes that the plane strain fracture toughness is for Mode I crack displacement, as illustrated in Figure 38.19. As K_{Ic} is independent of the sample thickness, it is a well-accepted intrinsic material property to characterize fracture resisting capability. Similarly, the plane strain fracture energy for Mode I is denoted as G_{Ic} .

38.8.2.2 Elastic–Plastic Fracture Mechanics When nonlinear elastic deformation or large-scale plastic deformation has been developed in the vicinity of crack tip, the above LEFM approach no longer applies. Instead, as one form of the elastic–plastic fracture mechanics (EPFM), a J -integral concept was developed to calculate the energy parameter for elastic–plastic materials (Brostow, 2001). The J -integral is defined as the contour line integral which is independent of the integration path and can be expressed as

$$J = -\frac{1}{B} \frac{dU}{da}$$

where B is the thickness and U is the potential energy. The critical J -integral value at crack initiation in Mode I crack displacement is denoted as J_{Ic} , which is an intrinsic material property of resistance to fracture.

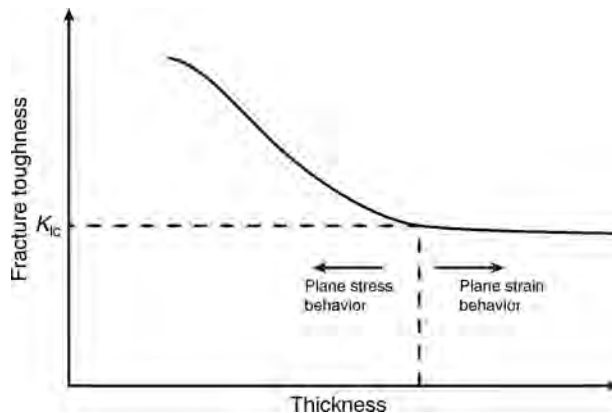


FIGURE 38.18 Schematic of the effect of sample thickness on fracture toughness, K_c .

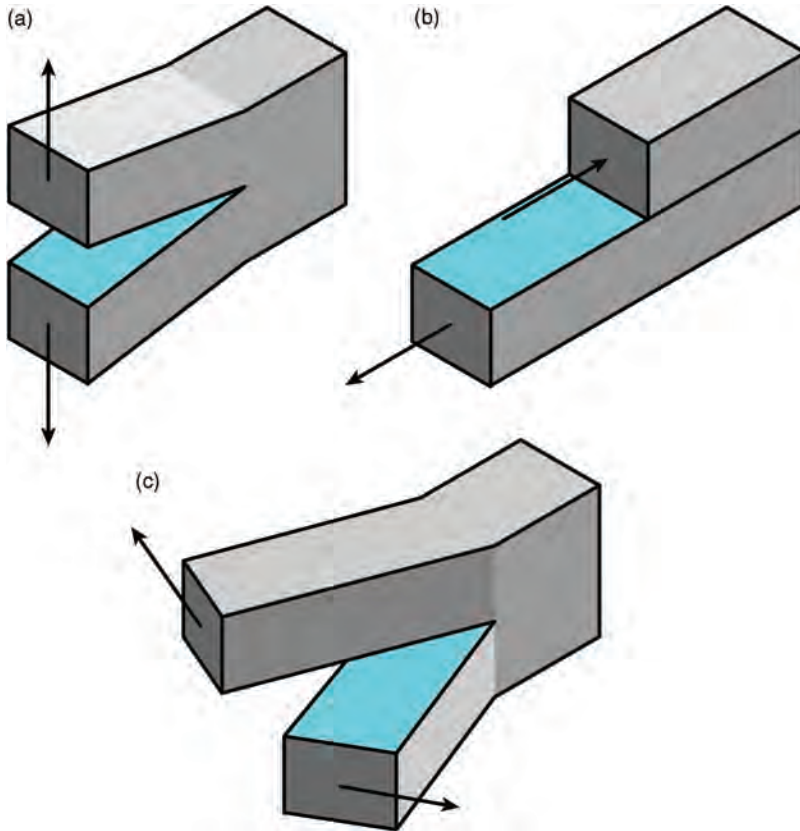


FIGURE 38.19 Three modes of crack displacement: (a) Mode I: opening (or tensile) mode, (b) Mode II: sliding (or in-plane shearing) mode, and (c) Mode III: tearing (or anti-plane shearing) mode.

To experimentally measure the J_{Ic} value, multiple specimens are loaded to various levels to obtain crack extensions, Δa . This procedure is outlined in ASTM E813. For certain types of specimens, the J -integral can be obtained from the load-displacement curve

$$J = \frac{2A}{W - a}$$

where A is the area under the load-displacement curve and $(W - a)$ is the ligament thickness. Then, a crack growth resistance curve (R-curve) can be drawn by plotting the J -integral values against corresponding Δa . In a nonlinear fracture, the crack tip will be blunted before the crack can further extend. The crack blunting line is defined by

$$J = 2\sigma_y \Delta a$$

where σ_y is the yield stress of the material. The intersection of the R-curve and the crack blunting line defines the value of J_{Ic} .

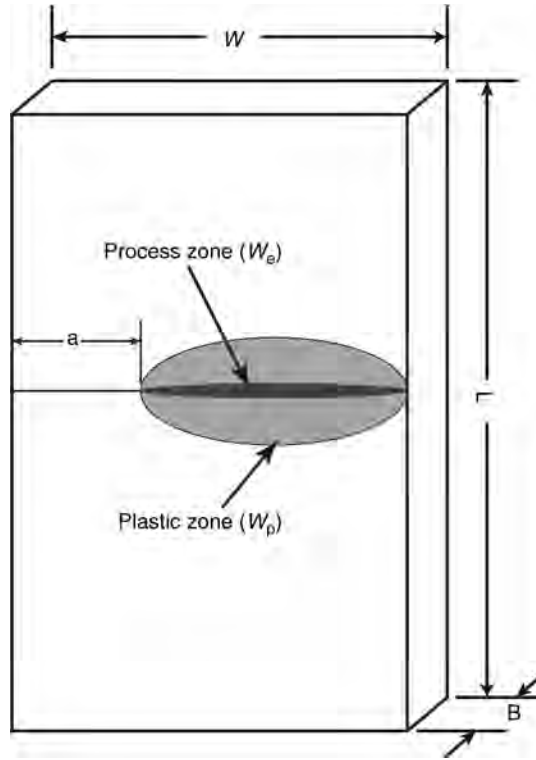


FIGURE 38.20 Schematic of the process zone and the plastic zone around the crack tip in a ductile material.

38.8.2.3 Essential Work of Fracture When the material is extremely ductile, essential work of fracture (EWF) can be used (Brostow, 2001). This method divides the deformation zone around the crack tip into two different regions, the inner process zone and the outer plastic zone, as shown in Figure 38.20. Consequently, the total work of fracture, W_f , can be separated into two parts, the essential work of fracture, W_e , and the plastic work of fracture, W_p , which is written as

$$W_f = W_e + W_p$$

W_e is proportional to the ligament thickness, $W - a$, while W_p is proportional to $(W - a)^2$. Thus, W_f can also be written as

$$W_f = w_f B(W - a) = w_e B(W - a) + \beta w_p B(W - a)^2$$

where w_f is the specific total work of fracture per unit surface area, w_e is the specific essential work of fracture per unit surface area, w_p is the specific plastic work per unit volume, and β is a plastic zone shape factor. The equation can be re-written as

$$w_f = w_e + \beta w_p(W - a)$$

When plotting w_f against $(W - a)$, w_e can be obtained by extrapolating this linear correlation to zero $(W - a)$.

38.9 MODIFYING MECHANICAL PROPERTIES

38.9.1 Toughening of Polymers

Toughening technologies have been researched for decades to improve the fracture toughness of brittle polymers. Shear yielding and crazing are the two most important toughening mechanisms for polymers, but other mechanisms, including croiding (crazing/voiding), crack pinning, bifurcation, crack deflection, crack bridging, and segmental crack growth are recognized (Arends, 1996). To improve fracture toughness, it is important to facilitate deformation processes that can involve toughening mechanisms extensively throughout the material, so more fracture energy can be dissipated.

One toughening technique is molecular flexibilization. This can be achieved by incorporating flexible segments into the macromolecular chains, or increasing the molecular weight between crosslinks for thermosets. The flexibilization of molecular architectures leads to the relaxation of polymer chains or networks, which contributes to the energy dissipation and toughness improvement. However, one of the major drawbacks of this method is compromised strength and T_g , and the effectiveness is limited.

Another popular toughening technique is the incorporation of a second-phase toughening agent. The most widely accepted technique in the industry is adding rubber particles. The industrial production of impact polymers like HIPS and ABS are examples of this type of rubber-toughened polymer. This approach can apply to thermosetting polymers as well. For example, both reactive rubber particles, such as carboxyl terminated butadiene acrylonitrile (CTBN), and nonreactive core-shell particles, such as a butadiene-styrene core with a styrene-methylmethacrylate-acrylonitrile-glycidyl methacrylate shell, have exhibited great toughening effect in matrices of diglycidyl ether of bisphenol A (DGEBA) epoxies (Pearson et al., 2000). Unfortunately, significant increase in viscosity with the addition of rubber particles may largely compromise the processability in certain applications. As an alternate, thermoplastic particles, such as nylon spheres, are sometimes used as toughening agents.

More recently, a novel technique using nanosized amphiphilic di- and tri-block copolymer (BCP) micelles as a toughening agent has shown great effectiveness in improving epoxy fracture toughness. When mixing the BCP with the matrix, the compatible block will dissolve into the matrix, while the incompatible block microphase separates from the mixture, thus forming a micellar structure. Because of the unique structure, extremely small size of the micellar particles, and overall low concentration required, BCP-toughened epoxies retain T_g and modulus, as well as suitable viscosities for processing (Liu,).

38.9.2 Reinforcement of Polymers

The strength and stiffness of polymers can be modified with reinforcements. Depending on the type of the reinforcing agents, polymer composites can be classified as one of three different families: filler-reinforced, fiber-reinforced, and structural composites. In load-bearing applications of polymers, fillers are generally used to increase stiffness and reduce cost, whereas fibers are generally used to increase strength and fatigue performance.

Filler-reinforced composites usually refer to a random dispersion of small, hard fillers in the polymer matrix. The reinforcing fillers can be spherical particles (ZnO , SiO_2 , Al_2O_3 , CaCO_3 , and so on), platelets (montmorillonite, kaolin, zirconium phosphate, graphene, and so on), or randomly oriented small fibers/tubes (halloysite, carbon nanotube, and so on). An isotropic improvement in strength and modulus can be produced in this

case, while sometimes other properties, such as thermal stability, flame retardancy, and barrier property may also be improved. If at least one of the dimensions of these fillers is under 100 nm, this type of composite is referred to as a nanocomposite. For the past two decades, polymer nanocomposites have gained a great deal of interest and effort in both industry and academia for their outstanding performance and light weight (Boo, et al. 2006).

Fiber-reinforced composites usually consist of discontinuous or continuous fibers with an alignment in the polymer matrix. Glass and carbon are the two most popular fiber materials. Composites containing oriented fibers usually exhibit a high anisotropy and a uniaxial stress-strain response. The strength and stiffness in the longitudinal direction (parallel to the fiber orientation) can be significantly improved with the addition of fibers to the matrix, while the transverse strength can be extremely low, sometimes even lower than that of the matrix.

Another family of composites is structural composites, in which the inclusion material has a two-dimensional or three-dimensional structure. Laminate composites are one of the most common structural composites. These use laminated sheets of fabric materials, such as wood, paper, woven glass, or carbon fibers embedded in the polymer matrix. The final properties of the products are dependent not only on the physical properties of the materials of composition, but also on the geometrical design of the structures. Additionally, the processing techniques of structural composites are extremely important, because an improper processing procedure may result in critical defects in the material, such as delamination or air bubbles/pockets. Generally, structural composites can produce very high strength and stiffness, because of their high load transmission and distribution efficiency.

For both fiber-reinforced composites and structural composites, crack growth involves similar processes. Crack initiation usually occurs at the fiber/matrix interface. Then, crack propagation occurs along the interface, leading to debonding, fiber pull-out, and sometimes fiber breakage. Crack propagation can extend through the matrix, and matrix deformation and cavitation can occur.

38.10 LOAD-BEARING APPLICATIONS: CREEP, FATIGUE RESISTANCE, AND HIGH STRAIN RATE BEHAVIOR

38.10.1 Creep and Creep Rupture

At temperatures below T_g , creep effects are small, but as the temperature increases above T_g creep effects become significant. The mechanism of creep in amorphous polymers is reptation, in which the motion of molecules occurs in a time-dependent manner and occurs more rapidly at higher temperatures. These materials behave much like a viscous fluid, with molecules or groups of molecules moving relative to one another over time when stress is applied. Because higher temperatures cause more oscillations at the atomic level, in turn causing more molecular movements that constitute creep, this process can be expressed as an Arrhenius equation. The derived expression for creep due to viscous flow directly relates the strain rate to the applied stress (Dowling, 1999).

For semicrystalline polymers, creep deformation by viscous flow occurs at elevated temperatures, approaching the T_m . This flow of molecules past each other is easier for polymers with lower molecular weights, less crosslinking and fewer side-branches. These considerations are of practical importance to the processing and performance of different grades of polyethylene.

At more moderate temperatures (just above T_g), the molecules cannot move past each other as readily as at higher temperatures. This is due to the secondary bonds between the

polymer chains being more effective at lower temperatures. The result is that the molecules are more easily entangled and create increasing resistance as the creep deformation progresses. These same entanglements act like recovery springs when the load is removed, pulling the stretched molecules toward their original position. The end result is that at temperature moderately above T_g , creep behavior is very different from viscous flow, as the entanglements both limit the amount of creep deformation and provide deformation recovery.

Figure 38.21 schematically depicts typical creep behavior, in terms of strain, as a function of time. There are three stages of creep: an initial transient stage, which contains nearly instantaneous deformation that is predominantly elastic, but may contain some plastic strain; a secondary steady-state region, where there is a more gradual, but steady increase in deformation, and a tertiary unstable region, in which deformation escalates as rupture failure is approached. During the tertiary creep stage, the creep deformation becomes localized through formation crazes, cracks and void coalescence, or formation of a neck in the case of ductile polymers. The creep strain–time curve for a polymer is affected by the magnitude of applied stress and temperature.

Dowling (1999) presents the development of isochronous stress–strain curves and secant modulus curves, which are often used in component design for polymers. Material suppliers often provide such information. However, differences in polymer composition or processing affect these data, and it is recommended that data specific to the material and process of interest be obtained.

In engineering applications where creep occurs, the design and material selection must ensure that neither excessive deformation nor rupture occur within the design life. Creep data can be used to generate stress life curves for the stress and time required to reach a given limiting amount of deformation or rupture, as dictated by the design criteria.

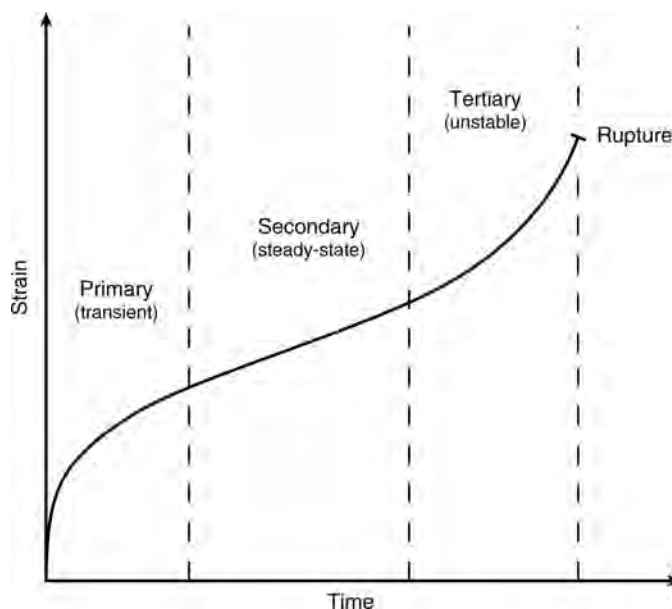


FIGURE 38.21 Three stages of creep behavior under constant engineering stress.

38.10.2 Fatigue Resistance

In most load-bearing applications, the applied stress does not remain constant. If the stress changes are infrequent and small, stress–life and time–temperature data can still provide reasonably accurate life estimates. However, if the stress changes are frequent enough, fatigue damage can start to occur. Fatigue of polymers is more complex than for metals, due to that fact that there is almost always a creep component involved due to the intrinsic viscoelasticity of polymers. Depending on the loading frequency, there can also be thermal damage generated in addition to the mechanical damage.

Dowling (1999), Atkins and Mai (1988), and Hertzberg (1983) provide both excellent background and details on fatigue of materials. The rest of this section touches on areas of fatigue behavior and analysis that are pertinent to polymers. Topics such as strain life versus stress life, the effect of mean stress, variable amplitude loading and overloads, which are all very important in plastics, are too complex to introduce here, but are aptly covered in the aforementioned texts.

38.10.3 Creep-Fatigue Interaction

The time–temperature dependence of polymers means that creep-relaxation phenomena occur in combination with cyclic loading. The slower the test frequency and higher the temperature, the more significant the creep component. This means that loading waveforms (sinusoidal versus square versus ramp) directly influence the overall resulting creep-fatigue behavior.

There are three main approaches to handling creep-fatigue interaction to assess long-term life:

- Sum of life fractions due to creep (time-fraction) and fatigue (Palmgren–Miner rule), which is expressed as $\sum \Delta t_i / t_{ri} + \sum \Delta N_i / N_{fi} = 1$. This method can also be based on attaining a given amount of strain.
- Frequency-modified fatigue approach by Coffin, in which the cyclic stress–strain and strain–life relations are generalized and material constants are functions of temperature and frequency.
- Strain-range partitioning method by Manson, which considers life fractions spent in each of the following four types of loading: plastic deformation with little creep, creep in tension, creep in compression, and creep in both tension and compression.

38.10.4 Fatigue Resistance of Polymers

Fatigue resistance of polymers is assessed by one of the following three methods: crack propagation until fracture; accumulation of damage, which is commonly used for fiber-reinforced polymers; and degradation of a mechanical property, such as modulus, yield, or unrecovered strain.

Similar to metals, polymers can exhibit fatigue striations. At high stress intensity (ΔK) levels, striations generally correspond to crack advancement due to one load cycle. However, in polymers, especially at low ΔK levels, striations generally correspond to discontinuous crack advancement following several hundred load cycles, during which the crack front remains stationary. Thus, macroscopic growth rate data or other supplemental

information is necessary to distinguish between fatigue markings created by a single cycle and markings created by up to hundreds or thousands of cycles.

Semicrystalline polymers have superior fatigue crack propagation (FCP) resistance over amorphous polymers. This is a direct consequence of the microstructure and morphology of semicrystalline polymers. Energy is dissipated by deforming crystallites; the reformed crystalline structure is strong; and the two-phase structure (crystalline and amorphous) impedes crack propagation. Many amorphous polymers, such as PS, PMMA, and polycarbonate (PC), can have an endurance limit of only one-fifth the static fracture stress, whereas crystalline polymers, such as polyethylene, polypropylene, and polytetrafluoroethylene (PTFE), can have an endurance limit of one-third to one-half the static fracture stress, and sometimes higher. These are generalized endurance limit values and actual values depend on the testing conditions under which they were obtained (strain-control or stress-control, loading waveform and frequency, and so on) and on the particular composition of the polymer.

There are several compositional factors that affect the fatigue resistance of polymers. Fatigue resistance is greater for higher molecular weights (MW) due to greater entanglement density and orientation hardening. The T_g and modulus of a polymer can plateau at higher MW (100,000 to 200,000), but fatigue life and strength can continue to improve (as can creep resistance). This applies to both amorphous and crystalline polymers. Fatigue resistance can also be improved with a broader molecular weight distribution (MWD), and particularly those with high-MW tails. Because chain ends can be sources of submicron cracks, fewer chain ends can lead to improved fatigue resistance.

A lower T_g of the amorphous region makes a polymer less sensitive to flaws at its use temperature, because of the increased chain and molecular mobility. Similarly, a polymer with a lower melt flow index will have improved fatigue resistance over its higher melt flow counterpart, because of entanglements that limit creep and crack propagation.

Crosslinking increase rigidity, reduces ductility, inhibits crazing and creep, and inhibits local plastic deformation and crack tip blunting. Consequently, less crosslinking leads to improved fatigue resistance.

Rubbery additions generally do not have as significant an effect on fatigue resistance as they do on impact resistance and toughness. However, the effect can depend on MW and amount added.

Fillers, such as talc and other minerals, can delay cyclic softening until near the end of life, but the lifetime is affected very little. Thus, caution should be used when measuring property degradation to assess fatigue life of filled polymers.

The environment in which the polymer is exposed to fatigue can also strongly affect fatigue performance. For example, testing a polymer in an unreactive, but volatile (evaporative) liquid (e.g., nylon in water or alcohol) can improve apparent fatigue resistance due to cooling. Likewise, testing under a nitrogen blanket as opposed to air can double the life of PS; PS is not sensitive to oxygen but fatigue fracture may include some bond fracture as craze fibrils rupture, and these free radicals interact with oxygen to produce oxidative degradation. Finally, crazing or cracking agents can be used to accelerate testing, for example, alcohols on PS or Igepal on high density polyethylene (HDPE).

Fatigue loading waveforms (shape and frequency) can strongly affect FCP rates due to associated strain rate and creep-relaxation phenomena. For un-notched specimens, hysteretic heating at higher test frequencies leads to a reduction in fatigue resistance due to the high capacity for energy absorption and low thermal conductivity (worse for thicker samples), and reduced elastic modulus from generalized heating. However, with notched

specimens, fatigue resistance can increase with increased test frequency; this is related to the frequency of movement of the main chain segments responsible for transition peak (resonance-like condition) at the temperature of concern, with localized crack tip heating leading to crack tip blunting. With regards to strain rate effects resulting from the loading waveform, higher frequencies and faster loading rates (ramp versus sinusoidal) result in faster rates of strain, which can affect the polymer differently.

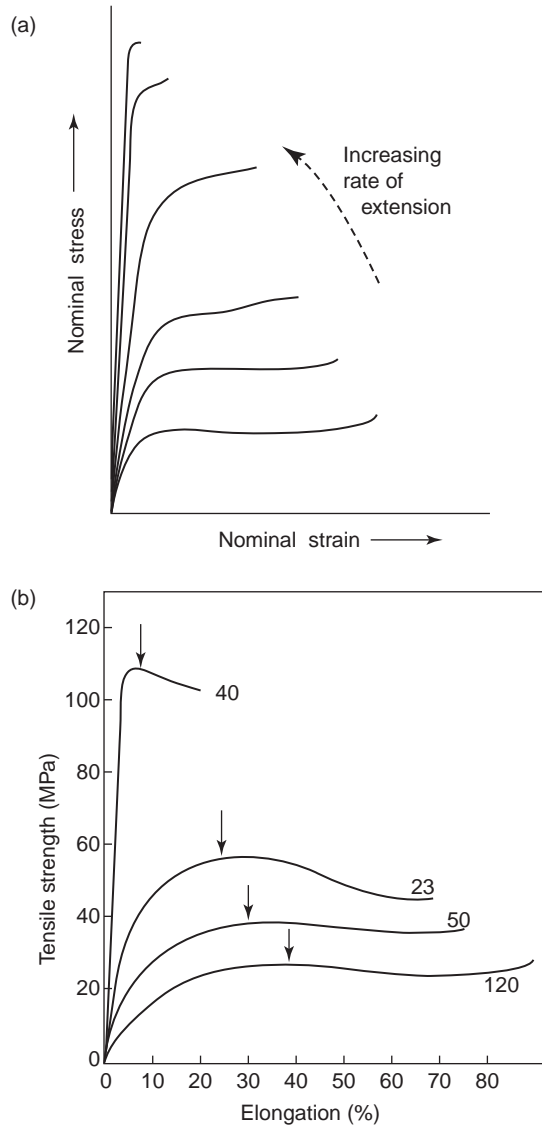


FIGURE 38.22 Stress-strain behavior of a polymer at different (a) strain rates and (b) temperatures (Rudin, 1999).

38.10.5 High Strain Rate Behavior

Because of the time–temperature dependence of polymers, it is understood that changes in behavior with decreasing temperature are similar to the change in behavior with decreasing time periods, or increasing strain rate. Figure 38.22 illustrates the general effects of decreasing temperature and increasing strain rate on the stress–strain response.

The deformation mechanisms in polymers depend on molecular mobility. At high strain rates (or loading rates), there is less time available for molecules to stretch, twist, and slide, and thus there is less resulting molecular movement. This is exhibited by the material response becoming stiffer and more brittle in nature. The most pronounced effect is often on the failure strain, with it becoming significantly lower with increasing strain rate. The failure stress and yield stress (if the polymer exhibits one) increase with increasing strain rate and the initial modulus often increases as well. Most polymers have a strain rate range over which the rate effects are most pronounced, and below and above that range the effects are more moderate.

It is difficult to acquire high quality stress–strain data for polymers at high strain rates, particularly if covering the entire range of behavior, from initial elastic modulus through yield and plastic deformation, to strain at failure. Numerous research activities have been underway to address these difficulties and provide protocol for experimental testing of polymers under high strain rates (Hill, 2005; Xiao, 2007; Pinnell, 2008). SAE J7149 provides guidance in experimentally determining high strain rate tensile behavior of polymers.

REFERENCES

- Arends CB. *Polymer Toughening*. Dekker; 1996.
- ASM International. *Characterization and Failure Analysis of Plastics*. ASM International; 2003.
- Atkins AG, Mai Y-W. *Elastic and Plastic Fracture*. Ellis Horwood Ltd; 1988.
- Boo WJ, Liu J, Sue HJ. Fracture Behavior of nanoplatelet reinforced polymer nanocomposites. *Materials Science and Technology* 2006; 22:829–834.
- Brostow W. *Performance of Plastics*. Hanser; 2001.
- Callister WD. *Materials Science and Engineering—An Introduction*. 6th ed. John Wiley & Sons; 2003. p. 493. Fig. 15.12; p. W-17. Fig. 8.8W; p. 202. Fig. 8.9.
- Dowling NW. *Mechanical Behavior of Materials: Engineering Methods for Deformation, Fracture, and Fatigue*. 2nd ed. Prentice-Hall, Inc. 1999.
- Hertzberg RW. *Deformation and Fracture Mechanics of Engineering Materials*. 2nd ed. John Wiley & Sons; 1983. p. 215. Table 6.5, Fig. 6.16 (reprinted with permission from McGraw-Hill Book Company), Fig. 6.30.
- Hill SI, Pinnell MF, Minch AJ. Standardization of high strain rate tensile testing of polymers. ANTEC Proceedings, Society of Plastics Engineers; 2005.
- Kinloch AJ, Young RJ. *Fracture Behavior of Polymers*. Applied Science Publishers; 1983.
- Liu J, Sue H-J, Thompson ZJ, Bates FS, Dettloff M, Jacob G, Verghese N, Pham H. Nano-cavitation in self-assembled amphiphilic block copolymer-modified epoxy. *Macromolecules* 2008; 41:7616–7624.
- Pearson RA, Sue HJ, Yee AF. *Toughening of Plastics*. American Chemical Society; 2000.
- Pinnell MF. Assessment of Techniques Used to Measure Strain During High Rate Tensile Testing of Polymeric Materials, SAE Paper 08AE-79; 2008.
- Rudin A. *Polymer Science and Engineering*. 2nd ed. Academic Press; 1999.
- Xiao X. *Stress Equilibrium in Dynamic Tensile Testing of Plastic Materials*. General Motors; 2007.

39

ELECTRICAL PROPERTIES OF POLYMERS

EVARISTO RIANDE AND RICARDO DIAZ-CALLEJA

- 39.1 Introductory remarks
 - 39.2 Polarity and permittivity
 - 39.3 Measurements of dielectric permittivity
 - 39.4 Polarization and dipole moments in isotropic systems
 - 39.4.1 Dielectric relaxations
 - 39.5 Thermostimulated depolarization currents
 - 39.6 Conductivity in polyelectrolytes and polymer-electrolytes as separators for low temperature fuel cells and electrical batteries
 - 39.7 Semiconductors and electronic conducting polymers
 - 39.8 Ferroelectricity, pyroelectricity, and piezoelectricity in polymers
 - 39.9 Nonlinear polarization in polymers
 - 39.10 Elastomers for actuators and sensors
 - 39.11 Electrical breakdown in polymers
- References

39.1 INTRODUCTORY REMARKS

Polymers are large molecules or macromolecules made up of repeating structural units usually connected by covalent chemical bonds. Polymers with structural units of the same type are called homopolymers and copolymers otherwise. Most skeletal bonds in macromolecules fluctuate between energy wells or rotational states giving rise to a nearly unlimited number of macromolecular conformations. As a result, the macromolecular nature of polymers vastly broadens the time scale for molecular adjustments to external

force fields. A relevant characteristic of polymers is their ability to withstand high electric fields with negligible conduction due to the high energy gap between the localized valence electronic states and the conduction band. This characteristic coupled with favorable processing properties makes polymers the obvious choice for insulating applications. The versatility of design of polymer chains has expanded the use of these materials to unexpected applications. Thus, polymers can be designed to be used as solid electrolytes for batteries and low temperature fuel cells, piezoelectric and pyroelectric sensors, electronic and ionic conductors, NLO materials, and so on. This chapter is focused on the description of the influence of the chemical structure on the polarity of molecular chains. Special attention is paid to the analysis of the influence of the physical nature of polymers (rubbery, glassy, crystalline, and liquid crystal) on their response to electrical force fields in wide temperature/frequency windows. Also chemical structures are described in relation with polymers capability to be used in ionic and electronic applications of great interest.

39.2 POLARITY AND PERMITTIVITY

A dipole is defined as an entity in which a positive charge q is separated by a distance \mathbf{r} from a charge of equal magnitude, but negative sign. The dipole moment is a first-order tensor or vector defined as $\boldsymbol{\mu} = q\mathbf{r}$. By convention, the direction of a dipole points from the negative toward the positive charge. Traditionally, the dipole moment resulting from two charges of 10^{-10} statc, separated by a distance of 1 Å, named Debye, is used as the dipole moment unit in the cgs system. Then, $1 \text{ D} = 10^{-18} \text{ stat C cm}$ and $3.338 \times 10^{-33} \text{ cm}$ expressed in cgs and SI units, respectively. Two electronic charges of different sign ($q = 1.6022 \times 10^{-19} \text{ C}$) separated by a distance of 1 Å have a dipole moment of 4.8 D. The dipole moment of rigid polyatomic molecules depends on their geometric structure and can be calculated as first approximation from the vectorial sum of bond dipoles.

In the International System of units, the electrical force lines crossing a surface S surrounding a positive electric charge q are given by (Jackson, 1974)

$$\oint_S \mathbf{E} \cdot d\mathbf{S} = \frac{q}{\epsilon_0 \epsilon} \quad (39.1)$$

where E is the electric field in V/m, ϵ_0 ($= 8.854 \text{ pF/m}$) is the permittivity in vacuum, and ϵ ($= \epsilon_a / \epsilon_0$) is the relative permittivity, ϵ_a being the absolute dielectric permittivity of the isotropic medium or dielectric under the electric force field. For a parallel capacitor containing a dielectric of relative permittivity, ϵ , Equation (39.1) becomes

$$E = \frac{q}{\epsilon \epsilon_0 S} = \frac{\sigma}{\epsilon \epsilon_0} \quad (39.2)$$

where q and S are, respectively, the charge and area of the capacitor arm plates and σ the surface-charge density. The decrease of the electric field in the capacitor by effect of the dielectric is given by

$$E_0 - E = \frac{\sigma}{\epsilon_0} \left(\frac{\epsilon - 1}{\epsilon} \right) \quad (39.3)$$

The reduction of the electric field arising from the decrease of the charge density in the capacitor, called polarization, can be written as

$$P = \sigma(\varepsilon - 1)/\varepsilon \quad (39.4)$$

The polarization of capacitors is caused by the alignment of the individual dipoles of the dielectric medium with the electric field. A parallel plate capacitor with polarization P , area of each arm plate S , and distance between plates d has a dipole moment given by $\mu = \mathbf{P}(\mathbf{S} \times \mathbf{d}) = \mathbf{P} V$, where V is the volume of the dielectric. Then the polarization can be defined as the dipole moment of the dielectric per volume unit.

Taking into account Equations (39.2) and (39.3), the polarization of an isotropic dielectric under an electric field E can be written in SI units as

$$\mathbf{P} = \sum_i \mu_i = e_0(\varepsilon - 1)\mathbf{E} \quad (39.5)$$

where μ_i is the dipole moment associated with the dipole of the molecular entity i of the dielectric. Then the electric susceptibility is given by

$$\chi = P/E = e_0(\varepsilon - 1) \quad (39.6)$$

Another important magnitude is the dielectric displacement, \mathbf{D} , defined as the charge density σ in the capacitor. From Equation (39.2), it follows that

$$\mathbf{D} = e_0\varepsilon\mathbf{E} \quad (39.7)$$

For anisotropic dielectrics, the relative dielectric permittivity becomes a second-order tensor and \mathbf{P} and \mathbf{D} should be written as

$$P_i = e_0(\varepsilon_{ij} - 1)E_j \text{ and } D_i = e_0\varepsilon_{ij}E_j \quad (39.8)$$

The SI units of \mathbf{P} and \mathbf{D} are C/m^2 . Notice that in the cgs system the factor $1/(4\pi)$ should replace the symbol e_0 in the equations indicated above. The units of electrical charge, electric field, voltage and capacity in the cgs system, and their equivalences with the respective SI units are, respectively, statc ($1 \text{ C} = 3 \times 10^9 \text{ statc}$), dyne/statc ($1 \text{ N/C} = 3.3 \times 10^{-5} \text{ dyne/statc}$), statv ($1 \text{ V} = 3.3 \times 10^{-3} \text{ statv}$), and statF ($1 \text{ F} = 9.1 \times 10^{11} \text{ statF}$).

39.3 MEASUREMENTS OF DIELECTRIC PERMITTIVITY

Since $E = -\text{grad } V$, where V is the electrical potential, and bearing in mind that the capacity of a capacitor is $C = q/\Delta V$, Equation (39.2) shows that the capacity of capacitors having parallel and cylindrical geometries are, respectively, $C = e_0\varepsilon S/d$ and $C = 2\pi e_0\varepsilon L/\ln(r_2/r_1)$. In these expressions, S and d are, respectively, the area of a single arm plate and the distance between the plates for parallel capacitors, whereas r_1 and r_2 are, respectively, the radii of the internal and external cylinders of length L for capacitors with cylindrical geometry. Notice that if $r_2 - r_1 = d$ is very small, $\ln(r_2/r_1) = \ln[1 + (r_2 - r_1)/r_1] \cong (r_2 - r_1)/r_1$. Since $2\pi r_1 L = S$, where S is

the lateral area of the cylinder, the formula for the capacity of a cylindrical capacitor is also $\epsilon_0 \epsilon S/d$. The SI units of the voltage, electric field, and capacity are, respectively, V, V/m, and F. The relative dielectric permittivity ϵ of a dielectric is C/C_0 , where C and C_0 are the capacitance of a capacitor containing, respectively, the dielectric material and vacuum.

It is an experimental fact that most capacitors with dielectrics between their arm plates lose a fraction of energy under an alternating electric current. The simplest model for a capacitor with a lossy dielectric is a capacitor with a perfect dielectric ($\epsilon = 1$) in parallel with a resistor accounting for the power dissipation (Figure 39.1a). The admittance of the circuit under an alternate voltage $V = V_0 \exp(j\omega t)$, where ω is the angular frequency of the force field, is a complex quantity that can be written as $Y^* = 1/R - j\omega C$. Since the current intensity flowing in the circuit is $C_0 \epsilon^*(\omega) dV/dt = VY^*$, where ϵ^* is the relative complex permittivity and C_0 is the capacity of the capacitor in vacuum, the values of the real ϵ' and imaginary ϵ'' components of ϵ^* are given by

$$\epsilon' = \frac{C}{C_0}; \quad \epsilon'' = \frac{1}{\omega RC_0} \quad (39.9)$$

The loss $\tan \delta$ for the equivalent parallel circuit is given by

$$\tan \delta = \epsilon''/\epsilon' = (\omega RC)^{-1} \quad (39.10)$$

An alternative model to the equivalent parallel circuit is an ideal capacitor ($\epsilon = 1$) in series with a resistor (Figure 39.1b). After pertinent calculations carried out following the procedure outlined above, ϵ' and ϵ'' are expressed by

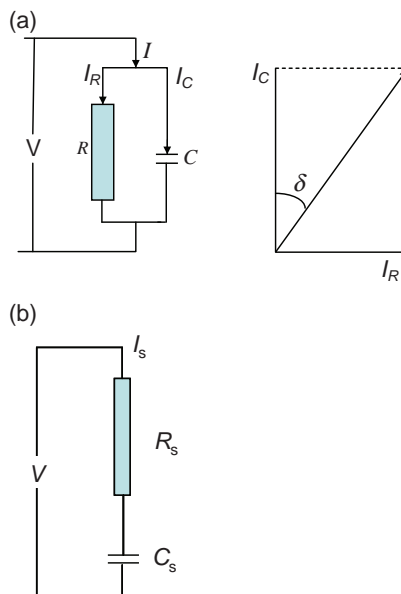


FIGURE 39.1 Schemes of the equivalent electric circuits for dielectrics: (a) in parallel and (b) in series.

$$\varepsilon' = \frac{C_s/C_0}{1 + \omega^2 R_s^2 C_s^2}$$

$$\varepsilon'' = \frac{\omega R_s C_s^2}{C_0 [1 + \omega^2 R_s^2 C_s^2]}$$
(39.11)

The subscript s reminds that the ideal capacitor and the ohmic resistance are in series in the circuit. The loss $\tan \delta_s$ can be written as

$$\tan \delta_s = \omega R_s C_s \quad (39.12)$$

The study of the electric behavior of the condensed matter in general, and polymers in particular, can be carried out in the frequency window $10^{-4} < f < 10^{11}$ Hz where $f = \omega/2\pi$ (Kremer and Arndt, 1997). However, this wide range of frequencies cannot be covered by a single apparatus. In the frequency range 10^{-4} –10 Hz, the permittivity can be measured using lumped circuit transient response. The dielectric permittivity in the intermediate frequencies range 10^{-1} – 10^6 Hz has traditionally been measured with an experimental device based on the Wheatstone bridge principle where the arms of the bridge are capacitance-resistance networks (Figure 39.2a).

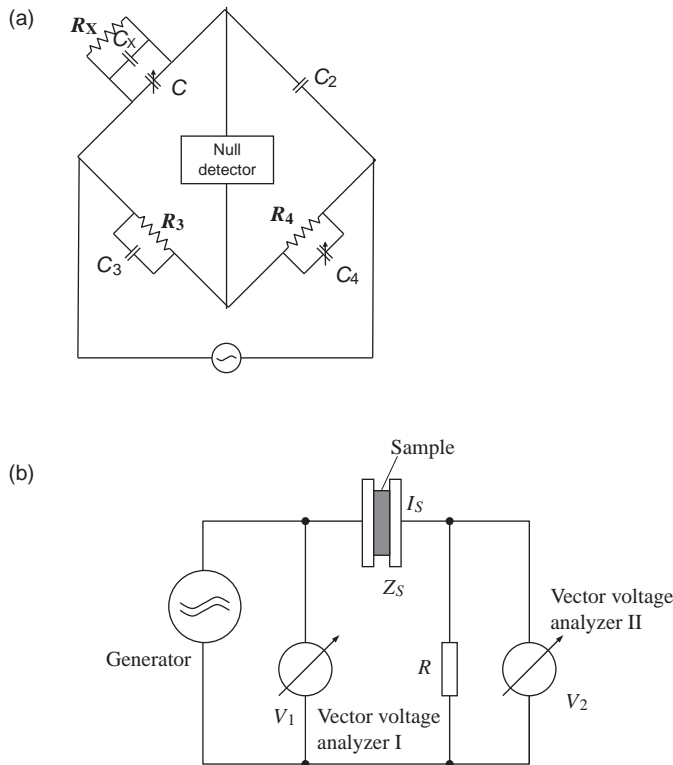


FIGURE 39.2 Schemes of (a) capacitance bridge (the resistance and capacitance of the equivalent circuit of the sample are indicated by R_x and C_x) and (b) impedance spectrometer.

The frequency range 10^{-2} – 10^6 Hz can also be covered utilizing impedance bridge devices. A scheme of a frequency response analyzer of this kind is shown in Figure 39.2b (Kremer and Arndt, 1997). An AC voltage V_1 is applied to the sample, and a resistor R or a current-to-voltage converter for low frequencies converts the sample current I_S into a voltage V_2 . By comparing the amplitude and the phase angle between these two voltages, the complex impedance of the sample is given by

$$Z_S^* = \frac{V_1 - V_2}{I_S} = \frac{V_1 - V_2}{V_2} R \quad (39.13)$$

For frequencies higher than 10^5 Hz, a buffer amplifier decouples the sample current of the analyzer and the current-to-voltage converter by means of an impedance variable. Owing to parasitic inductances, the high frequency limit is about 1 MHz.

At frequencies in the range 1 MHz–10 GHz, the inductance of the connecting cables contributes to the measured impedance. Resonance circuits have been used in the past in the range 10^5 – 10^8 Hz. The measurements performed in a broad range of temperatures require the lengths of the electric wires connecting the sample and the resonance circuit to be as short as possible. This problem can be circumvented by placing the sample capacitor as the termination of a high precision coaxial line.

A technique often used to obtain dielectric spectra at high frequencies is time domain reflectometry (Cole, 1975, 1977; Cole et al., 1980; Mopsik, 1984; Feldman, 1996). The technique is based on the reflection of an electric wave transported through a coaxial line in a dielectric sample cell attached to the end of the line. If the reflected wave corresponding to a step-pulse voltage incident signal is measured in the time domain, the ratio of the Fourier transforms of the reflected wave and the incident pulse is a function of the complex permittivity of the sample.

For frequencies above 1 GHz, the experimental devices move from interference optics to geometrical optics. Network analysis is used and the dielectric properties of materials from infrared to ultraviolet are studied in terms of the amplitude and phase of reflected and transmitted waves. A scheme of the range of frequencies covered by different experimental devices is shown in Figure 39.3.

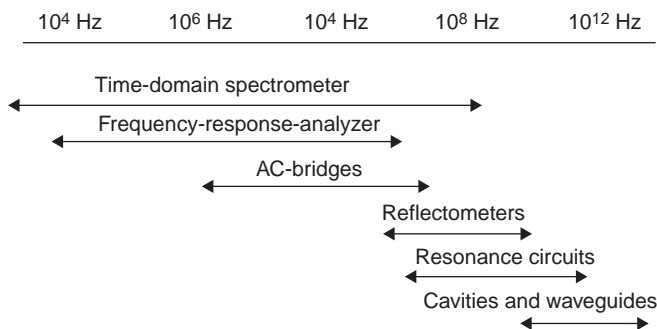


FIGURE 39.3 Experimental devices for dielectric spectroscopic measurements in different ranges of frequency.

39.4 POLARIZATION AND DIPOLE MOMENTS IN ISOTROPIC SYSTEMS

Let us consider a gas with randomly oriented N molecules per unit volume, each of them having permanent dipole moment μ_p . An electric field E produces a partial orientation of the permanent dipoles together with the separation of the residual charges in such a way that the dipole moment of the molecules is $\mu = \mu_p + \mu_d$, where μ_d arises from the residual charges displacement in the direction of the electric field. The parameter μ_d is related to E by $\mu_d = \alpha_d E$, where α_d is the molecular polarizability. In the cgs system, used in this section, the units of μ and α are, respectively, the Debye and cm^3 . The calculation of the dipolar contribution to the polarization was firstly proposed by Debye (1936). If E has the z -axis direction of the reference frame, the total polarization can be expressed by

$$\mathbf{P} = N(\mu_p \langle \cos \theta \rangle + \mu_d) \mathbf{k} \quad (39.14)$$

where \mathbf{k} is the vector unit along the z axis, and $\langle \cos \theta \rangle$ is the average of the cosine of the angle the permanent dipoles form with the electric field. This average can be obtained using the Boltzmann distribution for the dipoles at temperature T to yield $\langle \cos \theta \rangle = \mu_p E / 3k_B T$, where k_B is the Boltzmann constant. Then Equation (39.14) becomes

$$\mathbf{P} = N \left(\mu_p^2 E / 3k_B T + \mu_d \right) \mathbf{k} = N(\alpha_0 + \alpha_d) \mathbf{E}_1 \quad (39.15)$$

where $\alpha_0 = \mu_p^2 E / 3k_B T$ is the orientation polarizability and \mathbf{E}_1 is the true electric field acting on each molecule in the \mathbf{k} direction arising from the external electric field E together with the field E_{sph} produced by the charges surrounding the cavity where the molecule is located. Pertinent calculations find that $E_{\text{sph}} = (4/3)\pi P$ and the real field acting on the molecule in the \mathbf{k} direction is $E_1 = E + E_{\text{sph}}$. Since in the cgs system, the appropriate expression for Equation (39.6) is $\chi = P/E = (\epsilon - 1)/4\pi$, Equation (39.14) leads to

$$\frac{\epsilon - 1}{\epsilon + 2} = \frac{4\pi\rho N_A}{3M} (\alpha_0 + \alpha_d) = P_o + P_d \quad (39.16)$$

where ρ and M are, respectively, the density and molecular mass of the gas, and N_A is Avogadro's number. Moreover, $P_o = 4\pi\rho N_A \alpha_0 / 3M$ and $P_d = 4\pi\rho N_A \alpha_d / 3M$ are called, respectively, the orientation and induced molar polarizations. Notice that the molar polarization evidently differs from the polarization \mathbf{P} or dipole moment per volume unit. The parameter $\alpha_d (= \alpha_e + \alpha_a)$ arises from the distortion of the electronic cloud (electronic polarizability, α_e) and the small displacement of the atoms of the molecule (atomic polarizability, α_a) caused by the electric field. The molecular polarizability α_d can be obtained by writing $\alpha_0 = 0$ in Equation (39.16). Then the electronic polarizability α_e is usually calculated by utilizing the Maxwell relationship $\epsilon(\lambda) = n^2(\lambda)$. Owing to the fact that Equation (39.15) corresponds to a static electric field, n should be obtained at different wavelengths and its value extrapolated to $1/\lambda \rightarrow 0$. Except for silicon compounds, the parameter α_a is rather small, approximately $0.15\alpha_e$, and its contribution to the polarization is neglected. In these circumstances, Equation (39.16) can be written as

$$\frac{\varepsilon - 1}{\varepsilon + 2} - \frac{n^2 - 1}{n^2 + 2} = \frac{4\pi\rho N_A \mu_p^2}{3k_B T M} \quad (39.17)$$

This expression, known as Debye equation, holds for a variety of gases and vapors at ordinary pressures. However, Equation (39.17) presents some shortcomings, among them the prediction of a Curie temperature at which the material becomes ferroelectric. Moreover, Equation (39.17) fails for polar liquids. These facts led Onsager (1936) to reexamine the internal field in the cavity model introducing the concept of reaction field. Onsager succeeded in developing a fundamental modification of Debye's equation that gives a good account of the dipole moment of liquids in absence of real association between molecules. Kirkwood (1939) and Fröhlich (1948, 1958) refined even more Debye's formalism by taking into account short-range interactions between neighboring molecules.

By separating polar molecules from one another in the liquid state with nonpolar ones, the dielectric behavior of the solution resembles that of a gaseous condition. In this situation, the Debye model can be used to obtain the dipole moment of isolated polar molecules from their solutions in nonpolar solvents such as benzene, toluene, cyclohexane, and so on. In this case Equation (39.16) becomes (Guggenheim, 1949, 1951; Smith, 1950)

$$\mu_p^2 = \frac{27k_B T M}{4\pi\rho N_A (\varepsilon_1 + 2)^2} \lim_{w \rightarrow 0} \left(\frac{d\varepsilon}{dw} - \frac{dn^2}{dw} \right) \quad (39.18)$$

where M is the molecular weight of the polymer, ε and n are, respectively, the static dielectric permittivity and the index of refraction of the solution in which the weight fraction of polymer is w ; ρ and ε_1 are, respectively, the density and the dielectric permittivity of the solvent whereas N_A and k_B are, respectively, Avogadro's number and Boltzmann's constant. Notice that using cgs units in Equation (39.18), μ_p is given in Debye's equation.

Whereas rigid simple molecules have similar permanent dipole moments, flexible oligomers and polymers are continuously changing their spatial conformations. Actually, molecular chains are made up of thousands and even hundred of thousands of skeletal bonds. Each skeletal bond fluctuates in energy wells separated from contiguous wells by barrier energies. The energy wells are associated with discrete rotational states. By effect of the thermal energy $k_B T$, a fluctuating skeletal bond can overcome the barrier energy that separates it from another rotational state and a conformational transition takes place. Then conformational transitions can be considered to be produced by discrete rotations about skeletal bonds. For most skeletal bonds, the rotational states with minimum energy are *trans* (t), *gauche* positive (g^+), and *gauche* negative (g^-) (Figure 39.4). Accordingly, a

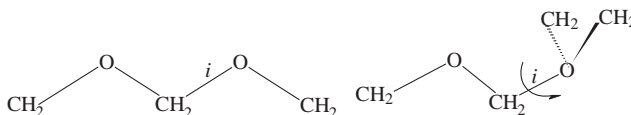


FIGURE 39.4 Conformations arising from rotations about skeletal bond i : *trans* (t , all the methylene groups are on the plane), *gauche* positive (g^+ , the last methylene group located behind the plane), and *gauche* negative (g^- , the last methylene group in front of the plane).

chain of N skeletal bonds has accessibility to 3^{N-2} spatial conformations, a nearly unlimited number of spatial conformations for large values of N . The dipole moment associated with each conformation may differ from that of another one, and μ_p^2 obtained for polymers using Equation (39.18) is an average value, currently written as $\langle \mu^2 \rangle$, where bracket angles mean average. Mean-square dipole moments obtained for polymers and oligomers using Equation (39.18), also known as the equation of Guggenheim (1949, 1951) and Smith (1950) show consistency among them, presumably because intra-molecular dipole-dipole interactions fade away among dipoles separated by four or more skeletal bonds. Then Equation (39.18) becomes one of the most reliable methods that can be used to determine mean-square dipole moments for isolated molecular chains.

Dipoles in molecular chains can be rigidly attached to the skeletal bonds and to flexible side chains. In the first case, dipoles can be classified into two types: (a) dipoles parallel to the chain contour (type A dipoles) and (b) dipoles bisecting skeletal bond angles, and therefore perpendicular to the chain contour (type B dipoles). Dipoles separated from the main chain by flexible segments are called type C dipoles (Stockmayer, 1967; Stockmayer and Burke, 1969). Dipoles of type A are correlated with the end-to-end distance of the chain, \mathbf{r} , in such a way that $\langle \boldsymbol{\mu} \times \mathbf{r} \rangle \propto \langle r^2 \rangle$, where $\langle r^2 \rangle$ is the mean-square end-to-end distance of the chains. As a consequence, the mean-square dipole moment measured in solution for this type of polymers is subject to the same excluded volume effects as $\langle r^2 \rangle$. For macromolecules with B and C type dipoles $\langle \boldsymbol{\mu} \times \mathbf{r} \rangle = 0$, and therefore, $\boldsymbol{\mu}$ and \mathbf{r} are uncorrelated parameters. This means that the mean-square dipole moment measured in solution does not present excluded volume effects for polymers containing either B or C type dipoles. The dipoles of most polymers are of types B or C.

The mean-square dipole moment of a chain of N skeletal bonds is given by

$$\langle \mu^2 \rangle = \langle \boldsymbol{\mu}_i \times \boldsymbol{\mu}_j \rangle = \left\langle \sum_i \mathbf{m}_i \times \sum_j \mathbf{m}_j \right\rangle = \sum_i m_i^2 + 2 \left\langle \sum_i \sum_{j < i} m_i m_j \cos \theta_{ij} \right\rangle \quad (39.19)$$

where m_i and m_j represent, respectively, the dipole moments of the skeletal bonds i j , whereas θ_{ij} is the angle between these two bonds. Dipole moments associated with skeletal or groups of skeletal bonds can be obtained from simple molecules. For example, from the dipole moments of dimethyl ether and dimethyl thioether, the dipole moments of ether ($\text{CH}_2\text{—O}$), thioether ($\text{CH}_2\text{—S}$), and $\text{CH}_2\text{—CH}_2$ skeletal bonds are found to be, respectively, 1.07, 1.21, and 0.00 D. For polyvinyl chloride, polyvinyl bromide and aliphatic polyesters, the dipole moments of ethyl chloride, ethyl bromide, and methyl acetate are used as dipole moments of the respective structural units. Detailed information concerning dipole moments of low molecular weight molecules used as models for molecular chains can be found elsewhere (McClellan, 1974).

For a freely jointed chain $\langle \cos \theta_{ij} \rangle = 0$ and $\langle \mu^2 \rangle = \sum_i m_i^2$. In general mean-square dipole moments are expressed in terms of the dipole moment ratio $g = \langle \mu^2 \rangle / \sum_i m_i^2$. The parameter g such as the characteristic ratio $\langle r^2 \rangle \langle r^2 \rangle_0 / \sum_i l_i^2$, where $\langle r^2 \rangle_0$ and l_i are, respectively, the unperturbed mean square end-to-end distance of the chains and the length of the skeletal bond i , depends on molecular weight for short chains. However, both ratios are independent on molecular weight for large chains.

Both the dipole moment ratio and its temperature dependence expressed as $d \ln \langle \mu^2 \rangle / dt$ may be very sensitive to the chemical structure of polymer chains (Riande and Mark, 1978). For example, polyoxymethylene (POM) exhibits a low polarity ($g = 0.2$) whereas syndiotactic polyvinyl chloride (PVC) presents a rather high polarity ($g \cong 4$)

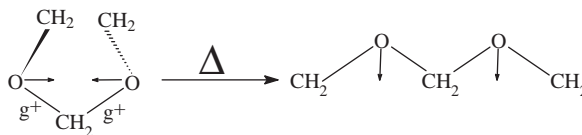


FIGURE 39.5 Conformational transition $g^+g^+ \rightarrow tt$ for four oxymethylene skeletal bonds. Arrows indicate dipoles associated with CH_2OCH_2 skeletal bonds.

(Riande and Saiz, 1992; Díaz-Calleja and Riande, 1997). The low value of g for POM arises from the fact that *gauche* states of the same sign are favored over the alternative *trans* states. In the former case, dipoles are in antiparallel direction (low polarity), whereas in the latter are in parallel direction (high polarity). An increase in temperature causes conformational transitions from *gauche* to *trans* states and, as a result, the dipole ratio of POM chains exhibit an unusually high temperature dependence, of the order of $6 \times 10^{-3} \text{ K}^{-1}$ (Figure 39.5) (Riande and Mark, 1978).

The stereochemical composition of vinyl polymers containing polar groups in their structure affects the total polarity of the chains (Mark, 1971, 1972; Blasco Cantera et al., 1981; Saiz et al., 1982; Salmerón et al., 1984). For example, syndiotactic PVC is formed by $ttt \dots t$ sequences occasionally separated by *gauche* states. In $ttt \dots t$ conformations the dipole moments associated with the repeating units of syndiotactic PVC are in nearly parallel direction and as a result the dipole moment ratio of the polymer is rather high, of the order of 4. However, isotactic PVC is formed by $tg^+tg^+ \dots tg^+$ (or $tg^-tg^- \dots tg^-$) sequences, occasionally separated by *trans* states. In these conformations, the dipoles associated with the repeating units of the isotactic chains are in nearly opposite direction; hence, the low dipole moment ratio of isotactic PVC is rather low, of the order of 0.5.

The Rotational isomeric state (RIS) (Volkenstein, 1963; Flory, 1969) model, which assumes that each skeletal bond has accessibility to a limited series of rotational angles of minimum energy, has proved to be a useful tool to predict the polarity of molecular chains as a function of the chemical structure (Suter and Mattice, 1994). Also, the model is useful to determine the conformational energy of *gauche* states relative to the corresponding *trans* states by comparing experimental mean-square dipole moments with theoretical ones calculated using the RIS model. For illustrative purposes, the variation of the dipole moment ratio with the energy of *gauche* states about $\text{OCH}_2\text{—OCH}_2$ bonds in the case of alternating copolymers of methylene oxide and ethylene oxide are shown in Figure 39.6 (Riande and Saiz, 1992) Notice that as the energy decreases the fraction of low polarity $g^\pm g^\pm$ conformations increases at expenses of the high polarity tt conformations (see Figure 39.5). Good agreement between theoretical and experimental results is found for values of $E(g^\pm) - E(t) = -1.2 \text{ kcal/mol}$.

39.4.1 Dielectric Relaxations

In response to electric field perturbations, the dipoles of polar molecules of liquids rotate toward the direction of the field until an equilibrium distribution of the dipoles around the direction of the field is accomplished. The dielectric polarization decreases with the molecular size, the viscosity of the medium and the frequency of the electric field. All these variables hinder the orientation of the molecules in the field. As a result, the

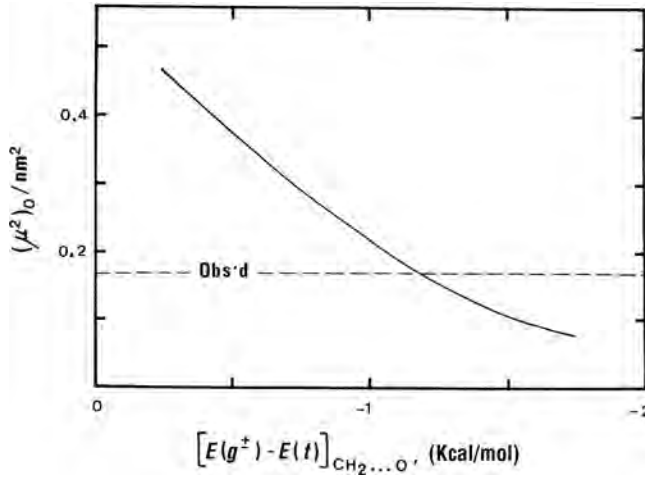


FIGURE 39.6 Dependence of the dipole moment ratio of poly(1,3-dioxane) on the energy of gauche states relative to trans about oxymethylene skeletal bonds (Riande and Mark, 1978).

polarization is a complex quantity with a capacitance perpendicular to the field, and another component, in phase with the field, accounting for the displacement current that gives rise to thermal dissipation of energy. Theories have been developed (Fatuzzo and Mason, 1967; Titulaer and Deutch, 1974) relating the polarization with the frequency force field, but here only a phenomenological approach will be described.

In linear conditions, experimental evidence (Riande and Díaz-Calleja, 2005) indicates that the dielectric displacement under a constant force field, E , varies with time as $D(t) = e_0[\varepsilon_\infty + (\varepsilon_0 - \varepsilon_\infty)\varphi(t)]E$, where $\varphi(t)$ is a normalized continuous monotonously increasing function or buildup function that describes the orientation of the dipoles with time, whereas ε_0 and ε_∞ are the dielectric permittivities at time ∞ and zero, respectively. The limit values of $\varphi(t)$ are $\varphi(0)=0$ and $\varphi(\infty)=1$. In linear systems, the Boltzmann superposition principle holds so that the total dielectric displacement at time t caused by a variable electric field $E(\theta)$ can be written as (Riande and Díaz-Calleja, 2005; MacCrum et al., 1967)

$$D(t)/e_0 = \int_{-\infty}^t \varepsilon(t-\theta)[dE(\theta)/d\theta]d\theta = \varepsilon_\infty E(t) + (\varepsilon_0 - \varepsilon_\infty) \int_0^\infty E(t-u)[d\varphi(u)/du]du \quad (39.20)$$

Under an alternating electric field $E(t)=E_0 \operatorname{Im} \exp(j\omega t)$, where ω is the angular frequency of the electric field, Equation (39.20) leads to

$$\varepsilon^*(\omega) = D(t)/(e_0 E(t)) = \varepsilon_\infty + (\varepsilon_0 - \varepsilon_\infty) \int_0^\infty \exp(-j\omega u)[-d\varphi(u)/du]du \quad (39.21)$$

In this expression, $\varphi(t)$ is a continuous monotonously decreasing function or memory function whose limit values are $\varphi(0)=1$ and $\varphi(\infty)=0$; moreover, $d\varphi(t)/dt + d\varphi(t)/dt = 0$. Hence, Equation (39.21) can be rewritten in the following way

$$\frac{\varepsilon^*(\omega) - \varepsilon_\infty}{\varepsilon_0 - \varepsilon_\infty} = L \left[-\frac{d\phi(t)}{dt} \right] \quad (39.22)$$

where the symbol L represents the Laplace transform (Williams, 1979). Notice that ε_0 and ε_∞ in the frequency domain represent the relaxed ($\omega = 0$, $t \rightarrow \infty$) and unrelaxed ($\omega \rightarrow \infty$, $t = 0$) dielectric permittivities. The right-hand side of Equation (39.22) is $\phi(0) - s\phi(s) = 1 - s\phi(s)$, where $\phi(s)$ is the Laplace transform of $\phi(t)$. Then Equation (39.22) can be written as $f(\omega) = 1 - s\phi(s)$, where $f(\omega)$ represents the left-hand side of this equation. In the complex plane, $s = j\omega$ and taking into account that $\phi(t) = \phi^{-1}(s)$, the memory function can be expressed in terms of the complex permittivity by

$$\phi(t) = \frac{1}{2\pi} \lim_{R \rightarrow \infty} \int_{-R}^R \frac{\varepsilon_0 - \varepsilon^*(\omega)}{\varepsilon_0 - \varepsilon_\infty} \exp(j\omega t) d \ln \omega \quad (39.23)$$

For a simple system defined by a single relaxation time, the rate of return to equilibrium of an observable property depending on time, $X(t)$, is proportional to the distance of the observable from its value at equilibrium represented by X_{eq} . As a first order approximation, $(X(t) - X_{eq})/(X(0) - X_{eq}) = \exp(-kt)$, where the proportionality constant k must be the reciprocal of time, that is, $k = 1/\tau$, τ being the so-called relaxation time. The relaxation time is defined as the time necessary for the observable quantity X to reduce $1/e$ the value it had at $t = 0$. Then, the time dependence of the normalized relaxation function with a single relaxation time is $\phi(t) = \exp(-t/\tau)$. By substituting this expression into Equation (39.22), the real and loss components of the complex dielectric permittivity for systems with a single relaxation time or Debye systems are obtained as

$$\varepsilon'(\omega) = \varepsilon_\infty + (\varepsilon_0 - \varepsilon_\infty)/(1 + \omega^2 \tau^2) \quad (39.24)$$

$$\varepsilon''(\omega) = (\varepsilon_0 - \varepsilon_\infty)\omega\tau/(1 + \omega^2 \tau^2) \quad (39.25)$$

According to Equation (39.24), the real component of the complex dielectric permittivity is a continuous decreasing function of ω , its limit values being ε_0 and ε_∞ at $\omega \rightarrow 0$ and $\omega \rightarrow \infty$, respectively. Illustrative plots for the components of the complex permittivity are shown as a function of $\omega\tau$ in Figure 39.7a. Three regions can be observed in the isotherm corresponding to ε' . In the high frequency region, ε' undergoes only a slight increase as frequency decreases. This region is followed by another one where ε' substantially increases with decreasing frequency. In the low frequencies region, ε' is nearly independent on frequency. The dielectric loss ε'' is an absorption that reaches a maximum at $\omega\tau = 1$, just at the inflexion point of the ε' versus $\omega\tau$ curve. The values of ε'' at the frequency limits ($\omega \rightarrow 0$, and $\omega \rightarrow \infty$) are zero.

The so-called complex electric modulus $M^*(\omega)$ is the reciprocal of the complex permittivity. Its components expressed in terms of those of $\varepsilon^*(\omega)$ are

$$M'(\omega) = \varepsilon'(\omega)/[\varepsilon'^2(\omega) + \varepsilon''^2(\omega)]; \quad |M''(\omega)| = \varepsilon''(\omega)/[\varepsilon'^2(\omega) + \varepsilon''^2(\omega)] \quad (39.26)$$

Plots with the components of $M^*(\omega)$ for a material with a single relaxation time are presented in Figure 39.7b. In the low frequency region, the value of the real component M'

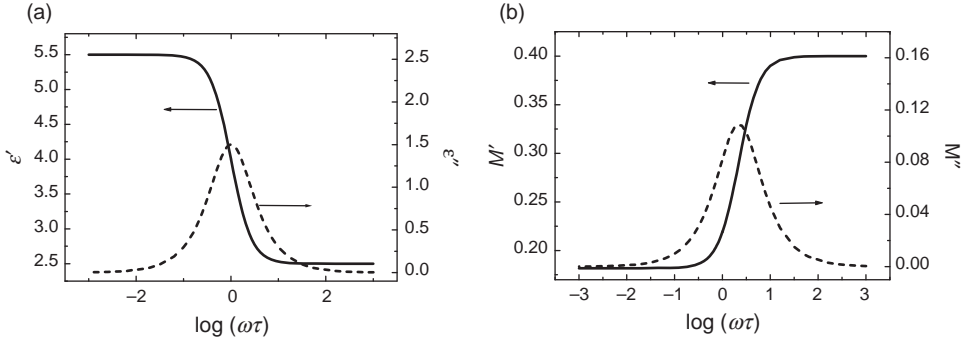


FIGURE 39.7 (a) Real and imaginary components of the complex dielectric permittivity ϵ^* against $\log \omega\tau$ for a system described by the following parameters: $\epsilon_0 - \epsilon_\infty = 3$, and $\epsilon_\infty = 2.5$. (b) Components of the complex relaxation modulus M^* for the same system.

slightly increases with increasing frequency until a frequency is reached at which the modulus undergoes a sharp increase, remaining nearly constant in the high frequencies region. On the other hand, the shape of M'' versus $\omega\tau$ is similar to that of ϵ'' reaching a maximum at the inflexion point of the curve M' versus $\omega\tau$. It is worth noting that the shapes of ϵ' and M' remind those of the creep compliance function (J') and relaxation modulus (G') in mechanical dynamic experiments.

For most systems, including low molecular weight liquids and polymers, dielectric dispersions in the frequency domain cover more than the 2.29 decades predicted by Debye systems. This fact suggests that a multitude of dielectric mechanisms i may intervene in the relaxations of real systems, each of them associated with a different relaxation time τ_i . Accordingly, the normalized decay function should be expressed as a weighted sum of exponential decay functions, $\phi(t) = \sum_i w_i \exp(-t/\tau_i)$ where w_i is the mass fraction of the mechanism i and $\sum_i w_i = 1$. Then Equation (39.22) can be written as

$$[\epsilon^*(\omega) - \epsilon_\infty]/(\epsilon_0 - \epsilon_\infty) = \sum_i w_i / (1 + \omega^2 \tau_i^2) \quad (39.27)$$

The integral analogue of Equation (39.27) involves a continuous distribution of retardation times $L(\ln \tau)$, and Equations (39.24) and 39.25 become

$$[\epsilon'(\omega) - \epsilon_\infty]/(\epsilon_0 - \epsilon_\infty) = \int_{-\infty}^{\infty} L(\ln \tau) d\ln \tau / (1 + \omega^2 \tau^2) \quad (39.28)$$

$$\epsilon''(\omega)/(\epsilon_0 - \epsilon_\infty) = \int_{-\infty}^{\infty} L(\ln \tau) \omega \tau d\ln \tau / (1 + \omega^2 \tau^2) \quad (39.29)$$

These equations can directly be obtained by taking into account that a continuous monotonously decreasing function of time, such as the memory function $\phi(t)$, can be expressed in terms of the Laplace transform of an unknown function $N(s)$ (see Riande and Díaz-Calleja, 2005 for details). Following the same strategy, the dielectric permittivity in the time domain can be written in terms of the retardation spectrum as

$$\varepsilon(t) = \varepsilon_{\infty} + (\varepsilon_0 - \varepsilon_{\infty}) \int_{-\infty}^{\infty} L(\ln \tau) (1 - e^{-t/\tau}) d \ln \tau \quad (39.30)$$

In Equations (39.28–39.30), $L(\ln \tau)$ is the normalized retardation times spectrum. It has been emphasized that the numerical fitting of weighted sum of different exponential decays to the decay function does not necessarily mean that a distribution of relaxation times really exists. The substitution of a decay function with a stretch exponent $\bar{\beta}$ lying in the range $0 < \bar{\beta} \leq 1$ into Equation (39.22) also gives a good account of relaxation processes of complex molecules (Williams and Watts, 1971).

The loss component of the complex dielectric permittivity for amorphous polymers, at a given frequency, can be represented as a function of temperature. The resulting isochrone presents a prominent absorption or α relaxation associated with the glass transition temperature T_g , named glass–rubber relaxation. Below T_g , one or more secondary absorptions appear, which in decreasing order of temperature are called β , γ , δ , . . . (Figure 39.8) (Diaz-Calleja et al., 1992). The dipoles of polymers, such as poly(propylene oxide), poly(propylene sulfide), and trans poly(*cis*-1,4 isoprene), have a component of the dipole associated with the repeat unit parallel to the chain contour. As result, a relaxation appears associated with chains length, named normal mode process, located at higher temperature than the α relaxation (Figure 39.9) (Boese and Kremer, 1990). As

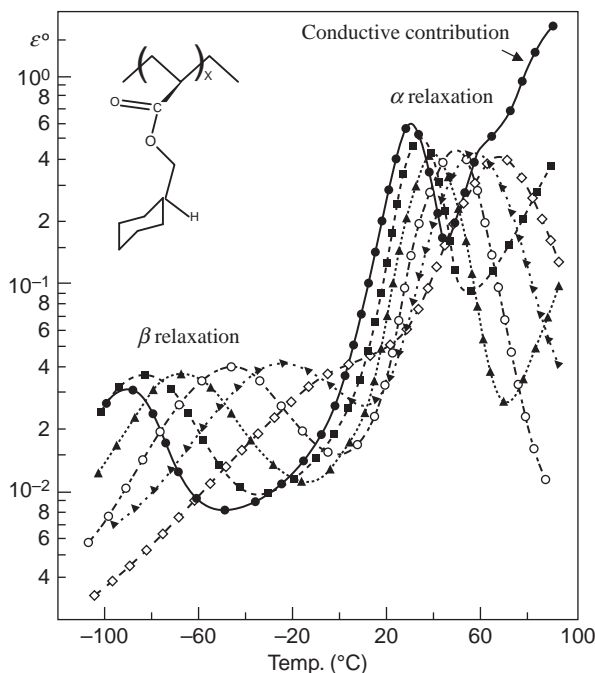


FIGURE 39.8 Dependence of the dielectric loss on temperature for poly(cyclohexyl acrylate) at several frequencies: (●) 0.1, (◆) 1, (▲) 10, (○) 10^2 , (▼) 10^3 , and (◇) 10^4 Hz. The repeat unit of the polymer is shown in the figure (Diaz-Calleja et al., 1992).

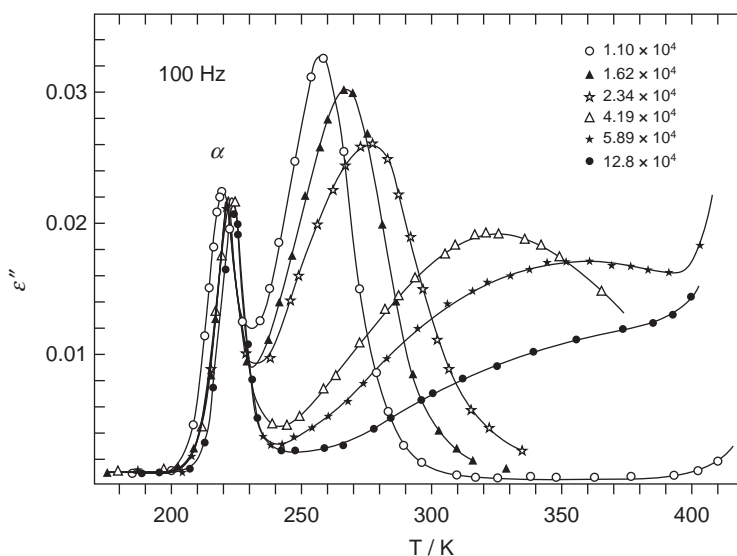


FIGURE 39.9 Isochrones, at 100 Hz, showing the dependence of the dielectric loss on temperature for several fractions of poly(cis-1,4-isoprene), the molecular weights of which are given in the figure (Boese and Kremer, 1990).

molecular weight increases, the location of the normal mode process in the isochrones shifts to higher temperatures, whereas the location of the glass–rubber relaxation and secondary processes are not affected by molecular weight. It is worth noting that the statement “the location of the α relaxation is independent on molecular weight” only holds if the molecular weight of the chains is high enough so that the T_g of the system is independent on molecular weight.

For polymers in the glassy state, the imaginary component or dielectric loss ϵ'' of the complex permittivity in the frequency domain exhibits one or more peaks associated with secondary or subglass relaxation processes named in increasing order of frequency β , γ , δ Above T_g , an ostensible absorption in most systems appears, just in the frequency region where a significant change in the value of ϵ' occurs, named α or glass–rubber relaxation. The α relaxation is followed at higher frequencies by the secondary absorptions, also detected in the glassy state, called in increasing order of frequency β , γ , δ An increase in temperature shifts the locations of the α relaxation and the secondary absorptions to higher frequencies. As the activation energy associated with the α relaxation is higher than that of the β , the distance between the peaks of both relaxations decreases with increasing temperature until a temperature is reached at which a single relaxation, named $\alpha\beta$ relaxation, appears. This behavior is presented in Figure 39.10 where dielectric loss in the frequency domain for poly(5-ethyl-1,3-dioxacyclohexane acrylate) (Huang et al., 2002), at several temperatures, is shown. Below the glass transition temperature of the polymer, 285 K, the isotherms show a single relaxation, named β process. However, at temperatures slightly above T_g , the isotherms present the α relaxation well separated from the β process appearing at higher frequency. As temperature increases, the intensity of the β relaxation increases and that of the α decreases. Moreover, the degree of overlapping between the α and β relaxations increases forming the $\alpha\beta$ relaxation in the vicinity of 343 K.

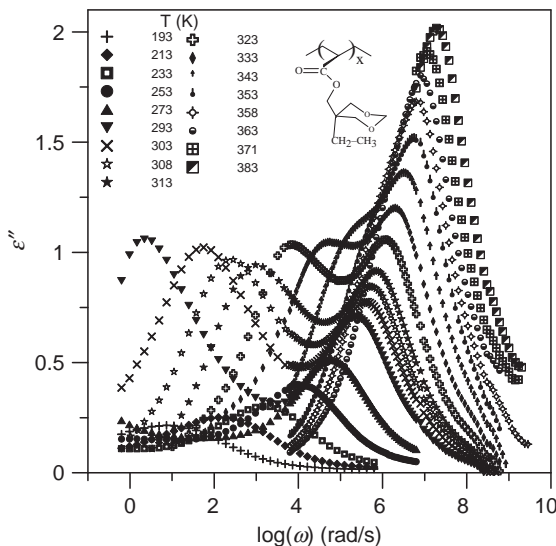


FIGURE 39.10 Dielectric loss in the frequency domain for poly(5-ethyl-1,3-dioxacyclohexane acrylate) at the temperatures indicated in the figure. The isotherms below 293 K correspond to the polymer in the glassy state. The T_g of the polymer lies in the vicinity of 290 K. The chemical structure of the repeat unit of the polymer is shown in the figure (Huang et al., 2002).

In the low frequency side of the α relaxation, the dielectric loss in polymeric systems may undergo a relevant increase with decreasing frequency caused by humidity and ionic impurities in the materials. In this case, the ionic contribution to the dielectric loss becomes dominant, and the scaling law $\varepsilon''(\omega) \sim \omega^{-1}$ holds. To account for this ionic conductive contribution, Equation (39.30) should be written as

$$\varepsilon''(\omega) = (\varepsilon_0 - \varepsilon_\infty) \int_{-\infty}^{\infty} L(\ln \tau) \omega \tau / (1 + \omega^2 \tau^2) + \sigma / (e_0 \omega) \quad (39.31)$$

where σ is the ionic conductivity in S/m. The ionic conductivity contribution to the dielectric loss is negligible at high frequencies.

In the low frequency region, the loss spectra of polymers containing dipoles of type A or type AB display the normal mode relaxation associated with chains disentanglement and flow (Adachi and Kotaka, 1993). As shown in Figure 39.11, the normal mode at a given temperature shifts to lower frequencies as molecular weight increases, while the location of the α relaxation in the loss spectra is nearly independent on molecular weight.

The energy dissipated per cycle and per volume unit by the dipoles under the electric field is $\pi \omega E_0^2 \varepsilon''$, where E_0 is the amplitude of the alternate electric field.

Owing to the fact that a Debye type process in the relaxation spectra extends over more two decades in the frequency domain, whereas it is a Dirac delta in the time domain, the retardation time spectra $L(\ln \tau)$ are better defined than the loss relaxation spectra in the frequency domain. This behavior can be observed in Figure 39.12 where the loss dielectric spectra and the retardation time spectra for poly(3-methylbenzyl methacrylate),

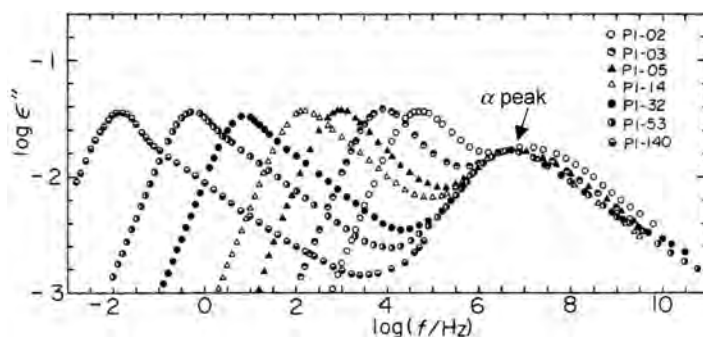


FIGURE 39.11 Dependence of the dielectric loss on frequency for fractions of poly(cis-1,4-isoprene) of different molecular weights. The peak associated with the normal process shifts to lower frequencies as molecular weight increases (Adachi and Kotaka, 1993).

poly(3-fluorobenzyl methacrylate), and poly(3-chlorobenzyl methacrylate) are shown. The better definition of the retardation spectra facilitates the deconvolution of overlapped α and β relaxations (Domínguez-Espinosa et al., 2005, 2006a, 2008; Diaz-Calleja et al., 2007; Alvarez et al., 2005). Since the buildup function $\varphi(t)$ corresponds to the integral of Equation (39.30) and taking into account that $\phi(t) + \varphi(t) = 1$, the normalized memory function can be obtained from the non-normalized retardation time spectra by means of the following expression

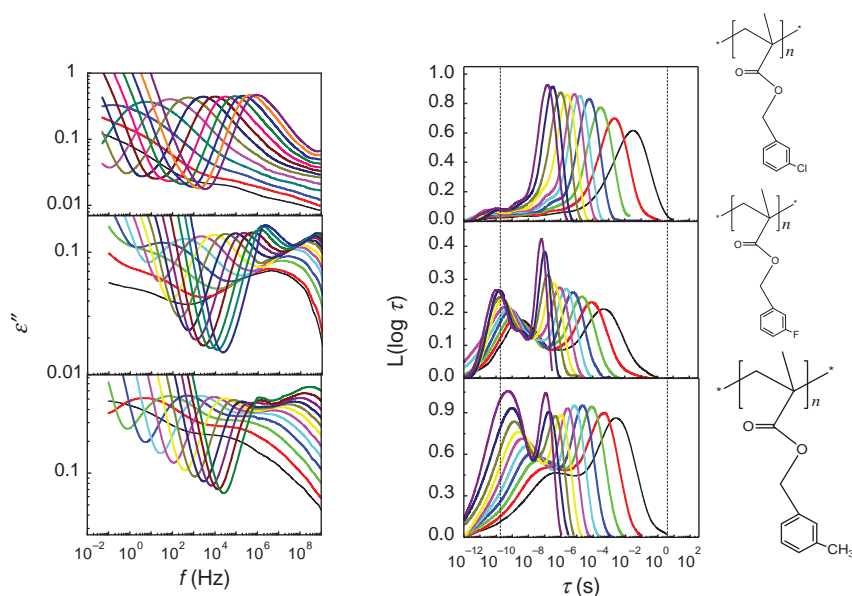


FIGURE 39.12 Dielectric loss spectra (left) and retardation time spectra (right) for poly(3-methylbenzyl methacrylate) (top), poly(3-fluorobenzyl methacrylate) (middle), and poly(3-chlorobenzyl methacrylate) in the temperature ranges 283–403, 313–423, and 293–403 K, respectively, at 10 K steps (Domínguez-Espinosa et al., 2005, 2006a; Diaz-Calleja et al., 2007; Alvarez et al. 2005).

$$\phi(t) = \frac{\int_{-\infty}^{\infty} L(\ln \tau) e^{-t/\tau} d \ln \tau}{\int_{-\infty}^{\infty} L(\ln \tau) d \ln \tau} \quad (39.32)$$

The shapes of both the glass–rubber relaxation and secondary processes in the frequency domain are described by empirical Havriliak–Negami type equations (Havriliak and Negami, 1987; Havriliak and Havriliak, 1997)

$$[\varepsilon^*(\omega) - \varepsilon_{\infty}]/(\varepsilon_0 - \varepsilon_{\infty}) = [1 + (\omega\tau)^b]^{-a} \quad (39.33)$$

The parameters a and b are shape parameters that lie in the range $0 < a, ab \leq 1$. For secondary relaxations $a = 1$ and the complex plots are arcs. For glass–rubber relaxations, the complex ε'' versus ε' plots are skewed arcs that fit to a straight line intersecting with the abscissa axis in the high frequency region. As the complexity of the relaxation system increases, the values of the shape parameters decrease.

The temperature dependence of the secondary absorptions obeys to Arrhenius behavior. The Arrhenius plots for the β absorptions of many flexible polymers lie in straight lines that intercept the ordinate axis at $f_0 = 1/2\pi\tau_0 \cong 13.5$. Then the activation energy can be estimated as $E_a \cong 0.258 T_{\max}$, where T_{\max} is the absolute temperature at the peak maximum of the β relaxation, at 1 Hz (Heijboer, 1972). Accordingly, the higher the activation energy of the β absorption, the higher the temperature at which the relaxation is centered.

The glass–rubber relaxation is governed by the volume and its temperature dependence is described by the Vogel–Fulcher–Tammann–Hesse (VFTH) equation (Vogel, 1921; Fulcher, 1925; Tammann et al., 1926)

$$\log \tau_{\max} = -\log(1/2\pi f_{\max}) = A + m/(T - T_V) \quad (39.34)$$

where T_V is the Vogel temperature, about 50 K below T_g , considered as the temperature at which the configurational entropy of the glassy system vanishes. Illustrative plots showing the temperature dependence of the α and β relaxations for poly(5-ethyl-1,3-dioxacyclohexane acrylate) (Huang et al., 2002) are shown in Figure 39.13. Notice that the activation energy associated with the β process above and below the glass transition temperature seems to have the same value. The temperature dependence of the ionic conductivity follows Arrhenius behavior and the activation energy associated with this process is rather high (about 30 kcal/mol).

The normalized glass–rubber relaxation in the time domain inevitably obeys to a stretch exponential, called the Kohlrausch–Williams–Watts (KWW) equation (Williams, 1979; Williams and Watts, 1971; Kohlrausch, 1847).

$$\phi_{\alpha}(t) = \exp[-(t/\tau_{\text{KWW}})^{\bar{\beta}}]; \quad 0 < \bar{\beta} \leq 1 \quad (39.35)$$

where $\bar{\beta}$ is the stretch exponent and τ_{KWW} is a characteristic relaxation time whose temperature dependence obeys to Equation (39.34). The lower $\bar{\beta}$, the higher the complexity of the relaxation process is. Illustrative plots showing the normalized memory functions for the total ($\phi(t)$) and $\alpha(\phi_{\alpha})$ relaxations of poly(2,4-difluorobenzyl methacrylate) are shown in Figure 39.14 (Domínguez-Espinosa et al., 2006a). The

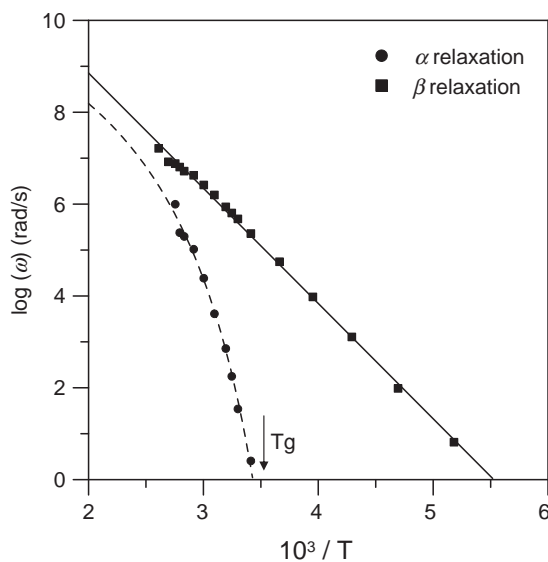


FIGURE 39.13 Temperature dependence of the relaxation times associated with the α and β relaxations for poly(5-ethyl-1,3-dioxacyclohexane acrylate) (Huang et al., 2002).

decrease of $\phi(t)$ at short times is caused by the β relaxation and that taking place at longer time arises from the glass–rubber relaxation. Notice that whereas the memory function corresponding to the glass–rubber relaxation is described by Equation (39.35), the total memory function is not.

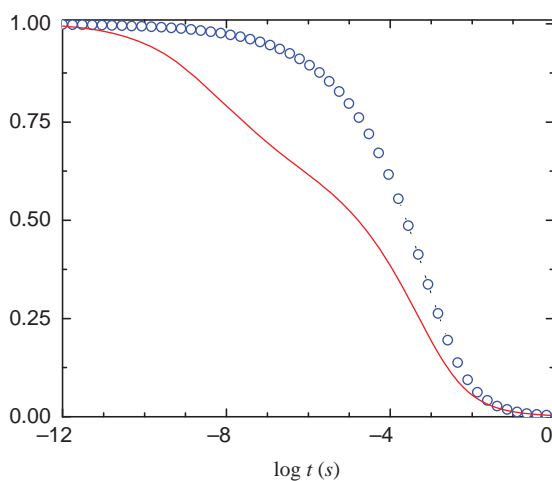


FIGURE 39.14 Memory relaxation function for the whole spectrum ($\phi(t)$, continuous line) and the α relaxation ($\phi_\alpha(t)$, open circles) for poly(2,4-difluorobenzyl methacrylate) (Domínguez-Espinosa et al., 2006a).

The fact that the relaxation times associated with secondary or subglass absorptions, such as the β process, are independent on molecular weight suggests that these absorptions arise from motions in the side groups of the molecular chains. For polymers without side groups, such as polyvinyl chloride and poly(oxyethylene), secondary absorptions are produced by local motions in the main chain. The mean relaxation time associated with the glass–rubber relaxation is also independent on the chains length for moderate and high weight polymers in which the concentration of chain ends is negligible. This behavior suggests that the glass–rubber relaxation is produced by cooperative micro-Brownian motions, or segmental motions, involving a significant number of skeletal bonds (~ 40 – 50 in flexible polymers) (Adachi and Kotaka, 1993; Schonhals, 1997; Kremer, 2003; Ferry, 1980). The secondary β absorption and the α relaxation are general phenomena in dynamics of supercooled liquids, independently of their molecular weight. The relaxation time associated with the dielectric normal mode process scales with molecular weight as $\tau \sim M^{3.4}$ for $M > M_c$ where $M_c \cong 2M_e$, M_e being the molecular weight between entanglements. For $M < M_c$, $\tau \sim M$ (Schonhals, 1997; Kremer, 2003; Ferry, 1980). The temperature dependence of the normal mode process is also described by the VFTH equation.

The dielectric strength of relaxation processes can be obtained from Equation (39.23) taking into account that $\phi(0) = 1$. The pertinent expression is $\epsilon_{0,i} - \epsilon_{\infty,i} = (2/\pi) \int_{-\infty}^{\infty} \epsilon''(\omega) d \ln \omega$, where the subscript i represents the type of relaxation α , β , In general, the relaxation strength of secondary relaxations increases with increasing temperature because new polar mechanisms associated with the process become available. However, the dielectric strength of the glass–rubber relaxation decreases with increasing temperature owing to the fact that the thermal energy hinders the alignment with the electric field of the dipoles intervening in the segmental motions. This behavior is depicted in Figure 39.15 where the dielectric strengths of the α and β relaxations for poly(5-ethyl-1,3-dioxacyclohexane acrylate) (Huang et al., 2002) are shown.

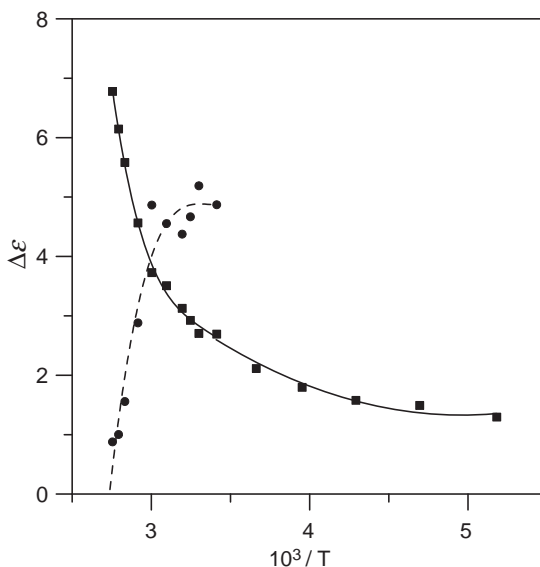


FIGURE 39.15 Dependence of the dielectric strength for β (filled squares) and α (filled circles) relaxations of poly(5-ethyl-1,3-dioxacyclohexyl acrylate) (Huang et al., 2002).

The dielectric strengths of relaxation processes can also be obtained from complex dielectric plots ϵ'' versus ϵ' , described by Equation (39.29). The intersection of the curves with abscissa axis at low and high frequencies gives, respectively, the relaxed (ϵ_0) and unrelaxed (ϵ_∞) dielectric permittivities. The strength of the relaxation i can be expressed in terms of the mean-square dipole moments by means of the Fröhlich equation (Fröhlich, 1948)

$$\epsilon_{0;i} - \epsilon_{\infty;i} = \frac{\epsilon_{0;i}(\epsilon_{\infty;i} + 2)^2}{(2\epsilon_{0;i} + \epsilon_{\infty;i})} \frac{4\pi\rho N_A g_i \mu_u^2}{9Mk_B T} \quad (39.36)$$

where μ_u^2 and M are, respectively, the square of the dipole moment and the molecular mass of the repeat unit, k_B and N_A are, respectively, the Boltzman constant and Avogadro's number, ρ is the density of the polymer, and T is the absolute temperature. The product $g_i \mu_u^2$ denotes the fraction of μ_u^2 relaxing through the relaxation process i . By obtaining the extreme values of the permittivity for the relaxation i from the experimental results using pertinent complex dielectric plots, the amount of μ_u^2 relaxed in each relaxation can be estimated.

The dielectric loss spectra are very sensitive to structure. Figure 39.16 shows the isochrones of the components of the complex dielectric permittivity of conventional (methylacrylate) (PMMA), at 10 Hz, in the temperature range -100 – 140°C . As temperature increases, the real component of the dielectric permittivity increases reaching a plateau; then ϵ' increases again, reaching a second plateau (Schönhals, 1997; Kremer, 2003; Ferry, 1980). The dielectric loss isochrone presents a prominent β relaxation below T_g ($\sim 105^\circ\text{C}$) which is believed to be associated with molecular motions about the $\text{CH}(\text{CH}_3)\text{—CO}(\text{O})$ bonds of the side groups. Above T_g , the isochrone shows a less developed absorption associated with the α or glass–rubber relaxation (Tetsutani et al., 1982). The fact that the β relaxation is more prominent than the α in conventional PMMA is caused by the methyl group of the repeat unit that hinders the cooperative conformational transitions in the

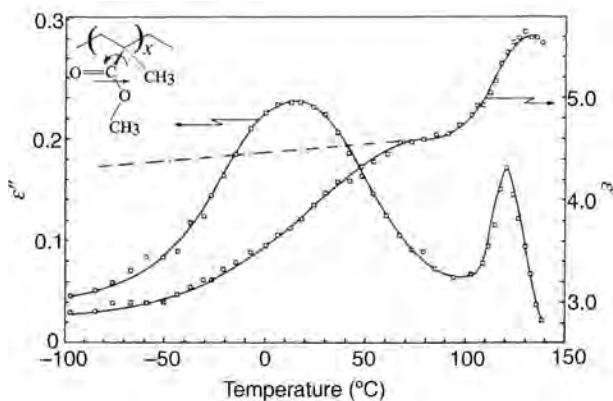


FIGURE 39.16 Real and imaginary isochrones for conventional poly(methyl methacrylate) at 10 Hz. Notice that the strength of the β relaxation (left peak) is significantly larger than that of the glass–rubber relaxation (right peak). The arrow in the inset indicates the dipole associated with the ester group (Tetsutani et al., 1982).

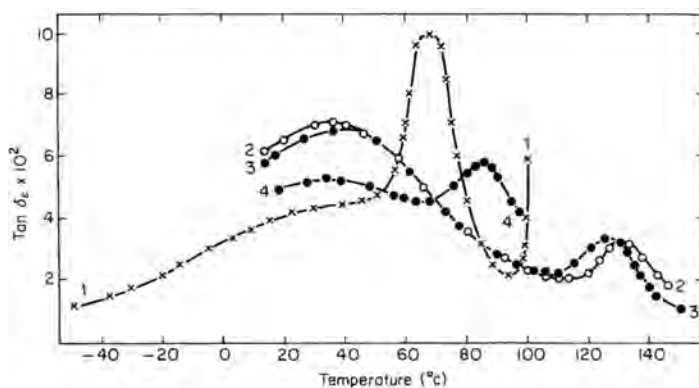


FIGURE 39.17 Dielectric loss factor as a function of temperature for conventional (2), isotactic (3), and syndio (1) polymethyl methacrylate. Curve 4 corresponds to a mixture (50/50) of iso and syndio poly(methyl methacrylate) (Mikhailov and Borisova, 1961).

segmental motions. As a result, the T_g rises from about 0°C for poly(methyl acrylate) (PMA) to ca. 105°C for PMMA. It is worth noting that the stereoregularity of the chains affects the response of PMMA to perturbation fields as Figure 39.17 shows (Mikhailov and Borisova, 1961). For example, the spectrum of iso-poly(methyl methacrylate) is similar to that of conventional PMMA. However, the α relaxation for syndio-PMMA is much more prominent than the β . As usual, the β relaxation of PMA and poly(n -alkyl methacrylates) with $n \geq 2$ is weaker than the α .

Aside from the chemical nature and the flexibility of the main chain, the chemical structure of the side groups of molecular chains has a rather high influence on the secondary dielectric relaxations of polymers. For example, the ratio between the strengths of the β and α relaxations $\Delta\epsilon_\beta/\Delta\epsilon_\alpha$, for poly(chlorophenyl acrylate) depends of the location of the chlorine atom in the phenyl group (Díaz-Calleja et al., 1991a, 1991b). The ratio decreases in the order P2CMA > P3CMA > P4CMA, where the numbers in the acronyms denote the location of the chlorine atom in the phenyl group of the side-chains. This behavior can be explained taking into account that dielectric activity in these polymers arises from motions about $\text{CH}(\text{CH}_3)\text{—CO}(\text{O})$ bonds coupled with motions about the $\text{O—C}_6\text{H}_5\text{Cl}$ bonds. For each $\text{CH}(\text{CH}_3)\text{—CO}(\text{O})$ conformation, significant changes in polarity come from rotations about $\text{O—C}_6\text{H}_5\text{Cl}$ bonds, that decrease in the order P2CMA > P3CMA > P4CMA. This fact suggests that motions about $\text{O—C}_6\text{H}_5\text{Cl}$ bonds play an important role in the development of the β relaxation of these poly(chlorophenyl acrylates).

The dielectric relaxation behavior of polymers is strongly influenced by the crystallinity that in turn depends on the undercooling $\Delta T = T_m - T_c$, where T_m and T_c are, respectively, the thermodynamic melting temperature and the crystallization temperature. The crystallization rate is small at low undercoolings, being zero for $\Delta T = 0$. On the other hand, the crystallization rate also tends to zero at high undercoolings where molecular transport governs the crystallization process (Mandelkern, 2004). Then by quenching from the melt semicrystalline polymers with low crystallization rate, such as poly(ethylene terephthalate) (PET) and aromatic polycarbonates, amorphous materials are obtained. Polymers of this type do not reach crystallization extents beyond 50%. There are also

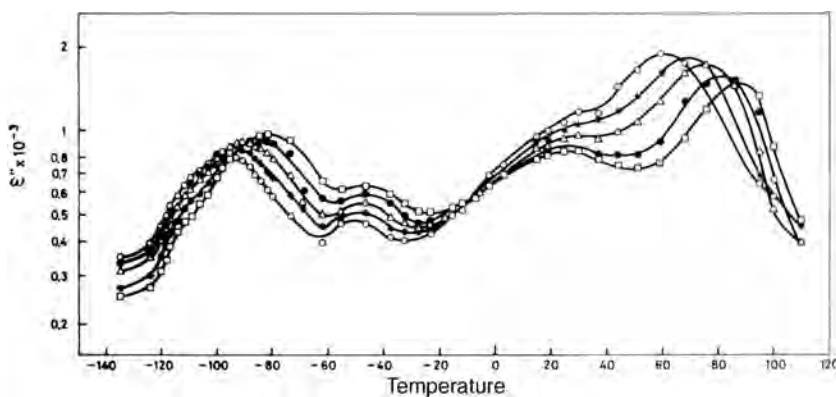


FIGURE 39.18 Isochrones representing the temperature dependence of the dielectric loss for polyethylene with 68% crystallinity. The frequencies of the isochrones are 5×10^2 , 10^3 , 2×10^3 , 5×10^3 , and 10^4 (Hz) (Ribes-Greus and Diaz-Calleja, 1989).

polymers that are difficult to crystallize to extents higher than 60%, but they are not quenchable to the amorphous state. Examples of these medium crystallinity polymers are aliphatic polyamides and polyesters. A third kind of semicrystalline polymers integrated by polyethylene, polyethers, and so on exhibit crystallinity after quenching hardly below 50%. The crystallinity of these polymers lies in the range 60–80%.

Figure 39.18 shows the isochrones for the dielectric loss of polyethylene with 68% crystalline content (Ribes-Greus and Diaz-Calleja, 1989). The relaxation appearing at higher temperature, named α process, is connected with motions in the crystalline entities of the material. In decreasing order of temperature appears the so-called β absorption, developed in the amorphous fraction of the material, which is associated with the glass–rubber relaxation. The intensity of this relaxation decreases as crystallinity increases. The lower temperature process, called γ absorption, emerges from molecular local motions in the amorphous phase though the assumption that it may have an important component from the crystalline phase should not be ruled out. Notice that in semicrystalline polymers the glass–rubber relaxation is denoted β relaxation instead of α , the nomenclature used for this process in amorphous polymers.

The effect of the crystallinity on the dielectric relaxation behavior is made evident in PET, and other semicrystalline polymers that can be obtained in the amorphous state. Illustrative plots showing the components of the complex dielectric permittivity in the frequency domain for both amorphous PET and 50% crystalline PET are presented in Figure 39.19 (Boyd, 1985, 1997). An inspection of the curves shows that the relaxation strength of the dielectric glass–rubber or β relaxation is smaller than that of the same process or α relaxation in the amorphous polymer. Moreover, the width of the β peak increases with the presence of crystalline entities, and therefore, the relaxation processes are less defined in semicrystalline polymers than in amorphous ones. The presence of crystallites entities in PET shifts to lower frequencies the location of the dielectric glass–rubber β process. The strength of the secondary γ absorption, not shown here, decreases as the crystallinity increases, but its location in the frequency domain is independent on the degree of crystallization. The comparatively low dielectric strengths of the β

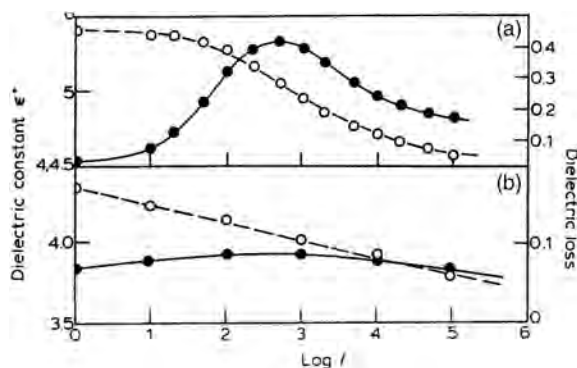


FIGURE 39.19 Components of the complex dielectric permittivity in the frequency domain for amorphous PET (a) and PET with crystallinity 50% (b). Open and filled symbols represent, respectively, the real and loss components of ϵ^* (Boyd, 1985, 1997).

relaxations of semicrystalline polymers arise from the fact that crystalline entities hinder segmental and local motions in the amorphous regions.

Besides the amorphous and crystalline state, condensed matter presents an additional state of aggregation, called liquid crystalline order (Collins, 1990; de Gennes and Prost, 1995). In this state molecular rods are free to move as a liquid, though as doing so they spend a little more time pointing toward a definite direction than along some other direction. An arrow, called the director, represents the direction along which the rod molecules or segments tend to align. Poly(*n*-alkyl isocyanates) having a planar or helicoidal rigid backbone, to which relatively flexible side groups are attached, develop liquid crystalline order in solution at room temperature and in bulk at elevated temperatures (Aharoni and Walsh, 1979; Aharoni, 1979). Thorough studies (Moscicki et al., 1982) on the dependence of the dielectric behavior of poly(*n*-alkyl isocyanates) on toluene concentration show that for concentrations lying in the range $0 < c < 15\%$, the relaxed dielectric permittivity ϵ_0 increases with concentration, ϵ_0 being a linear function of concentration. In this region, the system is isotropic because all the molecules through the Brownian motion randomize the dipole vectors associated with the highly polar rods into a 4π solid angle. In the concentration range $15\% < c < 25\%$, ϵ_0 increase with concentration, but this quantity falls below the linear relation because the orientation of the rods is hindered by the viscosity of the medium. In the range $22\% < c < 35\%$, a critical concentration is reached at which liquid crystalline phase coexisting with an isotropic phase appears. The component of the biphasic displaying liquid crystalline order is much larger than that of the isotropic component. As a result, a severe drop in ϵ_0 occurs because the permittivity of the biphasic may be considered as the weighted sum of the relaxed dielectric permittivities corresponding to the amorphous region, which is high, and the ordered region, which is low.

Most dielectric studies on liquid crystalline polymers were carried out on polymers containing mesogenic groups in the side chains with capability to develop mesomorphic order. In these polymers, called side chain liquid crystalline polymers (SCLCP), the mesogenic groups can be aligned by magnetic or electric fields. Obtaining SCLCP thermotropic polymers requires the placement of a flexible spacer between the main chain and the mesogenic group. These polymers develop smectic mesophases. The isochrones of

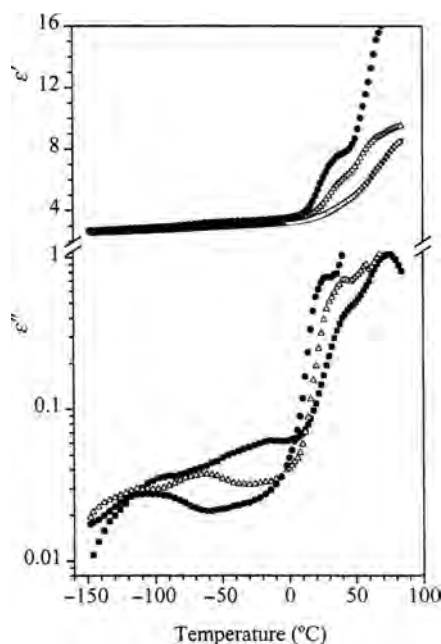


FIGURE 39.20 Illustrative isochrones showing the temperature dependence of the components of the complex dielectric permittivity for a thermotropic SCLC polymer at several frequencies: (filled circles) 1 Hz, (open triangles) 10^2 Hz, and (filled squares) 10^4 Hz (Díaz-Calleja et al., 1999).

SCLCP polymers present subglass absorptions followed in increasing order of temperature by two overlapped relaxations formed by the glass–rubber relaxation or α process and the δ peak (Figure 39.20) (Díaz-Calleja et al., 1999; Attard et al., 1988). The latter peak is assigned to librational fluctuations of the mesogen around the short molecular axis. Different extents of alignment between fully homeotropic and fully planar (homogeneous) structures can be achieved by slowly cooling the isotropic material back into the liquid crystalline state in the presence of AC fields of given amplitude and frequency. Whereas the loss dielectric curve is bimodal for the isotropic material, the loss curve seems to be made up of a δ process with only a small intensity α relaxation for the homeotropic aligned material.

To develop ferroelectric behavior, the mesogenic group in SCLCPs must contain a dipole perpendicular to the main axis of that group. Moreover, a chiral moiety must be attached to the end of the mesogenic group or to the spacer. To facilitate mesogenic alignment, the polymer backbone should be a flexible polymer, such as polyacrylates or polysiloxanes. A representative polymer of a ferroelectric SCLCP polymer is shown in Figure 39.21. The relaxation spectra of ferroelectric SCLCPs in the frequency domain present a strong absorption below 1 MHz, called Goldstone mode, typical of S_C^* mesophases, arising from fluctuations of the phase of helical structures. By imposing a DC field, the Goldstone mode decreases, eventually disappearing. With the suppression of the Goldstone mode, the soft mode emerges at higher frequencies associated with unwound state (Figure 39.22) (Schönfeld et al., 1994; Kremer, 1997). It is worth noting that dielectric spectroscopy is a useful tool to study mesophase transitions as a function of temperature in ferroelectric polymers.

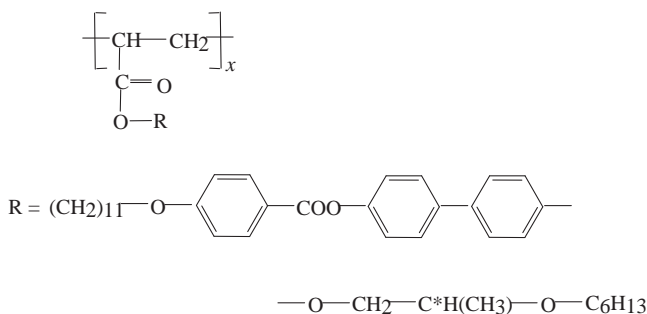


FIGURE 39.21 Scheme of a SCLC polymer with the requirements necessary to display ferroelectric behavior. Notice that the asterisk in the carbon of the end segments denotes chirality.

39.5 THERMOSTIMULATED DEPOLARIZATION CURRENTS

Electrical charges are virtually immobile at low temperatures in glassy systems. To stimulate their motion and eventually produce dipole libration, which is the origin of the dielectric relaxation, polymers must be heated. If an electrical field is applied during cooling, and then removed at the lowest temperature, the motions of the dipoles in the system can be thermally stimulated. This procedure gives rise to a thermogram in which depolarization peaks of current appear. Illustrative thermograms are shown in Figure 39.23 (Domínguez-Espinosa et al., 2006b). The thermograms exhibit secondary peaks followed by a prominent peak corresponding to the glass–rubber relaxation. The molecular origin of the peaks in the thermograms, especially at temperatures below the glass transition, is the same as that observed for loss peaks in broadband dielectric spectroscopy. The conversion of TSDC into ϵ'' data has been described elsewhere (Hino, 1975).

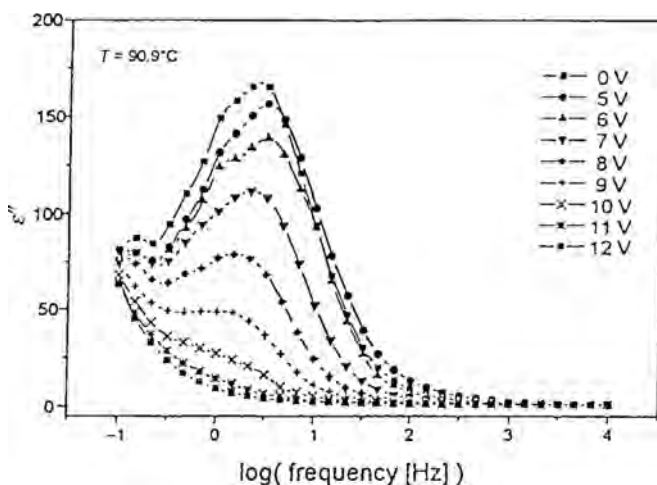


FIGURE 39.22 Dielectric loss in the frequency domain for a thermotropic SCLC thermotropic polymer under orientation promoted by the electric voltages indicated in the inset (Schönfeld et al., 1994; Kremer, 1997).

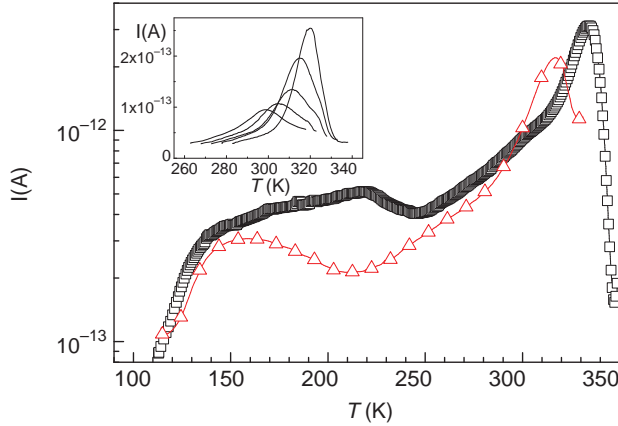


FIGURE 39.23 Global thermo stimulated depolarization current (TSDC) curves for poly(2,4-difluorobenzyl methacrylate) (open squares) and poly(3-fluorobenzyl methacrylate) (open triangles). The inset shows partial TSDC curves for P24FM obtained at poling temperatures from 273 to 298 K (5 K step) (Domínguez-Espinosa et al., 2006b).

Another convenient way to obtain the dielectric permittivity from TSDC results, called thermal windowing, is described elsewhere (Perlman and Unger, 1974; Shimizu and Nakayama, 1991). The method essentially consists in performing partial depolarizations in the material in the glassy state, in small temperature windows covering the experimental range of temperatures where the relaxation peaks appear. Illustrative partial thermal stimulated depolarization curves are shown in the inset of Figure 39.23. The real and imaginary parts of the complex permittivity can be obtained from the partial depolarization curves by

$$\begin{aligned}\varepsilon'(\omega) &= \varepsilon_{\infty} + \sum_{i=1}^n \frac{\Delta\varepsilon_i}{1 + \omega^2\tau_i^2} \\ \varepsilon''(\omega) &= \sum_{i=1}^n \frac{\Delta\varepsilon_i(\omega\tau_i)}{1 + \omega^2\tau_i^2}\end{aligned}\quad (39.37)$$

where τ_i are given by the following equation (van Turnhout, 1972)

$$\tau_i = \frac{\int_0^t i(t') dt'}{i(t)} = \frac{\int_{T_i}^T i(T') dT'}{hi(T)} \quad (39.38)$$

where T is the absolute temperature, T_i is the initial temperature of each individual depolarization curve, and h is the inverse of the heating rate or $1/(dT/dt)$. The dielectric strength $\Delta\varepsilon_i$ of partial depolarization curves used in Equations (39.37) are obtained as

$$\Delta\varepsilon_i(T_p) = \frac{1}{e_0 h E} \int_{T_0}^{T_f} i(T') dT' \quad (39.39)$$

From partial depolarization curves, the frequency dependence of the components of the complex dielectric permittivity can be estimated for glassy polymers in the region of very low frequencies.

39.6 CONDUCTIVITY IN POLYELECTROLYTES AND POLYMER-ELECTROLYTES AS SEPARATORS FOR LOW TEMPERATURE FUEL CELLS AND ELECTRICAL BATTERIES

Polymers with anionic or cationic groups covalently anchored to the structure of molecular chains are called, respectively, cation- and anion-exchange polyelectrolytes. For strong cation-exchange resins, the fixed ionic groups can be $-\text{SO}_3\text{H}$ or $-\text{PO}_3\text{H}_2$ and $-\text{N}(\text{CH}_3)_3\text{OH}$ for anionic-exchange resins. The fixed groups dissociate in water as $-\text{SO}_3\text{H} \rightarrow -\text{SO}_3^- + \text{H}^+$, $-\text{PO}_3\text{H}_2 \rightarrow -\text{PO}_3\text{H}^- + \text{H}^+$, $-\text{N}(\text{CH}_3)_3\text{OH} \rightarrow -\text{N}(\text{CH}_3)_3^+ + \text{OH}^-$ (Sata, 2004).

Usually, ion-exchange membranes are cast from polyelectrolyte solutions. The characteristics of ion-exchange membranes are: (a) ionic conductivity, (b) hydrophilicity, and (c) the existence of a fixed carrier (ion-exchange group). Mobile ions with opposite sign to that of the ionic fixed groups (anions and cations, respectively, for cation and anion-exchange membranes) are called counterions, whereas those with the same sign are coions. Traditionally, ion-exchange membranes have been used in electrodialysis, demineralization being the most promising application of this technique. Two cases can be distinguished in the use of electrodialysis: (a) demineralization of solutions containing only electrodialysable electrolyte solutes and (b) demineralization of solutions containing an electrolyte and a nondialysable electrolyte, such as a protein. The experimental devices for electrodialysis are made up of electrolytic cell groups, each of them separated from the two contiguous ones by a cation-exchange membrane at one side and an anion-exchange membrane, at the other side (Sata, 2004). Anion- and cation-exchange membranes to be used in electrodialysis must have high conductivity and high permselectivity. In an ideal permselective membrane, each Faraday of current carries one equivalent of counterions across the membrane.

Owing to the ever increasing demand of new kinds of energy sources for portable electronic products, such as cellular phones, and the stringent environmental regulations requiring zero emission traction vehicles, academic and industrial research laboratories are focused their interests on the development of electrochemical power sources that can provide high energy density and long cyclability. Moreover, the power sources must be reliable and safe. Two new sources of energy that use polyelectrolytes and electrolytes solvated in polymers are, respectively, low temperature fuel cells and rechargeable lithium batteries.

Low temperature fuel cells are energy source devices in which fuel (hydrogen, methanol, etc.) is reduced in the anode of the cell producing protons that travel across a separator, usually an acidic ion-exchange membrane, to the cathode where they react with reduced oxygen forming water as subproduct. The configuration of a fuel cell is anode (catalyst, i.e., platinum, dispersed in an electronic conducting substrate), | fuel (hydrogen, methanol), | cation-exchange membrane in the acid form | oxygen or air | cathode (catalyst, i.e., platinum dispersed in an electronic conducting substrate). Electrons travel from the anode to the cathode by an external circuit, whereas protons do that across a cation-exchange membrane in the acid form. Using hydrogen as fuel, the overall chemical reaction is $2\text{H}_2 + (1/2)\text{O}_2 \rightarrow \text{H}_2\text{O}$. Taking into account that $\Delta g^0 = -2FE_0$, where Δg^0 , F and E_0 are, respectively, the change in standard free energy of the redox reaction, Faraday's constant and standard electromotive force of the cell, the value of E_0 is 1.24 V. The voltage of the fuel cell can be expressed in terms of the current density by (O'Hayre et al., 2006; Larminie and Dicks, 2003).

$$V = E_0 - A \ln \frac{i}{i_0} - ir + m \ln(n \cdot i) \quad (39.40)$$

The second term on the right-hand side of Equation (39.40) denotes the reduction in voltage due to the redox reactions at the electrodes. The third term ir ($r = RS$, where R and S are, respectively, the resistance and area of the acidic membrane acting as separator) represents the loss in voltage caused by the transport of protons across the membrane. The last term in Equation (39.40) is a correction in the voltage arising from the slight reduction of reactants at the electrodes. An illustrative curve showing the variation of the voltage of a fuel cell with the current density is shown in Figure 39.24.

The solid polyelectrolyte separating the anode from the cathode in fuel cells is a cation exchange membrane in the acid form. Acidic membranes should exhibit high proton conductivity and good chemical stability at the operating fuel cell conditions, about 80°C in available low temperature fuel cells. However, to reduce platinum poisoning by effect of traces of CO in the fuel, it would be necessary to develop membranes with good chemical and mechanical properties in the vicinity of 150°C. This is one of the major problems facing the commercial development of fuel cells technology as a source of energy. There are many reviews that address membranes performance in relation with their chemical stability and proton conductivity (Hickner et al., 2004; Kreuer, 1996; Rozière and Jones, 2003; Schuster and Meyer, 2003; Rikukawa and Sanui, 2000). Up to now, the most successful membranes are based on perfluorosulfonate electrolytes, Nafion being the principal one. Nafion, the chemical structure of which is shown in Figure 39.25, exhibits rather high proton conductivity lying in the range 0.01–0.1 S/cm at 25°C. However, Nafion membranes present some drawbacks such as high permeability to methanol, high cost, environmental problems and a decrease in both chemical stability and mechanical properties at temperatures above 100°C. Proton exchange membranes are being developed based on high chemical stability polymers such as polysulfones, poly(ether-ketone), and polyimides (Hickner et al., 2004; Kreuer, 1996; Rozière and Jones, 2003; Schuster and Meyer, 2003; Rikukawa and Sanui, 2000). Illustrative sulfonated

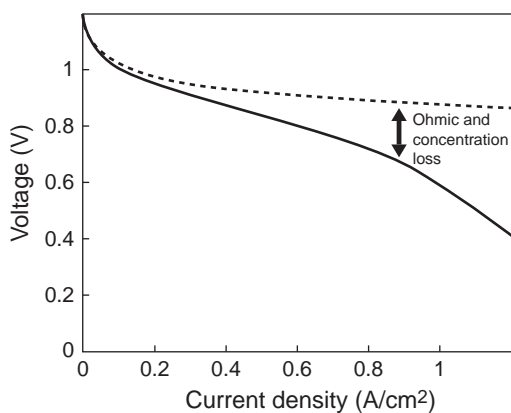
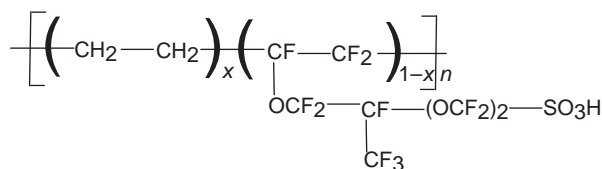
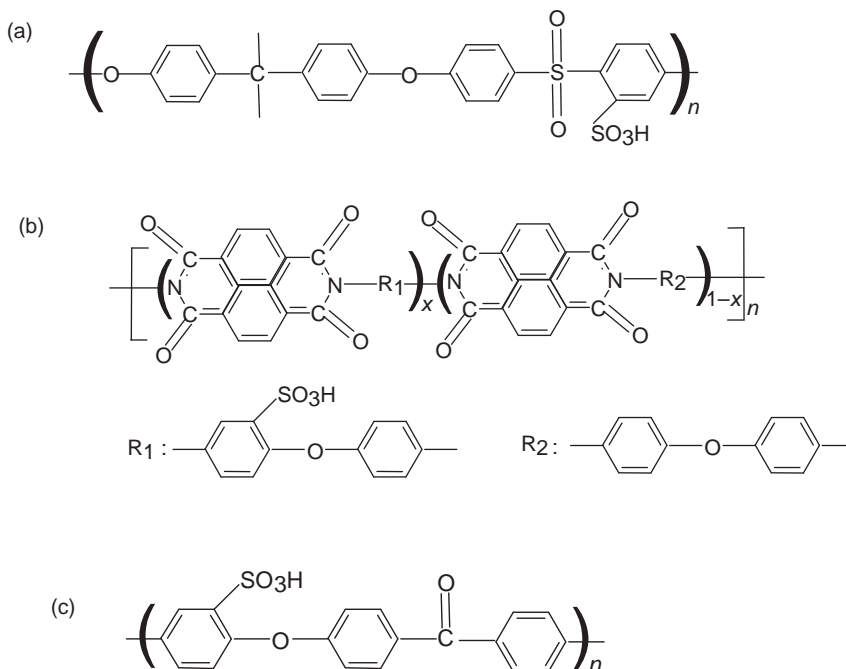


FIGURE 39.24 Illustrative curve (continuous line) showing the dependence of the voltage on the current density. The activation loss (discontinuous line) represents the loss of voltage caused by the reactions at the electrodes.

**FIGURE 39.25** Chemical structure of Nafion.

structures are shown in Figure 39.26. Conductivities lying in the range 10^{-3} – 10^{-2} S/cm, at room temperature, have been reported for membranes of this type.

Proton conduction in acidic membranes is very sensitive to the ion exchange capacity and water uptake. Water promotes segregation of the fixed acid groups forming hydrophilic nano-size domains and hydrophobic domains that contains the amorphous part of the chains. In this way, hydrophilic percolation paths are formed through which protons transport takes place. The conductivity of membranes is strongly dependent on the water content. Dry acidic membranes have conductivity lower than 10^{-8} S/cm as a consequence of the fact that dissociation reaction $-\text{SO}_3\text{H} \leftrightarrow -\text{SO}_3^- + \text{H}^+$ does not take place in the fixed $-\text{SO}_3\text{H}$ groups unless the membrane contains about 2 moles or more of water/ $-\text{SO}_3\text{H}$ group (Paddison, 2003). A dramatic increase of several decades occur in proton conductivity as the water content in the membrane increases. Proton transport in acidic membranes is a rather complex process that is discussed in detail elsewhere (Kreuer et al., 2004).

**FIGURE 39.26** Structures of representative cation-exchange membranes prepared from high chemical stability polymers: (a) polyarylene ether sulfone; (b) naphthalenic copolyimide; (c) polyetherether ketone.

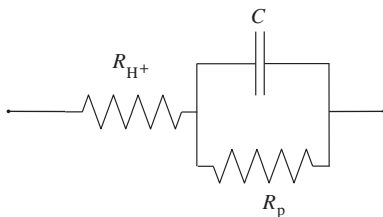


FIGURE 39.27 Equivalent electric circuit for a cation-exchange membrane in the acid form.

The conductivity of acidic membranes is usually measured by impedance spectroscopy. To carry out the measurements, the membrane is sandwiched between two blocking electrodes. The equivalent electrical circuit of the acidic membrane consists of a resistance accounting for the proton resistance of the acidic membrane in series with circuit made up of a resistance, R_p , in parallel with a capacitor that describe polymer-electrode polarizations (see Figure 39.27). The impedance of the circuit is given by

$$Z^* = R_{H^+} + \frac{R_p}{1 + j\omega R_p C} = R_{H^+} + \frac{R_p}{1 + (\omega R_p C)^2} - j \frac{\omega R_p^2 C}{1 + (\omega R_p C)^2} \quad (39.41)$$

The plot Z' versus Z'' , called Nyquist diagram, is an arc (Figure 39.28) displaced from the origin of the coordinate axis by R_{H^+} . Notice that $\lim_{\omega \rightarrow \infty} Z'(\omega) = R_{H^+}$ and $\lim_{\omega \rightarrow 0} Z'(\omega) = R_{H^+} + R_p$, whereas in both limit situations $Z'' = 0$. This analysis assumes that the polarization process has a single relaxation time given by $\tau = R_p C$. Real Nyquist diagrams are not described by Equation (39.41) as Figure 39.28 shows owing to the fact that polarization processes are associated with a wide distribution of relaxation times (Compañ et al., 2007). Fitting the data would require an equivalent circuit formed by N circuits in series, each of them formed by a resistance R_{pi} in parallel with a constant phase element of admittance $Y_i^* = Y_{0i}(j\omega\tau_i)^{n_i}$ where $0 < n_i \leq 1$. The impedance of the circuit is given by

$$Z^*(\omega) = R_{H^+} + \sum_{i=1}^N \frac{R_{pi}}{1 + R_{pi}Y_{0i}(j\omega\tau_i)^{n_i}} \quad (39.42)$$

In the extreme frequency limits $Z'' = 0$, whereas $Z' = R_{H^+}$ and $Z' = R_{H^+} + \sum_{i=1}^N R_{pi}$ as $\omega \rightarrow \infty$ and $\omega \rightarrow 0$, respectively. At high frequencies, the contribution of the polarization to the impedance is negligible and Equation (39.41) can be used to estimate the proton resistance of cation-exchange membranes. The proton conductivity of membranes is given by

$$\sigma = \frac{l}{R_{H^+} S} \quad (39.43)$$

where l and S are, respectively, the thickness and area of the membrane whose resistance is R_{H^+} . The dependence of the proton conductivity on temperature in acidic membranes follows Arrhenius behavior with 2–3 kcal/mol of activation energy.

Proton transport across acidic membranes is accompanied by water molecules, and the water thus transported is called electro-osmotic water. The electro-osmotic process dries

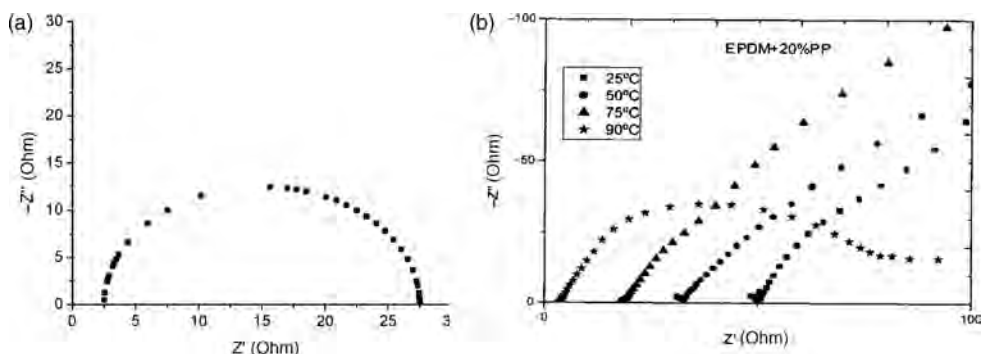
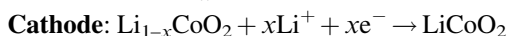
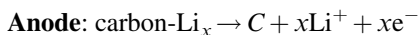


FIGURE 39.28 Nyquist diagram for a Debye type polyelectrolyte (a) and a real acidic membrane formed by blends of sulfonated EPDM and poly(propylene) (b). (Notice that the frequency increases from right to left) (Compañ et al., 2007).

the membrane in the anode side and as a result decreases the proton conductivity. Then besides high conductivity, membranes should exhibit low electro-osmosis (Lakshminarayanan, 1966; Riande, 1972).

Another recent breakthrough in power electric sources development in which polymers may play an important role has been the development of rechargeable commercial lithium batteries. These devices can exhibit high level of performance driven by their high volumetric and gravimetric energy density. Moreover, the environmental friendly condition of these power source devices makes them an attractive alternative to traditional battery technologies involving toxic metals (Pb/PbO₂, Ni/Cd). In principle, the anode of the battery is made up of Li or carbon–Li intercalation compound separated by means of an organic conducting electrolyte from the cathode, usually an inorganic compound with an open structure, that is, layered structure (i.e., LiCoO₂, V₂O₅) or a tunnel structure (i.e., V₆O₁₃, MnO₂) (Scrosati, 2001). Using carbon–Li intercalations in the anode, the typical electrochemical reactions are



The state of the art in Lithium batteries can be found in literature publications and reviews (Scrosati, 2001; MacCallum, 1987; Meyer, 1998; Koksang et al., 1994; Stephan and Nahm, 2006; Wright, 2002; Appetechi and Croce, 1998). The reasonable low conductivity, of the order of 10⁻² S/cm at room temperature, exhibited by rechargeable lithium batteries based on lithium metal electrodes and liquid organic electrolytes, is tarnished by safety problems these cells present. Safety problems could be circumvented using “plastics” electrolytes such as polyethylene oxide (PEO) complexed with a lithium salt. PEO crystallizes in helicoidal form and its melting point is about 70°C. At room temperature, the conductivity of PEO–LiX electrolytes is only about 10⁻⁷ S/cm, because only the amorphous phase of the semicrystalline polymer intervenes in the transport process. Complexes of type (PEO)_xLiClO₄ and (PEO)_xLiCF₃SO₃, with *x* in the range 8–50 salt molecules, exhibit an acceptable level of ionic conductivity, of the order of 10⁻³ S/cm at temperatures slightly above the melting temperature of PEO (Cairns and McLaren, 1993; Armand, 1987). Transport of Li⁺ in PEO involves ion solvation in the melt polymer and

further motion across the chains as Figure 39.29 illustrates (Meyer, 1998; Armand, 1987). In this context, molecular dynamics simulations suggest that Li^+ ions are solvated by PEO through five ether groups (Müller-Plathe and van Gunsteren, 1995), and the mobility of the ions in the chains is related to the motions of the complexing segments of the PEO chains. Notice that the electrolyte is not covalently bound to the polymer chains as occurs in the solid polyelectrolytes used as separators in low temperature fuel cells. The temperature dependence of the conductivity in the melt ($T > 70^\circ\text{C}$) obeys to the phenomenological Vogel–Fulcher–Tammann–Hesse (VFTH) equation (Armand, 1987) $\sigma(T) = AT^{-1/2} \exp[-m/(T - T_V)]$.

Power density is rather high for lithium polymer batteries as result of the thin laminate construction used. Owing to the fact that poly(propylene oxide) (PPO) is amorphous, and its T_g is below room temperature, one could think that this polymer could be used as complexing agent for Li^+ salts. However, the polymer exhibits limited solvation properties as a result of the fact that the rigidity of the chains promoted by the methyl solvation groups predominates over the increased donor power of the oxygen atoms linked to a secondary carbon. The conductivity of PPO–Li complexes lies in the range 10^{-5} – 10^{-6} S/cm (Watanabe and Ogata, 1987).

The preparation of separators with suitable conductivity at room temperature and below room temperature can be accomplished by adding organic liquids or plasticizers (i.e., propylene carbonate, ethylene carbonate, ethylene glycols of low molecular weight) to PEO. However, the liquid additives produce loss of mechanical stability, and moreover, they can react with lithium metal. Dispersion of nanoscale particles of selected ceramic powders, such as TiO_2 , SiO_2 , and Al_2O_3 , in the PEO–LiX matrix increases the amorphicity of the system without affecting the mechanical and interfacial properties of the electrolyte (Croce et al., 1998, 1999, 2000). The disperse ceramics prevents the crystallization of PEO and enhances the conductivity of the electrolyte composite that can reach values of 10^{-5} S/cm at room temperature.

The preparation of the electrolyte for lithium batteries involves the trapping in a polymer matrix of lithium salt solutions in organic solvent mixtures. Polymers commonly studied are polyacrylonitrile (PAN), poly(methylmethacrylate) (PMMA), or

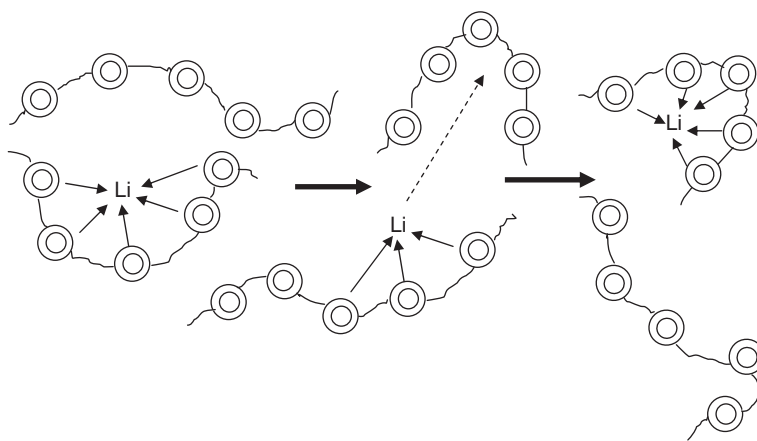


FIGURE 39.29 Scheme of the solvation and transport of Li^+ cations in amorphous poly(ethylene oxide).

poly(vinylidene fluoride) (PVDF) (Scrosati, 1997, 1998). Liquid immobilization of the Li^+ salts may involve UV cross-linking, casting, or gelation, the latter being the most widely used method. In this case, the gel membranes are not true polymer electrolytes, but hybrid systems containing a liquid phase within the polymer matrix. Ionic conductivities in a range 10^{-2} – 10^{-3} S/cm at 25°C have been reported for the gel formed by immobilizing an ethylene carbonate-dimethyl carbonate lithium hexafluorophosphate solution in polyacrylonitrile.

Lithium batteries can also be designed in which the cathode is an electronic conducting polymer. The cell configuration could be $\text{Li}|\text{electrolyte separator}(\text{A}^-\text{Li}^+)|\text{P}^+\text{A}^-$ where P^+ indicates p-type doping. Other configurations have been described but none of them is commercially available (Müller et al., 1997).

The conductivity of polymer electrolytes can be measured by impedance spectroscopy by sandwiching the electrolyte between two parallel blocking electrodes, such as platinum electrodes. Under the action of an alternating electric field, the lithium ions migrate back and forth in phase with the voltage. The migration of the lithium ions is represented by the resistance R_{Li} , whereas the polarization of the immobile polymer chains may be represented by a capacitor C in parallel with the resistor. Each electrode–electrolyte interface is similar to a parallel-plate electrode of capacitance C_e in such a way that the total capacitance of the two in series condensers is $1/C_e = 1/C_1 + 1/C_2$. The admittance of the parallel circuit is

$$Y^* = \frac{1}{R_{\text{Li}}} + j\omega C \quad (39.44)$$

where ω is the angular frequency of the alternating voltage. Since the C_e , capacitor is in series with the $R_{\text{Li}}-C$ parallel circuit, the complex impedance for the equivalent circuit of the electrolyte–electrodes system is

$$Z^* = \frac{1}{Y^*} + \frac{1}{j\omega C_e} = R_{\text{Li}^+} \left[\frac{1}{1 + (\omega R_{\text{Li}^+} C)^2} \right] - j \left[\frac{\omega R_{\text{Li}^+}^2 C}{1 + (\omega R_{\text{Li}^+} C)^2} + \frac{1}{\omega C_e} \right] \quad (39.45)$$

At high frequencies, $1/\omega C_e \cong 0$ and the complex Z'' versus Z' plot is an arc intersecting the abscissa axis at $Z' = 0$ ($\omega \rightarrow \infty$) and $Z' = R_{\text{Li}}$ (in the low frequency region). Then the conductivity, σ , of the polyelectrolyte separator can be obtained from

$$\sigma = \frac{l}{R_{\text{Li}^+} S} \quad (39.46)$$

where l is the thickness of the electrolyte and S is the area of the electrodes.

39.7 SEMICONDUCTORS AND ELECTRONIC CONDUCTING POLYMERS

Traditionally, semiconductors are of inorganic nature. However, the high flexibility, high impact resistance, and low density of polymers make these materials an attractive option to prepare semiconductors. Owing to the fact that each carbon is sp^2 hybridized in polydienes, polymers such as polyacetylene can be considered one-dimensional analogues of

graphite, an electronic conducting material. However, whereas the C—C bond lengths are equivalent in graphite, the backbone bond lengths of polydienes, such as trans-polyacetylene, are alternatively slightly longer and slightly shorter. This effect, known as Peierls distortion, opens a gap between the HOMO level of the fully occupied π bond (valence band) and the LUMO level corresponding to the empty π^* bond (conduction band) (Kohlman et al., 1996). These facts are illustrated in Figure 39.30. Trans-polyacetylene can be considered a semiconductor with an energy band of 1.5 eV.

Semiconductor character is not only exhibited by polymers of polyene type but also exhibited by conjugated aromatic polymers. For illustrative purposes, some of these polymers are shown in Figure 39.31 (Pron and Rannou, 2002; Frommer and Chance, 1996; Moratti, 1998). Under electric perturbations, these materials emit light, phenomenon known as electroluminescence. An electroluminescent diode in its simplest form consists of a single layer of polymer semiconductor sandwiched between two electrodes. The anode, which is a hole injector usually, is a thin layer of indium-tin oxide (ITO) transparent to the light, whereas the cathode, an electron injector, is a metal with a low work function such as calcium and magnesium (Yang and Pei, 1995; Kido et al., 1994). After application of an electric field, holes and electrons are injected, respectively, in the π and π^* orbitals. Injected charge carriers of different sign drift in opposite direction in the conjugated polymer matrix to form excited species, namely, singlet or triplet polaron-excitons. The radioactive decay of singlet excitons produces light whose frequency depends on the π – π^* energy gap. In increasing the energy gap, for example, by functionalization of the semiconductor polymer, the emitting light shifts to short-wave light (see Figure 39.32). The name of electroluminescent-diode given to these devices arises from

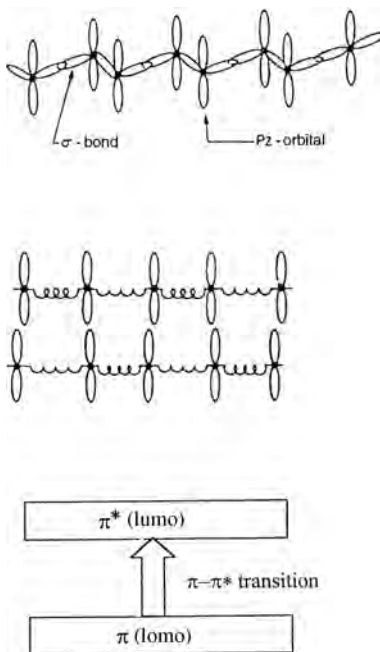


FIGURE 39.30 Schemes showing the sp^2 hybridization character of polydienes and the energy gap between the valence and conduction bands.

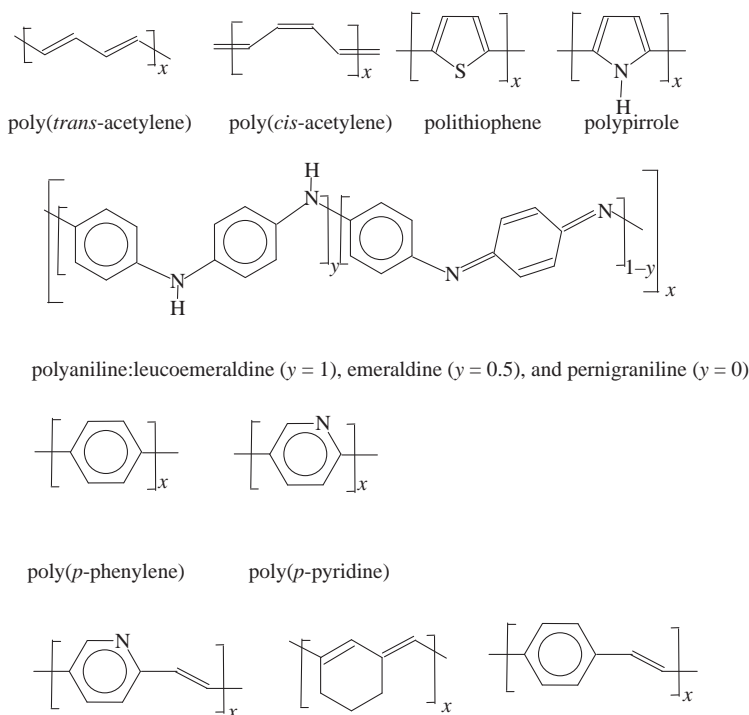


FIGURE 39.31 Representative conjugated polymers that exhibit semiconductor character.

the fact that no electroluminescence occurs until a voltage is reached above which electroluminescence rapidly increases with increasing voltage (Friend and Greenham, 1998; Leising et al., 1998). Conjugated polymers can also be used to construct photovoltaic cells (Yu et al., 1995; Hide et al., 1997; Tressler, 1999; McGehee and Heeger, 2000), polymer-based photopumped lasers, and polymer-based field-effect transistors (FET) (Würthner, 2001; Kraft, 2001; Sirringhaus et al., 1990). It is worth noting that functionalization of semiconductor polymers with moieties that increase the solubility is often performed to facilitate their processability.

In the early 1970s, Shirikawa and Ikeda (1971, 1974) reported the possibility of preparing strong, self-supporting films of polyacetylene by direct polymerization of

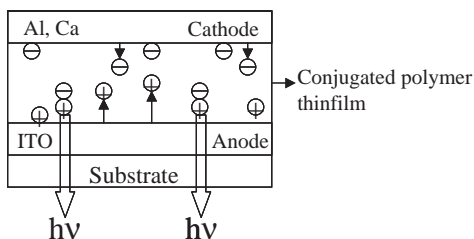


FIGURE 39.32 Scheme of a light-emitting diode.

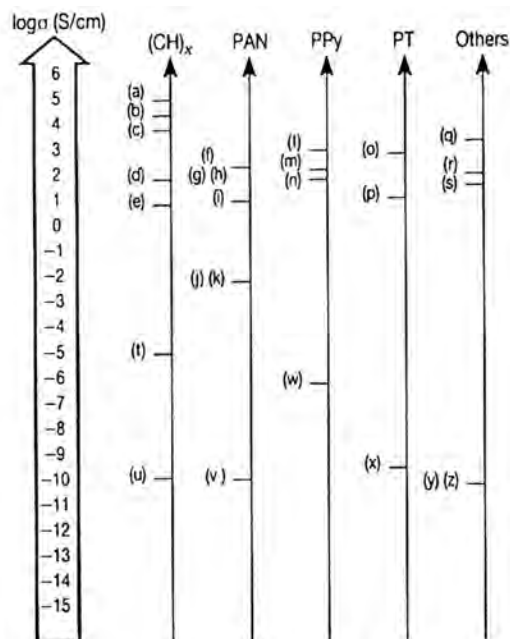


FIGURE 39.33 Electronic conductivities reported for doped poly(trans acetylene), polyaniline (protonated emeraldine, PAN), poly(pyrrol) (PPy), poly(thiophene) (PT), and others (Epstein, 1997).

acetylene. This finding attracted little attention until 1977 when Shirikawa et al. (1977) rediscovered that the treatment of polyacetylene with a Lewis acid or bases can increase its conductivity by up to 13 orders of magnitude. This process involves removal or addition of electrons. Then by oxidation or reduction of conjugated polymers using, respectively, an electron acceptor or an electron donor electronic conductor polymers are obtained. This process is called “doping.” Whereas the amount of dopant in inorganic semiconductors is of the order of parts per million, this amount can be up to 50% of the total weight in organic semiconductors. Trans-polyacetylene doping can change the conductivity of the polymer from 10^{-12} S/cm (undoped form) up to 10^5 S/cm (highly doped polymer), more than 14 orders of magnitude. Conductivities that present some conjugated polymers in different doping conditions are shown in Figure 39.33 (Epstein, 1997). Molecular order favors electronic conductivity of conjugated polymers (Joo et al., 1994). Atomic or molecular dopant ions are located interstitially between chains forming new three-dimensional structures. For polymers with degenerate ground states such as trans-polyacetylene, the charge transport added to the backbone can be described by the motion of a solitary wave, called a soliton in field theory notation (Mizes and Conwell, 1994; Yan et al., 1994). For nondegenerate systems (i.e., polypyrrole, polythiophene, poly(*p*-phenylene), etc.) the charges introduced at low doping process are stored as charged polarons (a radical cation or a radical anion plus a lattice distortion around the charges) or bipolarons (see Figure 39.34) (Bredas et al., 1982, 1984). The one-dimensionality of polymers leads to the localization of the electron wave functions. Then though experimental evidence suggests that the metallic states are three-dimensional, the transport properties are highly anisotropic. Macroscopic transport in conducting materials containing well-ordered

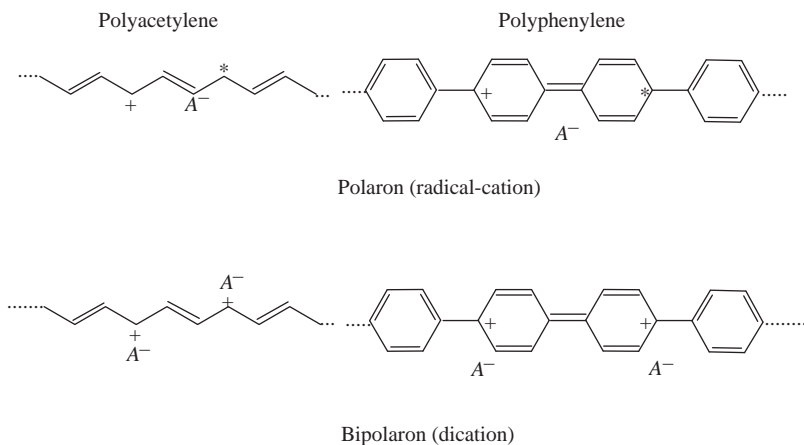


FIGURE 39.34 Electronic transport species in doped polyacetylene and doped poly(*p*-phenylene).

polymer chains is not possible unless the carriers are able to hop to an adjacent chain, otherwise resonant back staggering between chain ends confine electrons in one dimension.

Conducting polymers can be used for charge dissipation in textiles, resistive heaters provided that rather low power per m^2 is required, electromagnetic interference shielding, electrochromic applications, and so on. Other applications involve deposition of conducting polymers, for example neutral or doped polyaniline (protonated emeraldine), on metal surfaces to protect them against corrosion. In principle, conducting polymers can be used as electrodes in rechargeable batteries (Müller et al., 1997).

Carbon black mixed with an insulating material form a semiconducting layer that placed between the metallic conductor and the insulating layer in the technology of high-voltage coaxial cables. This configuration avoids electrical breakdown. Carbon black can be replaced by undoped polyaniline. For low-electric fields, polyaniline behaves like an insulator, whereas under electric fields exceeding 1 kV/mm the intensity is no longer a linear function of the voltage, leading to a significant increase in polymer conductivity. In the areas in which the field locally increases owing to imperfections of the material, polyaniline becomes conductive and the accumulation of charge is dissipated.

39.8 FERROELECTRICITY, PYROELECTRICITY, AND PIEZOELECTRICITY IN POLYMERS

Ferroelectric crystals such as pyroelectric crystals exhibit spontaneous polarization, whereas piezoelectric crystals become polarized only under stress. Centrosymmetric crystals and isotropic systems present neither piezoelectric nor pyroelectric effects. The conditions that anisotropic systems must hold to display ferroelectric behavior are described elsewhere (Nye, 1987).

For piezoelectric systems $dP_i = d_{ijk} d\sigma_{jk}$ where P and σ are, respectively, the tensors of first and second order representing the polarization and the stress whereas d_{ijk} is a third-order tensor whose number of independent components due to the symmetry of the stress tensor reduces from 27 to 18. As the symmetry of the crystals augments, the number

of components of the system equal to zero increases (Tashiro, 1995; Riande and Díaz-Calleja, 2005). Making reduction of indices ($11 \rightarrow 1$, $22 \rightarrow 2$, $33 \rightarrow 3$, $3 \rightarrow 4$, $13 \rightarrow 5$, and $12 \rightarrow 6$) the polarization can be written as $dP_i = d_{ij}d\sigma_j$. For pyroelectric systems, $dP_i = p_i dT$, where \mathbf{p} is the pyroelectric vector, and T is the temperature. The SI units of d_{ij} and p_i are, respectively, C/N and C/(m²K). Ferroelectric crystals present the additional property of reversing their polarization by applying a sufficient large electric field. Ferroelectric materials exhibit also pyroelectricity and piezoelectricity, and pyroelectric polymers also present piezoelectric behavior. Ferroelectric materials present a limit temperature, named Curie temperature, above that the ferroelectric effect is lost.

Although piezoelectricity and pyroelectricity were firstly studied in inorganic crystals, polyvinylidene fluoride (PVDF) is one of the most important ferroelectric materials. The preparation of ferroelectric PVDF films involves stretching the film and then poling. During the uniaxial or biaxial orientation, usually carried out in the range 60–100° C, crystallites in the α form (dipoles in antiparallel direction) are converted into the high polarity β form (dipoles in parallel direction) (Tashiro, 1995). The structures of the α and β crystallite forms are shown in Figure 39.35. Let us assume that there are N dipoles of moment μ dispersed in film of area A and thickness of a poled sample of PVDF or any other ferroelectric polymer. The remnant polarization is $P_r = N\mu\langle\cos\theta\rangle/Al$, where $\langle\cos\theta\rangle$ is the average of the angles between the dipoles and the orientation axis (Furukawa, 1989). By applying a pressure perpendicular to the film, P_r changes as a result of the variation of the film dimensions and, as a result, the system exhibits piezoelectricity.

Change in dimensions has little effect on the pyroelectric behavior of ferroelectric materials. In this case, disorder of the orientation of dipoles by change in temperature is the main responsible for the pyroelectric effect that these materials present.

Polymers with flexible backbone with chiral side groups containing mesogenic moieties with a dipole moment perpendicular to the rod may display ferroelectric behavior

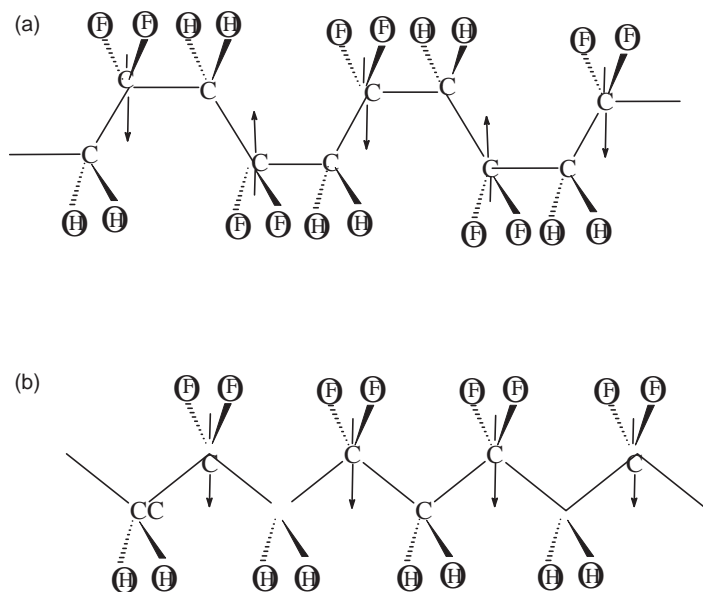


FIGURE 39.35 Configurations of PVDF chains in the α (a) and β (b) crystalline forms.

(Scherowsky, 1995). A scheme of a liquid crystal polymer displaying ferroelectric behavior is shown in Figure 39.21. Piezoelectric systems can also be prepared from amorphous polar polymers such as polyvinyl chloride and poly(vinyl acetate). At $T > T_g$, the dipoles of the polymer chains are oriented under the effect of an electric field, and then the temperature is lowered below T_g thus accomplishing frozen-in dipole orientation. The remnant polarization of the films in SI units is $P_r = \epsilon_0 \Delta \epsilon E_p$ where ϵ_0 is the dielectric permittivity in vacuum, $\Delta \epsilon$ is the dielectric relaxation strength and E_p is the polarization electric field.

Piezoelectricity is also displayed by uniaxial-oriented polymer chains containing chiral moieties. Systems of this kind include a variety of polymers such as poly(γ -benzyl-L-glutamate), cyanoethylcellulose, and isotactic polypropylene oxide. For illustrative purposes, the values of piezoelectric and pyroelectric coefficients for several polymers are shown in Table 39.1. PVDF and the VDF/TFE copolymers exhibit rather high piezoelectric and pyroelectric coefficients in comparison with other polymers. These polymers provide an alternate way to fabricate high quality ultra thin crystalline ferroelectric films (Bune, 1998; Duan, 2004).

The piezoelectric and pyroelectric properties of polymers are inferior to those exhibited by ceramic materials such as barium titanate and lead zirconate titanate (PZT). However, their brittle nature and inflexibility of inorganic materials limits their use (Mazur, 1995; Dias and Das-Gupta, 1996). A homogenous mixture of highly piezoelectric and pyroelectric ceramic material with a polymeric material, such as PVDF or PVDV, and Nylon, allows to obtain a composite that exhibits the piezoelectric and pyroelectric coefficients of ceramics with the flexibility, strength, and lightness of the polymer.

TABLE 39.1 Piezoelectric, Pyroelectric, and Permittivity Coefficients for Some Polymers (Furukawa, 1989)

Class 1: Poled Polymers		d_{31} (pC/N ⁻¹)	ϵ_{33}	p_3 (μ /Cm ² /K ⁻¹)
A. Ferroelectric polymers	PVDF ^a	30	0.2	35
	VDF/TrFE ^b	30	0.3	50
	VDF/TFE ^c	30	0.25	35
B. Polar polymers	VDCN/VAc ^d	8	0.25	20
	Nylon 11	4		5
	PVC ^e	0.2		1
C. Composite	PZT7PVDF ^f	-25	0.1	100
	PZT ^g	-180	0.6	270
Class 2: Drawn polymers		d_{14} , pC/N ⁻¹		
A. Chiral polymers	PBG ^h	1.7		
	CEC ⁱ	3.0		
	PHB ^j	1.5		

^aPoly(vinylidene fluoride).

^bVinylidene fluoride/trifluoroethylene copolymer.

^cVinylidene fluoride/tetrafluoroethylene copolymer.

^dVinylidene cyanide/vinyl acetate copolymer.

^ePoly(vinyl chloride).

^fLead zirconate titanate-PVDF composite.

^gLead zirconate titanate.

^hPoly(benzyl glutamate).

ⁱCyanoethyl cellulose.

^jPoly(β -hydroxy butyrate).

39.9 NONLINEAR POLARIZATION IN POLYMERS

Experiments indicate that the induced polarization of molecules easily polarizable is not a linear function of the electric field E and can be approximated by a Taylor series expansion (Buckingham, 1967)

$$P = \alpha E + (1/2)\beta EE + (1/6)\gamma EEE + \dots \quad (39.47)$$

where α is a second-order tensor, whereas β and γ named first and second hyperpolarizabilities, respectively, are third and fourth-order tensors, respectively.

Under electric fields of high intensity, second and third harmonics of polarization accompany the first or fundamental harmonic. Accordingly, a monochromatic laser beam at frequency ω traveling through a crystal produces a polarization 2ω , which radiates an electromagnetic wave of the same frequency (2ω), that propagates with the same velocity, monochromaticity and direction as the incident electromagnetic radiation of frequency ω . Symmetry operations dictate that neither isotropic system nor centrosymmetric crystals produce second-order harmonics. However, third-order harmonics can appear in isotropic systems.

It became clear in the 1980s that polymers with chromophore groups in their structure can display nonlinear optical behavior. However, owing to the fact that these polymers are amorphous, it is necessary to convert them into anisotropic materials in order that they can generate second-order harmonics or display electro-optical properties (see below). This can be accomplished by orientation of the dipoles of the chromophores in an electric field at temperatures well above T_g .

It has been suggested that a two-state model could be used to guide the design of second-order NLO chromophores (Buckingham, 1967). In the study of the hyperpolarizabilities, it was found (Oudar and Chemla, 1977) that $\beta = (\mu_{ee} - \mu_{gg})\mu_{ge}^2/E_{ge}^2$, where μ and E are the dipole matrix element and transition energy, respectively, between the ground state (g) and the excited state (e). In general (Chia-Cheng et al., 2005), molecules for second-order NLO applications were based simply on aromatic π -electron systems asymmetrically end-capped with electron donating and accepting groups to impart the directional bias. For example, 4-(*N,N*-dimethylamino)-4'-nitrostilbene (DANS). In DANS, $(\mu_\beta = (446 \sim 580) \times 10^{-48} \text{ esu})$ is a typical NLO chromophore. In this molecule, the two benzene rings and the double bond provide the conjugated π -system and the polarizable electrons; the dimethylamino group acts as the donor, and the nitro group acts as the acceptor.

NLO polymers exhibit large optical nonlinearities, which have ultrafast response because the electronic polarization is nearly instantaneous. At the macroscopic level, second-order harmonic generation under an electric field $E = E_0 \cos \omega t$ can be written as (Burland et al., 1994; Riande and Díaz-Calleja, 2005)

$$P_i(2\omega) = (1/2)e_0\chi_{ijk}(-2\omega, \omega, \omega)E_j(\omega)E_k(\omega) = e_0d_{ijk}(-2\omega, \omega, \omega)E_j(\omega)E_k(\omega) \quad (39.48)$$

where χ_{ijk} is the electro-optic susceptibility. Notice that χ_{ijk} is related to the first hyperpolarizability by

$$\chi_{ijk}(-2\omega, \omega, \omega) = (1/2)e_0Nf_i(-2\omega)f_j(\omega)f_k(\omega) < \beta_{ijk}(-2\omega, \omega, \omega) >_{ijk} \quad (39.49)$$

where N is the number of chromophores groups, f_i is the Lorentz field factor and the angular brackets denote the weighted sum of the polarizability of the molecule over all its orientations in the polymer. In SI units, the third-order susceptibility tensor χ is given in m/V.

Polymers containing chromophore groups with their dipoles duly oriented display the electro-optic effect, that is, the dielectric permittivity of the polymer changes by effect of an electric field. This phenomenon is known as Pockells effect. The change in permittivity is reflected in the index of refraction so that (Kaminov, 1974)

$$\Delta n_{IJ} = -(n^3/2)r_{IJK}E_K \quad (39.50)$$

where r_{IJK} is the electro-optic coefficient. Change in the permittivity (index of refraction) causes a change in the frequency of an electromagnetic radiation traveling through the poled material. Then the electro-optic effect plays an important role in the transmission of codified information, whereas the second-order effect is important in information storage. Making the reduction of indices $11 \rightarrow 1$, $22 \rightarrow 2$, $33 \rightarrow 3$, $3 \rightarrow 4$, $13 \rightarrow 5$, and $12 \rightarrow 6$, the d_{IJK} and r_{IJK} tensors are expressed as d_{IJ} and r_{IJ} , respectively. NLO chromophore that possesses a large first hyperpolarizability β is essential to ensure large optical nonlinearities in bulk polymers. The second harmonic coefficient d_{33} and the linear EO coefficient r_{33} are directly related to the second-order NLO susceptibility and EO susceptibility, respectively.

In the glassy state, micro-Brownian motions freeze and dipolar orientation is preserved. Therefore, to maintain the anisotropy of the system, the glass transition temperature of the poled polymers should be high. On the other hand, during devices processing, the temperatures of NLO polymers can be up to 300°C. Moreover, these materials may endure temperatures of the order of 100°C for long times during operation conditions (Kaminov, 1974; Samyn et al., 2001). As a result, NLO polymers should exhibit good mechanical and thermal properties. For frequency doubling devices, the second-order susceptibility should be of the order of 60 pm/V, whereas the electro-optical coefficient for electro-optical applications should be larger than 10 pm/V. Chromophore moieties often used in the preparation of NLO polymers based on poly(metacrylic acid) are shown in Figure 39.36. Values of the susceptibility and electro-optical polymers for copolymers with different chromophores are given in Table 39.2.

NLO polymers with suitable mechanical and thermal properties can be prepared from polyimides containing both rigid segments and chromophore groups in the main chain. However, the rigidity of this configuration hinders the orientation of the dipole associated with the chromophore in an electric field, the necessary step to convert the material into an anisotropic one. A better molecular design is the use of rigid segments in the backbone and chromophore groups as lateral substituents (Peng and Yu, 1994; Yang et al., 1994, 1996; Yu and Yu, 1994). A representative structure of this kind is shown in Figure 39.37. NLO polymers can also be obtained from epoxy resins (Jungbauer et al., 1990) and cross-linked polyurethanes (Chen et al., 1992; Shi et al., 1992) containing in both cases chromophore groups. Since molecular mobility in thermosets is severely hindered, orientation of the dipoles associated with the chromophore groups must be performed before gelification takes place. This can be accomplished by carrying out the cross-linking reaction under poling (see Figure 39.38). Dendrimeric structures containing NLO moieties with different configurations have also been reported (Ma et al., 2001; Kajzar et al., 2002).

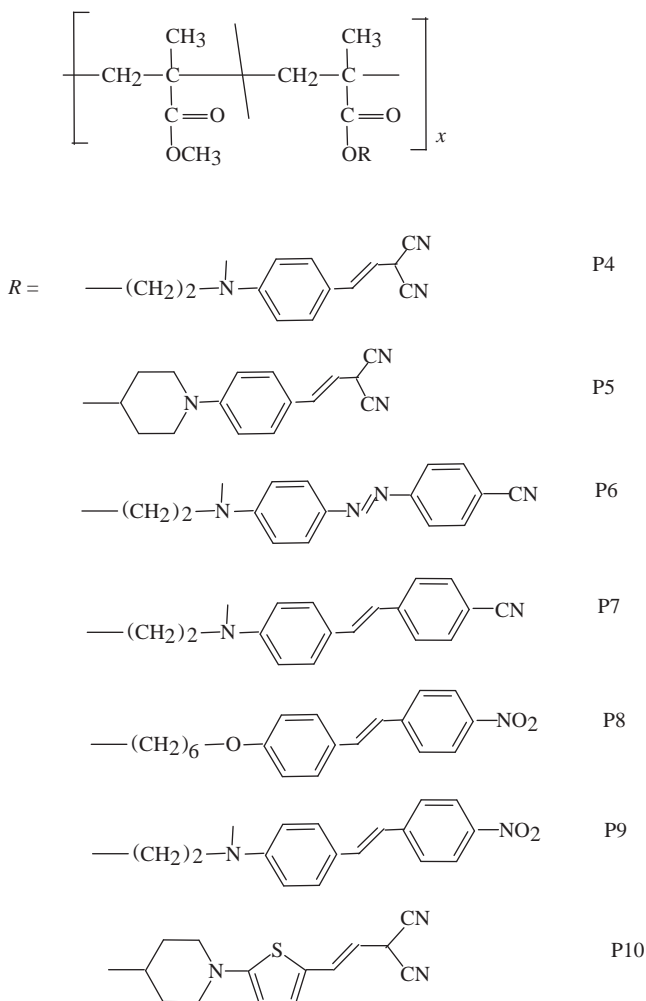


FIGURE 39.36 Copolymers of methyl methacrylate with methacrylates containing chromophore groups.

Ferroelectric polymers such as poly(vinylidene fluoride) and its copolymers vinylidene/fluoride, vinylidene cyanide copolymers and ferroelectric liquid crystalline polymers exhibit interesting nonoptical properties. For a survey with the description of NLO polymers (see Chia-Cheng et al., 2005).

39.10 ELASTOMERS FOR ACTUATORS AND SENSORS

Well above the glass transition temperature, cross-linked polymers or elastomers behave such as rubber materials, displaying large deformations under mechanical force fields. In this situation, stress-strain relationships in the elastomer become nonlinear. In the linear theory of elasticity, solutions of traction boundary problems are unique and the stress

TABLE 39.2 Glass Transition Temperatures and SHG Results for Polymethacrylate Copolymers of Figure 39.36 (Samyn et al., 2001)

Polymer	$\omega_{\text{dye-monomer}}$	T_g (°C)	d_{33} (pm/V)
P4-10	0.088	119	12.1
P4-20	0.18	118	21.8
P4-35	0.31	109	26.4
P4-100	1.00	120	65.8
P5-20	0.17	135	45.7
P5-40	0.41	142	20.1
P5-60	0.61	131	9.0
P6-10	0.119	129	5.2
P6-20	0.164	128	21
P6-35	0.315	134	68
P6-40	0.440	128	45
P6-60	0.490	125	32
P6-80	0.717	124	31
P6-100	1.000	181	26
P7-10	0.103	107	8.8
P7-20	0.164	107	11.5
P7-30	0.225	75	12
P7-50	0.430	75	12
P8-30	0.295	98	23.6
P9-20	0.190	120	31.4
P10-20	0.148	135	41.5

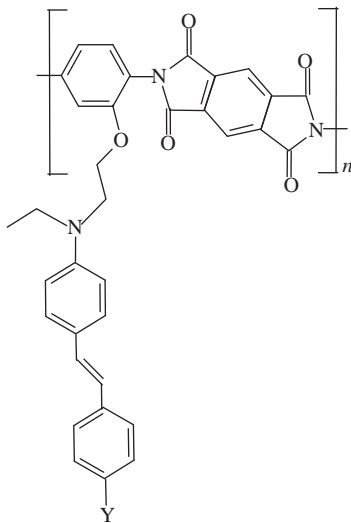


FIGURE 39.37 Polyimide containing a chromophore side group in the structural unit.

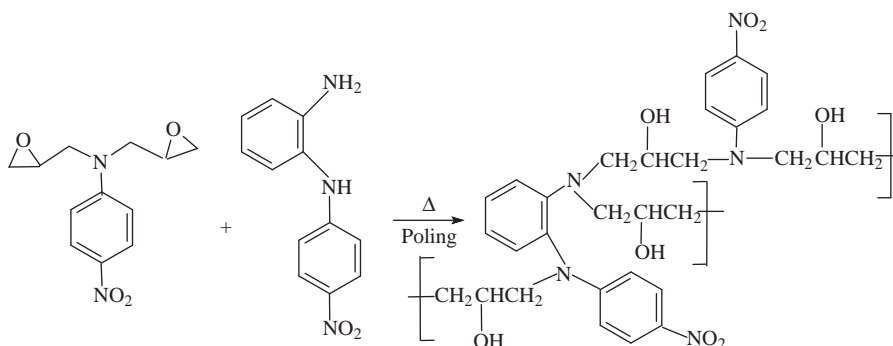


FIGURE 39.38 Scheme of the synthesis of a NLO epoxy resin containing chromophore groups. The condensation is carried out under the action of an electric field to accomplish dipoles orientation.

fields are also unique. This is a consequence of the linearity of the constitutive equations. However, under finite deformations, those relations are nonlinear and the uniqueness of the solutions is not warranted. In some circumstances, this could give rise to bifurcation phenomena (Liu, 2002).

Rubbery materials, as dielectric polymer materials, also present a dipolar moment that deforms under the action of electric force fields. These rubbers are referred as electromechanically active elastomers (EMAE) (Eringen, 1990). More recent theoretical foundations of the electroelasticity are described elsewhere (Dorfmann and Ogden, 2005, 2006). It is worth noting that the field may cause some dielectrics to become thinner and other thicker (Xu et al., 2002; Liu et al., 2002). They can convert electrical into mechanical energy and vice-versa (Xu et al., 2001). In some cases, the mechanical and electrical force fields enter in competition between them, thus giving rise to interesting changes in the instability and bifurcation phenomena in these materials (Díaz-Calleja et al., 2009).

Potential applications of EMAE include biomedical prostheses, actuators, energy harvesters, and robotics. In particular, the electrostriction of elastomeric polymer dielectrics with compliant electrodes is potentially useful as a small-scale, solid-state actuator technology (Peltine et al., 1998). These materials combine an efficient fast response with relatively high strains (>30%), good actuation pressures (up to 1.9 MPa), and high specific energy densities (up to 0.1 J/g).

The design of active structures (robots) based on actuators follows an approach in which a passive skeletal structure composed of rods and joints is augmented by actuators, which apply force to the skeleton, such that it changes its shape (Kofod and Wirges, 2007). In this case, the proposed technology rests on the principle of minimal surfaces found in nature. The principle of the complex structure involved is substantiated with a model actuator consisting of a sheet of stiff plastic bent to a shell to which ends the elastomer is attached. Compliant electrodes are applied to both sides of the elastomer sheet, thus forming an elastomeric capacitor driven by Maxwell stress. The analytical model of this structure is based on a free energy function, which is the addition of the free energy of the elastomer (the simplest case corresponding to the Gaussian limit of a freely jointed chain), the elastic energy of the curved beam, and the voltage-controlled energy of the capacitor. As usual, the minimum value of the energy corresponds to the equilibrium

state of the system. This equilibrium state can be modified by changing the applied voltage. In this way, minimum-energy structures can be produced. The advantages of this type of dielectric elastomers are simplicity in manufacture, an inherent push–pull configuration, and the existence of a saturation voltage.

However, these actuators have experienced in some cases high rates of failure that have prevented their practical application. The large-scale failure modes of dielectric elastomer actuators have been studied by Plante and Dubowsly (2006). Using a model elastomer proposed by Ogden (1972), these authors developed a theory based on continuum mechanics (Holzapfel, 2000), and establish three large-scale failure criteria: material strength, dielectric strength and pull-in. These criteria depend on the material and on the stretch rate. In spite of the promising applications, the properties of dielectric actuators still need to be improved.

39.11 ELECTRICAL BREAKDOWN IN POLYMERS

Dielectrics are not an exception to the “first rule” of materials science that states that “everything can be broken.” In increasing the voltage applied to a capacitor, a voltage may be reached at which an electrical breakthrough across the dielectric inside the capacitor occurs, a phenomenon called “electrical breakdown”. This fact is made evident by an irreversible—and practically always destructive—sudden flow of current within very short times, of the order of 10^{-8} s. The voltage at which electrical breakdown occurs is not a well-defined property because depends on several parameters, the most important being the nature of the chemical and physical characteristics of the material itself, production process, thickness, temperature, internal structure of the material, boundary surfaces, age, field stress history, and environment, especially the humidity of the environment in which the material is used.

Polymer materials are characterized for displaying a wide range of mechanical strength and stiffness. These characteristics are coupled with high corrosion resistance, ease of processing and low cost. Moreover, the very high breakdown strengths ($\sim 5 \times 10^6$ up to $\sim 10^8$ V/m) most polymers present, together with the high DC resistivities (typically $> 10^{16}$ Ω m) and low dielectric losses (in many cases $\tan \delta < 0.001$) make these materials suitable for insulating purposes. Obviously, the voltage at the electrical breakdown depends on the electrical strength in such a way that the voltage necessary for this phenomenon to occur decreases with decreasing films thickness.

Electrical treeing degradation has been detected in inorganic as well polymeric materials and the mechanism of this process must be independent on the chemical nature of the insulation. The origin of the electrical tree could be traced to asperities on the conductor or metallic particles embedded in the polymer during the extrusion. If semiconductor layers are simultaneously laid down by means of a triple extrusion technique over the inner conductor and the outer insulation, then electrical trees initiated in this way are substantially reduced. However, treeing degradation could still persist caused by water trees. Water trees occur in a variety of polymers (Xu et al., 2001) including polyethylene, (e.g., ethylene-polypropylene rubber [EPR]), poly(vinyl chloride) (PVC), polycarbonate, aromatic derivatives such as polystyrene, all of them below and above the glass transition temperature, as well as in thermosets. The feature making the path of a water tree is the presence of a large density of spherical microvoids of radius 1–5 μ m (Shaw and Shaw, 1984).

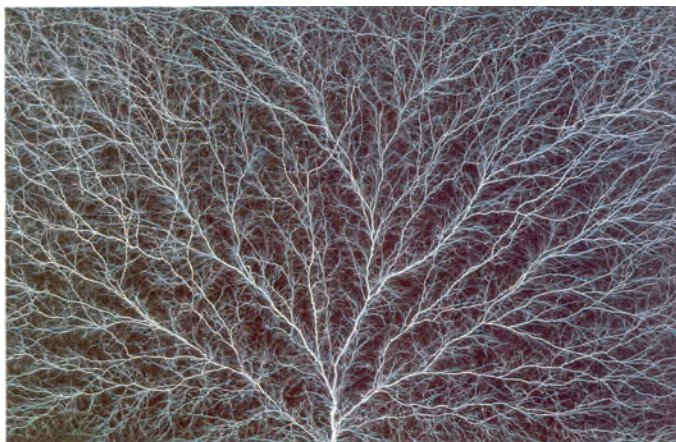


FIGURE 39.39 Discharge in a block of poly(methyl methacrylate) previously charged with a 2 MeV electron beam and discharged through a point contact. Taken from *Fractals and Disordered Systems* (Figure 8.1) (Shaw and Shaw, 1984).

Dielectric breakdown is the fastest way to prepare a fractal structure. For example, it is possible to obtain a decorative “frozen lightning” by electron beam charging of an insulating material such as poly(methyl methacrylate) followed by a discharge through a point connector (Knizhnik et al., 1982). The tree-like patterns resulting from the experiment, shown in Figure 39.39, have a fractal structure similar to of clusters described by the diffusion limited aggregation model (Bunde, 1991). A more detailed information on fractals related with electrical treeing can be found elsewhere (Dissado and Fotherhill, 1992).

Electrical aging or deterioration of the electric strength properties of polymers, such as polyethylene and polypropylene, by the actions of electric fields decreases their electrical breakdown strengths. Up to date (Zakrevskii et al., 2003), the mechanism responsible for electrical breakdown in insulating polymers is a not well-understood issue yet. For example, the short free path length in insulating polymers makes highly improbable that electron avalanches occur in them. As a result, it is inapplicable the concept that electrical breakdown arises from an electron avalanche occurring at some critical electric field intensity E_{cr} when conditions for the impact ionization of molecular chains by fast electrons are created. The experimental evidence at hand suggests that the electrical breakdown is not a critical phenomenon, but a kinetic process that develops in time. The breakdown itself involves the formation of a conducting channel, and it is a final stage of polymer degradation in the electric field. This fact arises from the damage accumulation process whose rate depends on the electric field intensity. It is believed that after a cavity or a low density region in the material is formed, erosion of the polymer sample induced by discharges gives rise to a rapid propagation of a conducting channel through the sample, that is, to the sample breakdown. The mechanism of formation of pore-like or low density structures under the action of an electric field is a nonsolved major problem in the theory of electrical strengths of polymers. A thorough information concerning the physics involved in electrical breakdown can be found elsewhere (Dissado and Fotherhill, 1992).

REFERENCES

- Adachi K, Kotaka T. *Progress in Polymer Science* 1993;18:585.
- Aharoni SM. *Macromolecules* 1979;12:537.
- Aharoni SM, Walsh EK. *Macromolecules* 1979;12:271.
- Alvarez C, Lorenzo V, Riande E. *Journal of Chemical Physics* 2005;122:194905.
- Appetechi GB, Croce F. *Science* 1998;394:456.
- Armand MB, *Polymer Electrolyte Reviews*, Vol. 1. England: Elsevier Essex; 1987. Chap. 162.
- Attard GS, Araki K, Moura-Ramos JJ, Williams G. *Liquid Crystals* 1988;3:861.
- Blasco Cantera F, Riande E, Almendro JP, Saiz E. *Macromolecules* 1981;14:138.
- Boese D, Kremer F. *Macromolecules* 1990;23:829.
- Boyd RH. *Polymer Journal* 1985;26:323;1133.
- Boyd RH. Dielectric spectroscopy of semicrystalline polymers. In: Runt HP, Fitzgerald JJ, editors. *Dielectric Spectroscopy of Polymeric Materials*. Washington, DC: American Chemical Society; 1997. Chap. 4.
- Bredas JL, Scott JC, Yakushi K, Street GB. *Physical Review B* 1984;30:1023.
- Bredas JL, Chance RR, Silbey R. *Physical Review B* 1982;26:58431.
- Buckingham AD. *Advances in Chemical Physics* 1967;12:107.
- Bunde A, Havlin S, editors. *Fractal and Disordered Systems*. Berlin Heidelberg: Springer; 1991. Chaps. 4 and 8.
- Bune AV, et al. *Nature (London)* 1998;391:874.
- Burland DM, Miller RD, Welsh CA. *Chemical Reviews* 1994;94:31.
- Cairns EJ, McLaren FR. Status of batteries for energy storage applications. 183rd Meeting of the Electrochemical Society, May 16–21, Honolulu, Hawaii; 1993.
- Chen M, Dalton LR, Xu LP, Shi XQ, Steier WH. *Macromolecules* 1992;25:4032.
- Chia-Cheng Chang; Chih-Ping Chen; Che-Chung Chou; Wen-Jang Kuo; Ru-Jong Jeng *Polymer Reviews* 2005;45:125.
- Cole RH. *The Journal of Physical Chemistry* 1975;79:1459, 1469.
- Cole RH. *Annual Review of Physical Chemistry* 1977;28:283.
- Cole RH, Mashimo S, Winsor P. *The Journal of Physical Chemistry* 1980;84:786.
- Collins PJ. *Liquid Crystals*. Bristol: Zadam Hilger; 1990.
- Compañ V, Fernández-Carretero FJ, Riande E, Linares A, Acosta JL. *Journal of the Electrochemical Society* 2007;154:B159.
- Croce E, Appetecchi GB, Persi L, Scrosati B. *Nature* 1998;394:456.
- Croce E, Persi L, Ronci F, Scrosati B. *The Journal of Physical Chemistry B* 1999;103:10632.
- Croce E, Persi L, Ronci F, Scrosati B. *Solid State Ionics* 2000;315:47.
- de Gennes PG, Prost J. *The physics of Liquid Crystals* Oxford: Calrendon Press; 1995.
- Debye P. *Polar Molecules*. New York: Dover; 1936.
- Dias CJ, Das-Gupta DG. *IEEE Transactions on Dielectrics and Electrical Insulation* 1996;3:706.
- Díaz-Calleja R, Domínguez-Espinosa G, Riande E. *Non-Crystalline Chalcogenides* 2007;353:719.
- Díaz-Calleja R, Riande E. Calculations of dipole moments and correlation parameters in polymers. In: Runt HP, Fitzgerald JJ, editors. *Dielectric Spectroscopy of Polymeric Materials*. Washington, DC: American Chemical Society; 1997.
- Díaz-Calleja R, Riande E, San Román J. *Journal of Non-Crystalline Solids* 1991a;131–133:852.
- Díaz-Calleja R, Riande E, San Román J. *Macromolecules* 1991;24:264.

- Díaz-Calleja R, Riande E, San Román J. *Macromolecules* 1992;25:2875.
- Díaz-Calleja R, Sanchis MJ, Riande E. *Journal Electrostatics* 2009;67:158.
- Díaz-Calleja R, Sanchis MJ, Riande E, Pérez E, Pinto M. *Journal of Molecular Structure* 1999;479:135.
- Dissado LA, Fotherhill JC. *Electrical Degradation and Breakdown in Polymers*. London: Peter Peregrinus Ltd.; 1992.
- Domínguez-Espinosa G, Díaz-Calleja R, Riande E. *Journal of Chemical Physics* 2005;123:114904.
- Domínguez-Espinosa G, Díaz-Calleja R, Riande E, Gargallo L, Radic D. *Macromolecules* 2006a;39:3071.
- Domínguez-Espinosa G, Díaz-Calleja R, Riande E. *Macromolecules* 2006b;39:5043.
- Domínguez-Espinosa G, Ginestar D, Sanchis MJ, Díaz-Calleja R, Riande E. *Journal of Chemical Physics* 2008;129:104513.
- Dorfmann A, Ogden RW. *Acta Mechanica* 2005;174:167.
- Dorfmann A, Ogden RW. *Journal of Elasticity* 2006;82:99.
- Duan Ch. et al. *Physical Review Letters* 2004;69:235106.
- Epstein AJ. *MRS Bulletin* 1997;22:16.
- Eringen AC, Maugin GA, editors. *Electromechanics of Continua*. Berlin, Heidelberg, New York: Springer; 1990.
- Fatuzzo E, Mason PR. *Proceedings of the Royal Society of London* 1967;90:741.
- Feldman Y. *Review of Scientific Instruments* 1996;67:3208.
- Ferry JD. *Viscoelasticity Properties of Polymers*, 3rd ed. New York: Wiley; 1980.
- Flory PJ. *Statistical Mechanics of Chain Molecules*. New York: Wiley; 1969.
- Friend RH, Greenham NC. Electroluminescence in conjugated polymers. In: Stokheim TA, Elsembaumer RL, Reynolds JR, editors. *Handbook of Conducting Polymers*. 2nd ed. New York: Marcel Dekker; 1998. p. 823–845.
- Fröhlich H. *Transactions of the Faraday Society* 1948;44:238.
- Fröhlich H. *Theory of Dielectrics*. 2nd ed. Oxford University Press, 1958.
- Frommer JE, Chance RR. *Encyclopedia of Polymers*. New York: Wiley; 1996;5:462.
- Fulcher GS. *Journal of the American Ceramic Society* 1925;8:339.
- Furukawa T. *IEEE, Transactions on Electrical Insulation* 1989;24:375.
- Guggenheim EA. *Transactions of the Faraday Society* 1949;45, 714.
- Guggenheim EA. *Transactions of the Faraday Society* 1951;47, 573.
- Havriliak Jr. S, Havriliak SJ. *Dielectric and Mechanical Relaxation in Materials*. Munich: Hanser; 1997.
- Havriliak Jr. S, Negami S. *Polymer Journal* 1987;8:161.
- Heijboer J. Mechanical properties of glassy polymers containing saturated rings. Ph.D. thesis. TNO, Leiden; 1972.
- Hickner MA, Ghassemi H, Kim YS, Einsla BR, McGrath JE. *Chemical Reviews* 2004;104:4587.
- Hide F, García MA, Schwartz BJ, Heeger AJ. *Accounts of Chemical Research* 1997;30:430.
- Hino TJ. *Journal of Applied Physics* 1975;46:1956.
- Holzapfel GA. *Nonlinear Solid Mechanics*. Chichester, UK: Wiley; 2000.
- Huang Y, Saiz E, Ezquerro T, Guzmán J, Riande E. *Macromolecules* 2002;35:2926.
- Jackson JD. *Classical Electrodynamics*. 2nd ed. New York: Wiley; 1974.
- Joo J, Oblakowski Z, Du G, Pouget JP, Oh EJ, Weisinger JM, Min Y, MacDiarmid AG, Epstein AJ. *Physical Review B* 1994;69:2977.

- Jungbauer D, Reck B, Twieg R, Yoon D, Wilson C, Swallen J. *Journal of Applied Physics Letter* 1990;56:2610.
- Kajzar, F, Lex, K-S, Jen, AK-Y. *Advances in Polymer Science* 2002;161:1.
- Kaminov I. *An Introduction to Electro-optic Devices*. New York: Academic Press; 1974. p. 56.
- Kido J, Hongawa K, Okuyama K, Nagai K. *Applied Physics Letters* 1994;64:815.
- Kirkwood JG. *Journal of Chemical Physics* 1939;7:911.
- Knizhnik EI, Onisco AD, Gaydamaka AV. *Radiation Physics and Chemistry* 1982;19:473.
- Kofod G, Wirges W. *Applied Physics Letters* 2007;90:81916.
- Kohlman RS, Joo J, Epstein AJ. Conducting polymers: electrical conductivity. In: Mark JE, editor. *Physical Properties of Polymers Handbook*. Woodbury, NY: American Institute of Physics; 1996.
- Kohlrausch R. *Annals of Physics* 1847;12:393.
- Koksang R, Olsen II, Shackle D. *Solid State Ionics* 1994;69:320.
- Kraft A. *Journal of Chemical Physics* 2001;2:163.
- Kremer F. Broadband dielectric spectroscopy on collective and molecular dynamics In: Runt JP, Fitzgerald JJ, editors. *Ferroelectric Liquid Crystals in Dielectric Spectroscopy of Polymeric Materials*. Washington, DC: American Chemical Society; 1997.
- Kremer F, Schönhals A, editors. *Broadband Dielectric Spectroscopy*. Berlin, Heidelberg: Springer; 2003.
- Kremer F, Arndt M. Broadband dielectric measurement techniques. In: Runt HP, Fitzgerald JJ, editors. *Dielectric Spectroscopy of Polymeric Materials*. Washington, DC: American Chemical Society; 1997. Chap. 2.
- Kreuer KD. *Chemistry of Materials* 1996;8:610.
- Kreuer K, Paddison SJ, Spohr E, Schuster M. *Chemical Reviews* 2004;104:4637.
- Lakshminarayanan N. *Transport Phenomena in Ion-Exchange Membranes*. London: Academic Press; 1966.
- Larminie J, Dicks A. *Fuel Cell Systems*. 2nd ed. West Sussex, England: Wiley; 2003.
- Leising G, Tasch S, Graupner W. Fundamentals of electroluminescence in paraphenylene type conjugated polymers and oligomers. In: Stokheim TA, Elsembaumer RL, Reynolds JR, editors. *Handbook of Conducting Polymers*. 2nd ed. New York: Marcel Dekker; 1998. p. 847–880.
- Liu I-S. *Continuum Mechanics*. Berlin: Springer; 2002.
- Liu SF, Park SE, Cross LE, Shrotr TR. *Journal of Applied Physics* 2002;92:461.
- Ma, H, Chen, BQ, Sassa, T, Dalton, LR, Jen, AK-Y. *Journal of the American Chemical Society* 2001;123:986.
- MacCallum JR, Vincent CA, editors. *Polymer Electrolyte Reviews*, Vol. 1. England: Elsevier Essex; 1987.
- MacCrum NG, Read BE, Williams G. *Anelastic and Dielectric Effects in Polymeric Solids*. New York: Wiley; 1967; Dover, 1990.
- Mandelkern L. *Crystallization of Polymers*. Cambridge University Press; 2004.
- Mark JE. *Journal of Chemical Physics* 1971;56:451.
- Mark JE. *Journal of Chemical Physics* 1972;57:2541.
- Mazur K. Polymer-ferroelectric ceramic composites. In: Nalwa HS, editor. *Ferroelectric Polymers*. New York: Marcel Dekker; 1995. Chap. 11.
- McClellan AL. *Tables of Dipole Moments*. El Cerrito, CA: Rahara Enterprises; 1974.
- McGehee MD, Heeger AJ. *Advanced Materials* 2000;12:1655.
- Meyer WH. *Advanced Materials* 1998;10:439.

- Mikhailov GP, Borisova TI. *Polymer Science USSR* 1961;2:387.
- Mizes HA, Conwell EM. *Physical Review B* 1994;50:11243.
- Mopsik FI. *Review of Scientific Instruments* 1984;55:79.
- Moratti SC. The chemistry and use of poly(*p*-phenylene vinylene) In: Skotheim TA, Elsenbaumer RL, Reynolds JR, editors. *Handbook of Conducting Polymers*. New York: Marcel Dekker; 1998. p. 341–361.
- Moscicki JK, Williams G, Aharoni SM. *Macromolecules* 1982;15:642.
- Müller K, Santhanam KSV, Haas O. *Chemical Reviews* 1997;97:207.
- Müller-Plathe F, van Gunsteren WF. *Journal of Chemical Physics* 1995;103:4745.
- Nye JF. *Physical Properties of Crystals*. Oxford: Clarendon Press; 1987.
- O'Hayre R, Cha S-W, Colella W, Printz FB. *Fuel Cell Fundamentals*. New York: Wiley; 2006.
- Ogden RW. *Proceedings of the Royal Society of London A* 1972;326:565.
- Onsager L. *Journal of Chemical Physics* 1936;58:1486.
- Oudar JL, Chemla DS. *Journal of Chemical Physics* 1977;66:2664.
- Paddison SJ. *Annual Review of Materials Research* 2003;33:289.
- Pelrine RE, Kornbluh RD, Joseph JP. *Sensors and Actuators A* 1998;64:77.
- Peng Z, Yu L. *Macromolecules* 1994;27:2638.
- Perlman MM, Unger SJ. *Journal of Applied Physics* 1974;45:2389.
- Plante JP, Dubowsky S. *International Journal of Solids and Structures* 2006;43:7727.
- Pron A, Rannou P. *Progress in Polymer Science* 2002;27:135.
- Riande E, Mark JE. *Macromolecules* 1978;11:956.
- Riande E. Transport phenomena in ion-exchange membranes. In: *Physics of Electrolytes*. London: Academic Press; 1972.
- Riande E, Díaz-Calleja R. *Electrical Properties of Polymers*. New York: Marcel Dekker; 2005. Chaps. 8, 12, 13.
- Riande E, Saiz E. *Dipole Moments and Birefringence of Polymers*. Englewood Cliffs, NJ: Prentice Hall; 1992. Chap. 5.
- Ribes-Greus A, Diaz-Calleja R. *Journal of Applied Polymer Science* 1989;38:1127.
- Rikukawa M, Sanui K. *Progress in Polymer Science* 2000;25:1463.
- Rozière J, Jones DJ. *Annual Review of Materials Research* 2003;33:503.
- Saiz E, Riande E, Delgado MP, Barrales-Rienda JM. *Macromolecules* 1982;15:1152.
- Salmerón M, Barrales-Rienda JM, Riande E, Saiz E. *Macromolecules* 1984;17:2728.
- Samyn C, Verbiest T, Persons A. *Macromolecular Rapid Communications* 2001;21:1.
- Sata T. *Ion-Exchange Membranes*. Cambridge: RSC; 2004.
- Schönfeld A, Kremer F, Poths H, Zentel R. *Molecular Crystals and Liquid Crystals* 1994;49:254.
- Schonhals A. Dielectric properties of amorphous polymers. In: Runt HP, Fitzgerald JJ, editors. *Dielectric Spectroscopy of Polymeric Materials*. Washington, DC: American Chemical Society; 1997.
- Scherowsky G. Ferroelectric liquid crystal (FLC) polymers. In: Nalwa HS, editor. *Ferroelectric Polymers*. Marcel Dekker; 1995. Chap. 10.
- Schuster MF, Meyer WH. *Annual Review of Materials Research* 2003;33:233.
- Scrosati B. *Chimica e Industria (Milan)* 1997;79:463.
- Scrosati B. *Polymer International* 1998;47:50.
- Scrosati B. *Chemistry Record* 2001;1:173.
- Sirringhaus H, Tessler N, Friend RH. *Science* 1990;280:1741.

- Shaw MT, Shaw SH. *IEEE Transactions Electrical Insulation* E-I-I9, 1984; 419.
- Shi Y, Steier WH, Chen M, Yu L, Dalton LR. *Applied Physics Letter* 1992;60:2577.
- Shimizu H, Nakayama KJ. *Journal of Applied Physics* 1991;74:1597.
- Shirakawa H, Ikeda S. *Polymer Journal* 1971;2:231.
- Shirakawa H, Ikeda S. *Journal of Polymer Science, Polymer Chemistry Edition* 1974;12:929.
- Shirakawa H, Louis EJ, MacDiarmid AG, Chiang CK, Heeger AJ. *Journal of the Chemical Society: Chemical Communications*. 1977; 578.
- Smith JW. *Transactions of the Faraday Society* 1950;46:394.
- Stephan AM, Nahm AS. *Polymer Journal* 2006;47:5952.
- Stockmayer WH. *Pure and Applied Chemistry* 1967;15:539.
- Stockmayer WH, Burke JJ. *Macromolecules* 1969;2:647.
- Suter UW, Mattice WL. *Conformational Theory of Large Molecules*. New York: Wayne; 1994.
- Tammann G, Hesse W. *Z Anorg Allgem Chemistry* 1926;156:245.
- Tashiro K. Crystal structure and phase transition of PVDF and related copolymers. In: Nalwa HS, editor. *Ferroelectric Polymers*. New York: Marcel Dekker; 1995. Chap. 2.
- Tetsutani T, Kakizaki M, Hideshima T. *Polymer Journal*. 1982;14:305.
- Titulaer UM, Deutch JM. *Journal of Chemical Physics* 1974;60:1502.
- Tressler N. *Advanced Materials* 1999;11:363.
- van Turnhout J. *Thermally stimulated discharge of polymer electrets*. Ph.D. thesis. Leiden; 1972.
- Vogel H. *Physik Z* 1921;27:645.
- Volkenstein MV. *Configurational Statistics of Polymeric Chains*. New York: Wiley; 1963.
- Vright PV. *MRS Bulletin* 2002;27:597.
- Watanabe M, Ogata N. *Polymer Electrolyte Reviews*, Vol 1. England: Elsevier Essex; 1987. Chap. 2.
- Williams G. *Advances in Polymer Science* 1979;33:59.
- Williams G, Watts DC. In: Diehl P, Flick E, Kosfeld E, editors. *NMR Basic Principles and Progress*. Vol. 4. Berlin: Springer; 1971. p. 271.
- Würthner F. *Angewandte Chemie, International Edition* 2001;40:1037.
- Xu H, Cheng Z-Y, Olson D, Mai T, Zang QM, Kavarnos G. *Applied Physics Letter* 2001;78:2360.
- Xu T-B, Cheng Z-Y, Zang QM. *Applied Physics Letter* 2002;80:1082.
- Yan M, Rothberg LJ, Papadimitrakopoulos F, Galvin ME, Miller TM. *Physical Review Letters* 1994;72:156.
- Yang Y, Pei Q. *Journal of Applied Physics* 1995;77:4807.
- Yang S, Peng Z, Yu L. *Macromolecules* 1994;27:5858.
- Yu G, Gao J, Hummelen JC, Wudl F, Heeger AJ. *Science* 1995;270:1789.
- Yu D, Gharavi A, Yu L. *Macromolecules* 1996;29:6139.
- Yu D, Yu L. *Macromolecules* 1994;27:6718.
- Zakrevskii VA, Sudar NT, Zaopo A, Dubitsky YA. *Journal of Applied Physics* 2003;93:2135, (and references therein).

40

NONDESTRUCTIVE INSPECTION*

ROBERT L. CRANE AND JEREMY S. KNOPP

- 40.1 Introduction
 - 40.1.1 Information on inspection methods
 - 40.1.2 Electronic references
 - 40.1.3 Future NDE capabilities
- 40.2 Liquid penetrants
 - 40.2.1 Penetrant process
 - 40.2.2 Reference standards
 - 40.2.3 Limitations of penetrant inspections
- 40.3 Radiography
 - 40.3.1 Generation and absorption of X radiation
 - 40.3.2 Neutron radiography
 - 40.3.3 Attenuation of X radiation
 - 40.3.4 Film-based radiography
 - 40.3.5 Penetrameter
 - 40.3.6 Real-time radiography
 - 40.3.7 Computed tomography
- 40.4 Ultrasonic methods
 - 40.4.1 Sound waves
 - 40.4.2 Reflection and transmission of sound
 - 40.4.3 Refraction of sound
 - 40.4.4 Inspection process
 - 40.4.5 Bond testers
- 40.5 Magnetic particle method
 - 40.5.1 Magnetizing field
 - 40.5.2 Continuous versus noncontinuous fields
 - 40.5.3 Inspection process
 - 40.5.4 Demagnetizing the part
- 40.6 Thermal methods
 - 40.6.1 IR cameras
 - 40.6.2 Thermal paints

* Reprinted from *Mechanical Engineers' Handbook*, Vol. 1, Wiley, New York, 2006, with permission of the publisher.

Handbook of Measurement in Science and Engineering. Edited by Myer Kutz.
Copyright © 2013 John Wiley & Sons, Inc.

- 40.6.3 Thermal testing
- 40.7 Eddy current methods
 - 40.7.1 Eddy current inspection
 - 40.7.2 Probes and sensors
- References

40.1 INTRODUCTION

This chapter deals with the nondestructive inspection of materials, components, and structures. The term nondestructive inspection (NDI) or nondestructive evaluation (NDE) is defined as that class of physical and chemical tests that permit the detection and/or measurement of significant properties or the detection of defects in a material without impairing its usefulness. The inspection process is often complicated by the fact that many materials are anisotropic, and most NDI techniques were developed for isotropic materials such as metals. The added complication due to the anisotropy usually means that an inspection is more complicated than it would be with isotropic materials.

Inspection of complex materials and structures is frequently carried out by comparing the expected inspection data with a standard and noting any significant deviations. This means a well-defined standard must be available for calibration of the inspection instrumentation. Furthermore, standards also must contain implanted flaws that mimic those that naturally occur in the material or structure to be inspected. Without a well-defined standard to calibrate the inspection process, the analysis of NDI results can be significantly in error. For example, to estimate the amount of porosity in a cast component from ultrasonic measurements, standard calibration specimens with calibrated levels of porosity must be available to calibrate the instrumentation. Without such standards, estimation of porosity from ultrasonic data is a highly speculative process.

This chapter covers some important and some less well-known NDI tests. Since information on the less frequently used tests is not generally in standard texts, additional sources of information are listed in the references.

Inspection instrumentation must possess four qualities in order to receive widespread acceptance in the NDI community:

1. *Accuracy*: The instrument must accurately measure a property of the material or structure that can be used to infer either its properties or the presence of flaws.
2. *Reliability*: The instrument must be highly reliable, that is, it must consistently detect and quantify flaws or a property with a high degree of reliability. If an instrument is not reliable, then it may not detect flaws that can lead to failure of the component, or it may indicate the presence of a flaw where none exists. The detection of a phantom flaw can mean that an adequate component is rejected, which is a costly error.
3. *Simplicity*: The most frequently used instruments are those used by factory or repair technicians. The inspection community rarely uses highly skilled operators due to the cost constraints.

4. *Low Cost*: An instrument need not be low cost in an absolute sense. Instead, it must be an inexpensive relative either to the value of the component under test or to the cost of a failure or aborted mission. For example, in the aircraft industry as much as 12% of the value of the component may be spent on inspection of a flight-critical aircraft component.

40.1.1 Information on Inspection Methods

To the engineer confronted by a new inspection requirement, there may arise the question of where to find pertinent information regarding an inspection procedure and its interpretation. Fortunately, many potential sources of information about instrumentation and techniques are available for NDI, and a brief examination of this literature is presented here. Many of these references were generated because of the demands of materials used in flight-critical aerospace structures. In this chapter, we will refer only to scientific and engineering books and journals that one would reasonably expect to find in a well-provisioned library. With the rise of the Internet, now there are many electronic sources of information available on the World Wide Web. These include library catalogs, societal home pages, online journals devoted to inspection, home pages of instrument manufacturers with online demonstrations of their capabilities and inspection services, inspection software, and online forums devoted to solving inspection problems. The references provide many such sources. However, with new electronic sources appearing daily, it is only a brief snapshot of those available at the beginning of the twenty-first century. For those new to the technology, American Society for Testing and Materials (ASTM) standards are particularly valuable because they give very detailed directions on many NDE techniques. More importantly, they are widely accepted standards for inspections. The references also provide sources for those situations where standard inspection methods are not sufficient to detect the material condition of interest.

40.1.1.1 General NDE Reference Books General overviews to NDE techniques are provided in Refs. (ASM International, 1989; Mitchell and Buck, 1992; Shapuk, 1997; Green, 1998; Altergott and Henneke, 1990; Halmshaw, 1991; Kline, 1992; Mallick, 1997; McGonnagle, 1961; Ruud, 1986; Sharpe, 1984; Summerscales, 1994; Thompson and Chimenti, 1982; Boogaard and van Dijk, 1989; Bray and Stanley, 1989; Geier, 1994) (*British Journal of Nondestructive Testing*, no longer published; American Society for Nondestructive Testing, <http://www.asnt.org>; Center for Nondestructive Evaluation, <http://www.cnde.iastate.edu>; Center for Quality Engineering & Failure Prevention, Center for Quality Engineering & Failure Prevention, <http://www.cqe.nwu.edu>; *Nondestructive Evaluation System Reliability Assessment*, <http://www.ihserc.com>; Nondestructive Testing Information Analysis Center, <http://www.ntiac.com>). The reader will note that some of these citations are not recent, but they are included because of their value to the engineer who does not possess formal training in the latest inspection technologies. Additionally, some older works were included because of their clarity of presentation, completeness, or usefulness to the inspection of complex structures.

40.1.1.2 NDE Journals The periodical literature is often a source of the latest research results for new or modified inspection methodologies (*British Journal of Nondestructive Testing*, no longer published; Online Journal Publication Service, <http://ojps.aip.org>; *Journal of Composite Materials*, 2004; electronic journals, <http://lib-www.lanl.gov/>

cgi-bin/ejrnlsrch.cgi; Elsevier Science, <http://www.elsevier.com/homepage/elecserv.htm>; *Japanese Journal of Nondestructive Inspection*, http://sparc5.kid.ee.cit.nihon-u.ac.jp/homepage_Eng.html; *Journal of Nondestructive Evaluation*, Kluwer Academic, Norwell, MA, 2004). Some excellent journals are no longer available but are still a valuable source of information or may contain data available nowhere else. Whenever possible, World Wide Web addresses are provided to give the reader ready access to this material.

40.1.2 Electronic References

There are many useful electronic references for those working in NDE technology. Only a few of the many useful sites on the World Wide Web are included here. Many sites contain links to other sites that contain information on a special topic of interest to the reader. Because the Web is constantly being updated, the list in the References represents a very brief snapshot of the information available to the NDE community. Some useful sites associated with government agencies were not included due to space limitations. The Web addresses provided are associated with NDE societies (*British Journal of Non-destructive Testing*, no longer published; American Society for Nondestructive Testing, <http://www.asnt.org>; *Japanese Journal of Nondestructive Inspection*, http://sparc5.kid.ee.cit.nihon-u.ac.jp/homepage_Eng.html; IFANT, International Foundation for the Advancement of Nondestructive Testing, <http://www.ifant.org>; Japan JSNDI, <http://sparc5.kid.ee.cit.nihon-u.ac.jp/homepageEng.html>; SPIE, <http://spie.org/>; Institute of Electrical and Electronic Engineers, <http://www.ieee.org/>; *IEEE-ASME, Journal of Microelectromechanical Systems*, Vol. 2000, 2004; American Society of Mechanical Engineers, <http://www.asme.org/>) (British Institute of Non-Destructive Testing, 1999) institutes, (Center for Nondestructive Evaluation, <http://www.cnde.iastate.edu>; Center for Quality Engineering & Failure Prevention, <http://www.cqe.nwu.edu>; Nondestructive Testing Information Analysis Center, <http://www.ntiac.com>; Center for Nondestructive Evaluation, <http://www.cnde.com>; Airport and Aircraft Safety Research & Development, <http://www.asp.tc.faa.gov>; Fraunhofer IZFP, <http://www.fhg.de/english/profile/institute/izfp/index.html>; Stasuk Testing & Inspection, <http://www.nde.net>), government agencies (Electronic journals, <http://lib-www.lanl.gov/cgi-bin/ejrnlsrch.cgi>; Airport and Aircraft Safety Research & Development, <http://www.asp.tc.faa.gov>; AFRL electronic journals, <http://www.wrs.afrl.af.mil/infores/library/ejournals.htm>), and general-interest sites (Elsevier Science, <http://www.elsevier.com/homepage/elecserv.htm>; SPIE, <http://spie.org/>; *IEEE-ASME, Journal of Microelectromechanical Systems*, Vol. 2000, 2004; Link, Springer Verlag, <http://link.springer-ny.com/>; IBM Intellectual Property Network, <http://www.patents.ibm.com>; Lavender International NDT, in *Lavender International*, 2004). There are also many references for the reader interested in using or modifying existing NDE techniques (IFANT, International Foundation for the Advancement of Nondestructive Testing, <http://www.ifant.org>; *Journal of Intelligent Material Systems and Structures*, <http://www.techpub.com>) (Krishnadas Nair, 1997; Rose and Tseng, 1988).

40.1.3 Future NDE Capabilities

At this point, the reader might be tempted to ask if there are new technologies on the horizon that will enable more cost-effective, anticipatory inspection or monitoring of materials and structures. The answer to this is an emphatic yes. There are new developments in solid-state detectors that should significantly affect both inspection capability

and cost. For example, optical and X-ray detectors now give the inspector the ability to rapidly scan large areas of structures for defects. Many new developments in these areas are the outgrowth of advances in noninvasive medical imaging. By coupling this technology with computer algorithms that search an image, the inspection of large areas can be automated, providing more accurate inspections with much less operator fatigue. Hopefully, this technological advance will remove much of the drudgery of detecting the rather small number of flaws in an otherwise large population of satisfactory components.

The area of data fusion is just beginning to be explored in the NDE field. This means that data collected with one technique can be combined with another technique to detect a range of flaws not detected when either is used independently. Data from several techniques can then be coupled at the basic physics level to provide a more complete description of the microstructural details of a material than is now possible.

Finally, the development of new semiconductor-based devices microelectromechanical systems (MEMS) and radio-frequency identification (RFID) allows the implantation of monitoring devices into a material at the time of manufacture to enable real-time structural health monitoring. These devices will permit the inspector to detect and quantify material or structural degradation remotely. This should also enable management of the components and structures for optimum usage over their lifetimes. Remote inspection and tracking of material degradation should reduce the burden of inspection while giving the inspector the ability to examine areas of structure that are now called “hidden.” For more information about this rapidly evolving area, the reader is referred to the literature (*Journal of Micromechanics and Microengineering*, 2004; SPIE, <http://spie.org/>; Smart Structures, <http://www.adaptive-ss.com/>; Smart Materials and Structures, <http://www.adaptive-ss.com/>, 2001; Smart Structures–Harvard, <http://iti.acns.nwu.edu/clear/infr/imatsmart.html>).

A brief review of the commonly used NDI methods are listed in Table 40.1 along with types of flaws that each method detects and the advantages and disadvantages of each technique. For detailed information regarding the capabilities of any particular method, the reader is referred to the literature. A good place to start any search for the latest NDE technology is the home page of the American Society for Nondestructive Testing (American Society for Nondestructive Testing, <http://www.asnt.org>).

40.2 LIQUID PENETRANTS

Liquid penetrants are used to detect surface-connected discontinuities, such as cracks, porosity, and laps, in solid, nonporous materials (Tracy, 1999). The method uses a brightly colored visible or fluorescent penetrating liquid that is applied to the surface of a cleaned part. During a specified “dwell time,” the liquid enters the discontinuity and is then removed from the surface of the part in a separate step. The penetrant is drawn from the flaw to the surface by a developer to provide an indication of surface-connected defects. This process is depicted schematically in Figures 40.1–40.4. A penetrant indication of a flaw in a turbine blade is shown in Figure 40.5.

40.2.1 Penetrant Process

Both technical societies and military specifications require a classification system for penetrants. Society documents (typically ASTM E165) (Tracy, 1999) categorize penetrants

TABLE 40.1 Capabilities of the Common NDI Methods

Method	Typical Flaws Detected	Typical Application	Advantages	Disadvantages
Radiography	Voids, porosity, inclusions, and cracks	Castings, forging, weldments, and structural assemblies	Detects internal flaws; useful on a wide variety of geometric shapes; portable; provides a permanent record	High cost; insensitive to thin laminar flaws, such as tight fatigue cracks and delaminations; potential health hazard
Liquid penetrants technique	Cracks, gouges, porosity, laps, and seams open to a surface	Castings, forging, weldments, and components subject to fatigue or stress-corrosion cracking	Inexpensive; easy to apply; portable; easily interpreted	Flaw must be open to an accessible surface, level of detectability operator dependent
Eddy current inspection	Cracks and variations in alloy composition or heat treatment, wall thickness, dimensions	Tubing, local regions of sheet metal, alloy sorting, and coating thickness measurement	Moderate cost; readily automated; portable	Detects flaws that change conductivity of metals; shallow penetration; geometry sensitive
Magnetic particles method	Cracks, laps, voids, porosity, and inclusions	Castings, forging, and extrusions	Simple; inexpensive; detects shallow subsurface flaws as well as surface flaws	Useful on ferromagnetic materials only; surface preparation required; irrelevant indications often occur; operator dependent
Thermal testing	Voids or disbands in both metallic and nonmetallic materials, location of hot or cold spots in thermally active assemblies	Laminated structures, honeycomb, and electronic circuit boards	Produces a thermal image that is easily interpreted	Difficult to control surface emissivity and poor discrimination between flaw types
Ultrasonic methods	Cracks, voids, porosity, inclusions and delaminations, and lack of bonding between dissimilar materials	Composites, forgings, castings, and weldments and pipes	Excellent depth penetration; good sensitivity and resolution; can provide permanent record	Requires acoustic coupling to component; slow; interpretation of data is often difficult

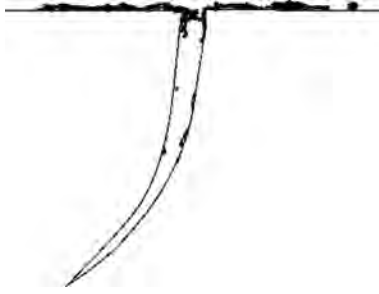


FIGURE 40.1 Representation of part surface before cleaning for penetrant inspection.

into visible and fluorescent, depending on the type of dye used. In each category, there are three types, depending on how the excess penetrant is removed from the part. These are water washable, postemulsifiable, and solvent removable.

The first step in penetrant testing (PT) or inspection is to clean the part. This critical step is one of the most neglected phases of the PT procedure. Since PT only detects flaws that are open to the surface, the flaw and part surface must be free of dirt, grease, oil, water, chemicals, and other foreign materials that might block the penetrant's entrance into a defect. Typical cleaning procedures use vapor degreasers, ultrasonic cleaners, alkaline cleaners, or solvents.

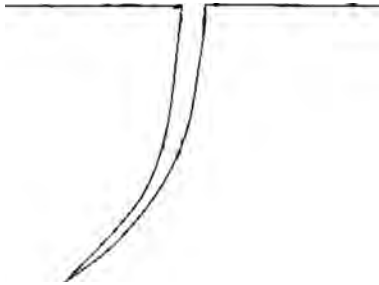


FIGURE 40.2 Part surface after cleaning and before penetrant application.

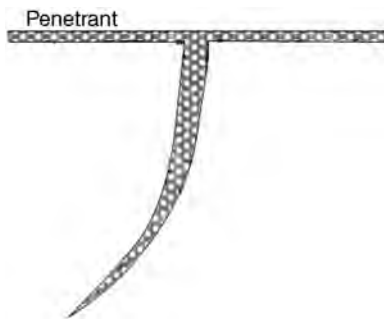


FIGURE 40.3 Part after penetrant application.

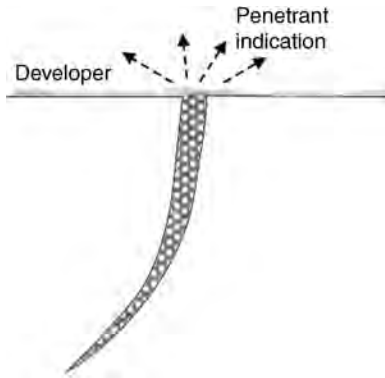


FIGURE 40.4 Representation of part after excess penetrant has been removed and developer has been applied.

After the surface is clean, a liquid penetrant is applied to the part by dipping, spraying, or brushing. In this step, the penetrant on the surface is wicked into the flaw. In the case of tight or narrow surface openings, such as fatigue cracks, the penetrant must be allowed to remain on the part for a minimum of 30 min to completely fill the flaw. High-sensitivity fluorescent dye penetrants are used for this type of inspection.

After the dwell time, excess penetrant is removed by one of the processes mentioned previously. For water-based penetrants, an emulsifier is sprayed onto the part and again a dwell time is observed. Water is then used to remove the penetrant from the surface of the part. In some cases, the emulsifier is included in the penetrant, so one only needs to wash the part after the penetrant has had time to penetrate the flaw. These penetrants are therefore called “water washable.” Of course, the emulsifier reduces the brightness of any flaw indication because it dilutes the penetrant. Ideally, only the surface penetrant is removed with the penetrant in the flaw left undisturbed.

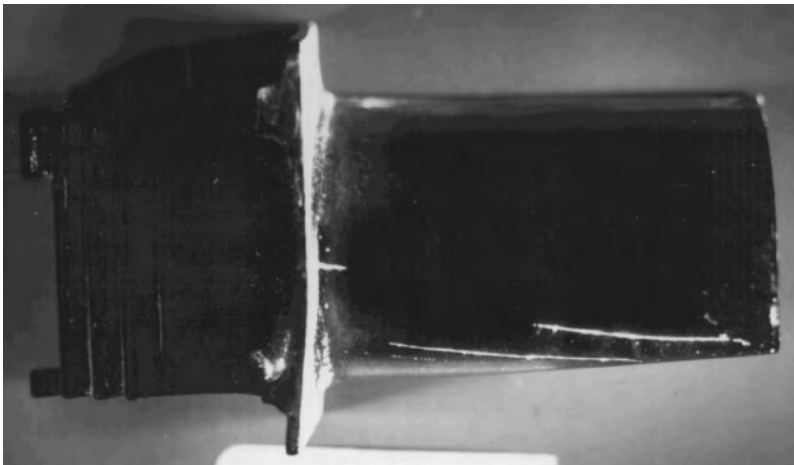


FIGURE 40.5 Penetrant indication of crack running along edge of jet engine turbine blade. UV illumination causes the extracted penetrant to fluoresce.

The final step in a basic penetrant inspection is the application of a fine powder developer. This may be applied either wet or dry. The developer aids in wicking the penetrant from the flaw and provides a suitable background for its detection. The part is then viewed under a suitable illumination—either an ultraviolet (UV) or a visible source. A typical fluorescent penetrant indication for a crack in a jet engine turbine blade is shown in Figure 40.5.

40.2.2 Reference Standards

Several reference standards are used to check the effectiveness of liquid penetrant systems. One of the oldest and most often used methods involves applying penetrant to hard chromium-plated brass panels. The panel is bent to place the chromium in tension, producing a series of cracks in the plating. These panels are available in sets containing fine, medium, and coarse cracks. The panels are used to classify penetrant materials by sensitivity and to detect degrading changes in the penetrant process.

40.2.3 Limitations of Penetrant Inspections

The major limitation of liquid penetrant inspection is that it can only detect flaws that are open to the surface. Other inspection methods must be used to detect subsurface defects. A factor that may inhibit the effectiveness of liquid penetrant inspection is surface roughness. Rough surfaces are likely to produce false indications by trapping penetrant, therefore, PT is not suited to the inspection of porous materials. Other penetrant-like methods are available for porous components—see the discussion of filtered particle inspection in Tracy (1999).

40.3 RADIOGRAPHY

In radiography used in NDE, the projected X-ray attenuations of a multitude of paths through a specimen are recorded as a two-dimensional image on a recording media, usually film. One might ask if the newer solid-state X-ray imaging technologies used in medicine also apply to NDE. The answer is yes, as will be discussed in the latter part of this section. The most used recording medium is still film because it is the simplest to apply and provides a resolution of subtle details not currently available with solid-state detectors. However, this situation may not be the case much longer as rapid progress is being made in the development of solid-state detectors with significantly enhanced resolution capabilities. Therefore, since this chapter is written at the beginning of the twenty-first century, when film usage for inspection is still commonplace, this portion of the chapter approaches radiography from the standpoint of film-based recording. Since most quantitative relationships for film also apply to solid-state detectors, the material presented should be applicable for the near future.

The radiography testing (RT) process is shown schematically in Figure 40.6. RT records any feature that changes the attenuation of the X-ray beam as it traverses the component. This local change in attenuation produces a change in the intensity of the X-ray beam, which translates into a change in the density, or darkness, on a film. This change in brightness may appear as a distinct shadow or in some cases a delicate shadow on the radiograph. The inspector is greatly aided in detecting a flaw or discrepancy in a part by

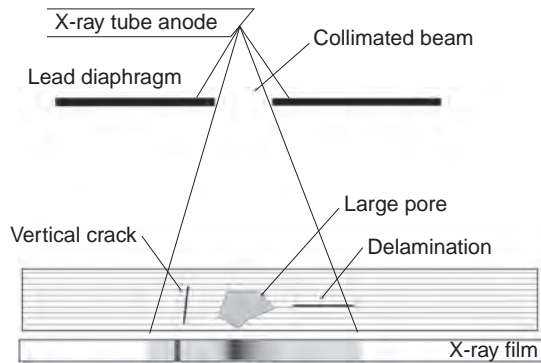


FIGURE 40.6 Schematic radiograph with typical flaws.

his or her knowledge of part shape and its influence on the radiographic image. Flaws, which do not change the attenuation of the X-ray beam on passage through the part, are not recorded. For example, a delamination in a laminated specimen is not visible because there is no local change in attenuation of the X-ray beam as it transverse the part. Conversely, flaws that are oriented parallel to the X-ray path do not attenuate the beam as much, allowing more radiation to expose the film and appearing darker than the surrounding image. An example of a crack in the correct orientation to be visible on a radiograph of a piece of tubing is shown in Figure 40.7.

40.3.1 Generation and Absorption of X Radiation

X radiation can be produced from a number of processes. The most common method of generating X rays is with an electron tube in which a beam of energetic electrons impacts a metal target. As the electrons are rapidly decelerated by this collision, a wide band of X radiation is produced, analogous to white light. This band of radiation is referred to as *Bremsstrahlung* or breaking radiation. These high-energy electrons produce short-wavelength energetic X rays. The relationship between the shortest wavelength radiation and the highest voltage applied to the tube is given by

$$\lambda = \frac{12,336}{\text{voltage}},$$

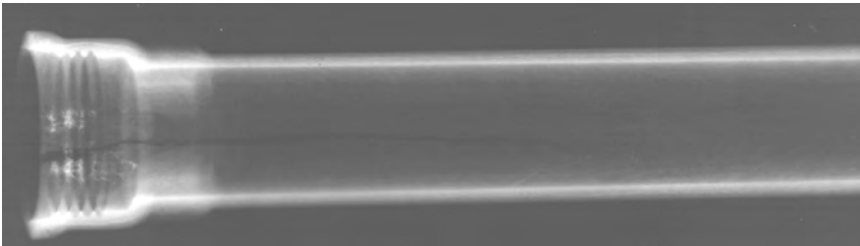


FIGURE 40.7 Radiograph of crack in end of aluminum tubing.

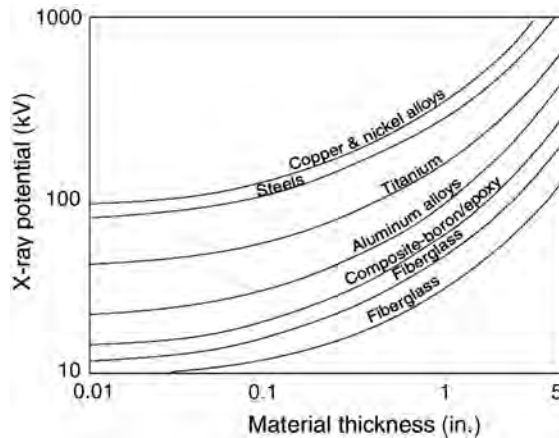


FIGURE 40.8 Plot of X-ray tube voltage versus thickness of several industrial materials.

where λ is the wavelength in angstroms and is the shortest wavelength of the X radiation produced. The more energetic the radiation, the more penetrating powers it possesses, and very high energy radiation is used on dense materials such as metals. While it is possible to analytically predict what X-ray energy would provide the best image for a specific material and geometry, a simpler method of determining the optimum X-ray energy is shown in Figure 40.8. Note that high-energy X-ray beams are used for dense materials, for example, steels, or for thick low-density materials, for example, large plastic parts. An alternate method to using this figure is to use the radiographic equivalence factors given in Table 40.2 (Quinn, 1980). Aluminum is the standard material for X-ray tube voltages below 100 keV, while steel is the standard above this voltage. When radiographing another material, its thickness is multiplied by the factor in this table to obtain the equivalent thickness of the standard material. The radiographic parameters are set up for this thickness of aluminum or steel. When used in this manner, good radiographs can be obtained for most parts. For example, assume that one must radiograph a 0.75-in.-thick piece of brass with a 400-keV X-ray source. The inspector should multiply the 0.75 in. of brass by the factor of 1.3 to obtain 0.98. This means that an acceptable radiograph of the brass plates would be obtained with the same exposure parameters as would be used for 0.98 in. (~ 1 in.) of steel.

Radiation for RT can also be obtained from the decay of radioactive sources. In this case, the process is usually referred to as gamma radiography. These radiation sources have several characteristics that differ from X-ray tubes. First, gamma radiation is very nearly monochromatic; that is, the spectrum of radiation contains only one or two dominant energies. Second, the energies of most sources are on the order of millions-of-volts range, making this source ideal for inspecting highly attenuating materials or very large structures. Third, the small size of these sources permits them to be used in tight locations where an X-ray tube could not fit. Fourth, since the gamma-ray source is continually decaying, adjustments to the exposure time must be made to achieve consistent results over time. Finally, the operator must always remember that the source is continually on and is therefore a persistent safety hazard! Aside from these differences, gamma radiography differs little from standard practice, so no further distinction between the two will be given.

TABLE 40.2 Approximate Radiographic Equivalence Factors

Energy Level	150kV	150kV	250 kV	400kV	400 kV	1 MeV	2 MeV	4–25 MeV	¹⁹² Ir	⁶⁰ Co
Metal										
Magnesium	0.05	0.05	0.08						0.35	0.35
Aluminum	0.08	0.12	0.18						0.35	0.35
Aluminum alloy	0.10	0.14	0.18						0.9	0.9
Titanium	—	0.54	0.54			0.9	0.9	0.9	1.0	1.0
Iron/all steels	1.0	1.0	1.0	1.0		1.0	1.0	1.0	1.1	1.1
Copper	1.5	1.6	1.4	1.4		1.1	1.1	1.2	1.1	1.0
Zinc	—	1.4	1.3	—		—	—	1.2	1.1	1.0
Brass	—	1.4	1.3	—		1.2	1.1	1.0	1.1	1.0
Inconel X	—	1.4	1.3	—		1.3	1.3	1.3	1.3	1.3
Monel	1.7	—	1.2	—						
Zirconium	2.4	2.3	2.0	1.7	1.5	1.0	1.0	1.0	1.2	1.0
Lead	14.0	14.0	12.0	—	—	5.0	2.5	2.7	4.0	2.3
Halfnium			14.0	12.0	9.0	3.0	—	—		
Uranium			20.0	16.0	12.0	4.0	—	3.9	12.6	3.4

Source: From Quinn (1980).

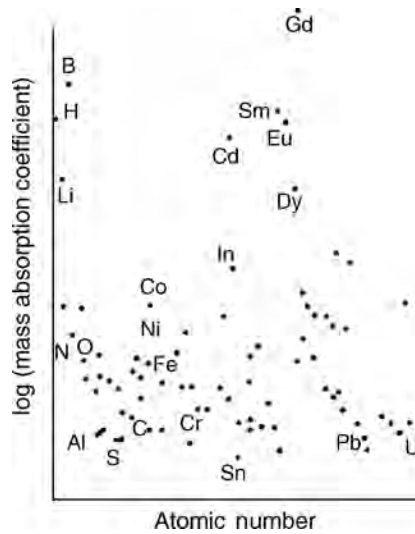


FIGURE 40.9 Plot of mass absorption coefficient for neutron radiography versus atomic number.

40.3.2 Neutron Radiography

Neutron radiography (Burger, 1965) may be useful to inspect some materials and structures. Because the attenuation of neutrons is not related to the elemental composition of the part, some elements can be more easily detected than others. While X rays are most heavily absorbed by high-atomic-number elements, this is not true of neutrons, as shown in Figure 40.9. In Figure 40.10 two aluminum panels are bonded with an epoxy adhesive. The reader can discern that hydrogen adsorbs neutrons more than aluminum does, and thus the missing adhesive is easily detectable.

Neutron radiography, however, does have several constraints. First, neutrons do not expose radiographic film and therefore a fluorescing medium is often used to produce light, which exposes the film. The image produced in this manner is not as sharp and well defined as that from X rays. Second, at present, there is no portable high-flux source of neutrons. This means that a nuclear reactor is most often used to

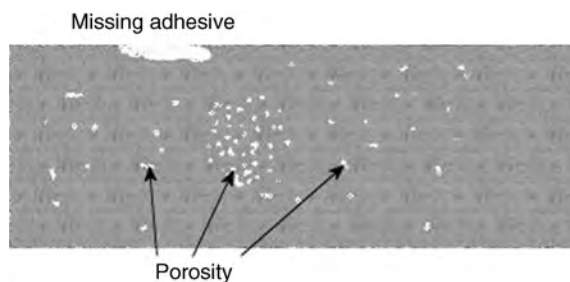


FIGURE 40.10 Representation of neutron radiograph showing flaws in adhesive bond.

supply the neutron radiation. Although neutron radiography has these severe restrictions, at times there is no alternate, and the utility of this method outweighs its expense and complexity.

40.3.3 Attenuation of X Radiation

An appreciation of how radiographs are interpreted requires a fundamental understanding of X-ray absorption. The relationship governing this phenomenon is de Beer's law:

$$I = I_0 e^{-\mu x},$$

where I , I_0 = transmitted and incident X-ray beam intensities, respectively; μ = attenuation coefficient of material, cm^{-1} ; x = thickness of specimen, cm.

Since the attenuation coefficient is a function of both the composition of the specimen and the wavelength of the X rays, it would be necessary to calculate or measure it for each wavelength used in RT. However, it is possible to calculate the attenuation coefficient of a material for a specific X-ray energy using the mass absorption coefficient μ_m as defined below. The mass absorption coefficients for most elements are readily available for a variety of X-ray energies (Bossi et al., 2002),

$$\mu_m = \frac{\mu}{\rho},$$

where μ = attenuation coefficient of an element, cm^{-1} ; ρ = density, g/cm^3 .

The mass absorption coefficient for the material is obtained, at a specific X-ray energy, by multiplying the μ_m of each element by its weight fraction in a material and summing these quantities. Multiplying this sum by the density of the material yields its attenuation coefficient for the material. This procedure is often not used in practice because the results are valid only for a narrow band of wavelengths. Radiographic equivalency factors are used instead. This process points out that each element in a material contributes to the attenuation coefficient by an amount proportional to its amount in the material.

40.3.4 Film-Based Radiography

The classical method of recording an X-ray image is with film. Because of the continued importance of this medium of recording and the fact that much of the technology associated with it is applicable to newer solid-state recording methods, this section explores film radiography in some detail.

The relationship between the darkness produced on an X-ray film and the quantity of radiation impinging on it is shown by log-log plots of darkness, or film density, and relative exposure (Figures 40.11 and 40.12). Varying the time of exposure, intensity of the beam, or specimen thickness changes the density, or darkness, of the image. The slope of the curve along its linear portion is referred to as the film gamma, γ . Film has characteristics that are analogous to electronic devices: The greater the gamma or amplification capability of the film, the smaller its dynamic range—the range of exposures over which density is linearly related to thickness. If it is necessary to use a high-gamma film to detect very subtle flaws in a part with a wide range of thicknesses, then it is necessary to

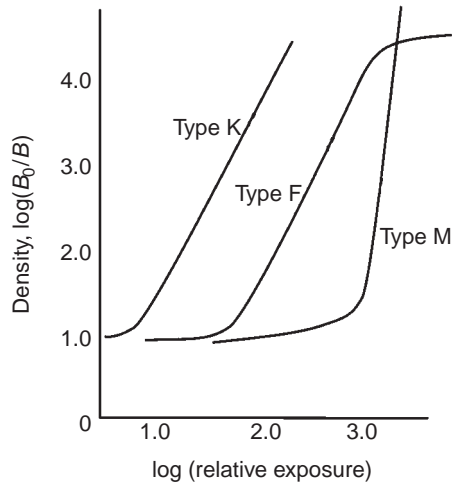


FIGURE 40.11 Density or darkness of X-ray film versus relative exposure for three common films.

use several different film types in the same cassette or package. In this way, each film will be optimized for flaw detection in a narrow thickness range of the part.

Using this information, one may calculate the minimum detectable flaw size for a specific RT inspection. A simple method is available to check the radiographic procedure to determine if this detectability has been achieved on the film. This method does not ensure that the radiograph was taken with the specimen in the proper orientation; it merely provides a method of checking for proper execution of a radiographic procedure (see Section 40.3.5).

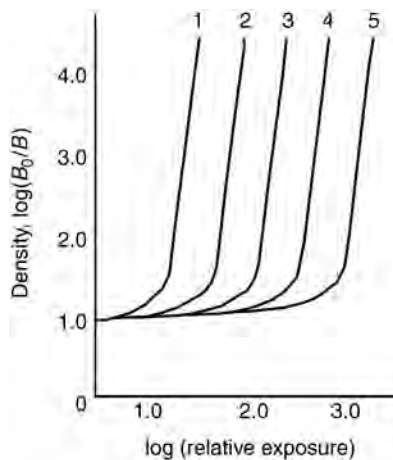


FIGURE 40.12 Density versus relative exposure for films that could be used in multiple film exposure to obtain optimum flaw detectability in complex part.

TABLE 40.3 Radiographic Sensitivity with Thinnest Penetrameter and Smallest Hole Visible on Radiograph

Sensitivity, S (%)	Quality Level (% T – Hole Diameter)
0.7	1 – 1 T
1.0	1 – 2 T
1.4	2 – 1 T
2.0	2 – 2 T
2.8	2 – 4 T
4.0	4 – 2 T

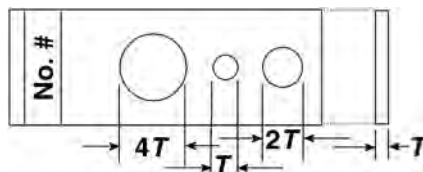
Using a knowledge of the minimum density difference that is detectable by the average radiographic inspector, the following equation relates the radiographic sensitivity, S , to radiographic parameters:

$$S = \frac{2.3}{\gamma \mu x},$$

where S is the radiographic sensitivity in percent, γ is the film gamma, μ is the attenuation coefficient of the specimen material, and x is the maximum thickness of the part associated with a particular radiographic film. The radiographer uses a penetrameter to determine if this sensitivity was achieved. Table 40.3 give the sensitivity S in percent and the expected RT performance in penetrameter values (see Section 40.3.5).

40.3.5 Penetrameter

An example of a penetrameter is shown schematically in Figure 40.13, while its image on a radiograph is shown in Figure 40.14. While there are many types of penetrameters, this one was chosen because it is easily related to radiographic sensitivity. The penetrameter is simply a thin strip of metal or polymeric material (Fassbender and Hagemaiier, 1983) in which three holes of varying sizes are drilled or punched. It is composed of the same material as the specimen and has a thickness 1, 2, or 4% of maximum part thickness. The holes in the penetrameter have diameters that are 1 T , 2 T , and 4 T . The sensitivity achieved for each radiographic is determined by noting the smallest hole just visible in the thinnest penetrameter on a film and using Table 40.3 to determine the sensitivity achieved. By calculating the radiographic sensitivity and then noting the level achieved in practice, the radiographic process can be quantitatively evaluated. While this procedure does not offer any guarantee of flaw detection, it is useful in evaluating the effectiveness of the RT process.

**FIGURE 40.13** Schematic of typical film penetrameter.

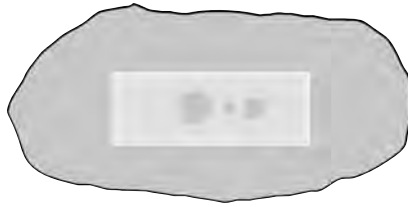


FIGURE 40.14 Radiograph of penetrameter shown in Figure 40.13. The 1T hole is just visible, indicating the resolution obtained in the radiograph.

Almost all variables of the radiographic process may be easily and rapidly changed with the aid of tables, graphs, and nomograms, which are usually provided by film manufacturers free of charge. For more information, the reader is referred to the commercial literature.

40.3.6 Real-Time Radiography

While film radiography represents the bulk of radiographic NDE performed at this time (the beginning of the twenty-first century), new methods of both recording the data and analyzing it are coming into widespread usage. For example, filmless radiography (FR) and real-time radiography (RTR) use solid-state detectors and digital signal processing (DSP) software instead of film to record and enhance the radiographic image. These methods have many advantages, along with some disadvantages. For example, FR permits viewing a radiographic image while the specimen is being moved. This often permits the detection of flaws that would normally be missed in conventional film radiography because of the limited number of views or exposures taken—remember that the X-ray beam must pass along a crack or void for it to be detectable. Additionally, the motion of some flaws enhances their detectability because they present the inspector with a different image as a function of time. Additionally, image enhancement techniques can now be economically and rapidly applied to these images because of the availability of inexpensive, fast computing hardware. The disadvantage of RTR is its lower resolution compared to film. Typical resolution capabilities of RTR or FR systems are in the range of 4 to perhaps 20 line pairs/mm, whereas film resolution capabilities are in the range of 10–100 line pairs/mm. This means some very fine flaws may not be detectable with FR and the inspector must resort to film. However, in cases where resolution is not the limiting factor, the benefits of software image enhancement can be significant. While the images on film may also be enhanced using the image processing schemes, they cannot be performed in real or near real time, as can be done with an electronic system.

40.3.7 Computed Tomography

Another advance in industrial radiography has been the incorporation of computed tomography (CT) into the repertoire of the radiographer. Unfortunately, CT has not been exploited to its fullest extent principally due to the high cost of instrumentation. The capability of CT to link NDE measurements with engineering design and analysis gives this inspection a unique ability to provide quantitative estimates of performance not associated with NDE.

The principal advantage of this method is that it produces an image of a thin slice of the specimen under examination. This slice is parallel to the path of the X-ray beam that passes through the specimen, in contrast to the shadowgraph image produced by traditional radiography shown in Figure 40.7. Whereas the shadowgraph image can be difficult to interpret, the computed CT image does not contain information from planes outside the thin slice.

A comparison between CT and traditional film radiography is best made with images from these two modalities. Figure 40.7 shows a typical radiograph where one can easily see the image of the top and bottom surfaces of the tube under inspection. The reader can contrast this with the image in Figure 40.15, a CT image of a flashlight. The individual components of the flashlight are easily visible and any misplacement of its components or defects in its assembly can be easily detected. An image with a finer scale that reveals the microstructural details of a pencil is shown in Figure 40.16. Clearly visible are not only the key features and even the growth rings of the wood. In fact, the details of the growth during each season are visible as rings within rings. The information in the CT image contrasted with conventional radiographs is striking. First, the detectability of a defect is independent of its position in the image. This is not the case with the classical radiograph, where the defect detectability decreases significantly with depth in the specimen because the defect represents a smaller change in the attenuation of the X-ray beam as the depth increases. Second, the defect detectability is very nearly independent of its orientation. This again is clearly not the case with classical radiography. New applications for CT are constantly being discovered. For example, with a digital CT image, it is possible to search for various flaw conditions using computer analysis and relieve the inspector of much of the tedium of examining structures for the odd flaw. In addition, it is possible to link the digital CT image with finite-element analysis software to examine precisely how the flaws present will affect such parameters as stress distribution, heat flow, and the like. With little effort, one could analyze the full three-dimensional performance of many engineering structures.

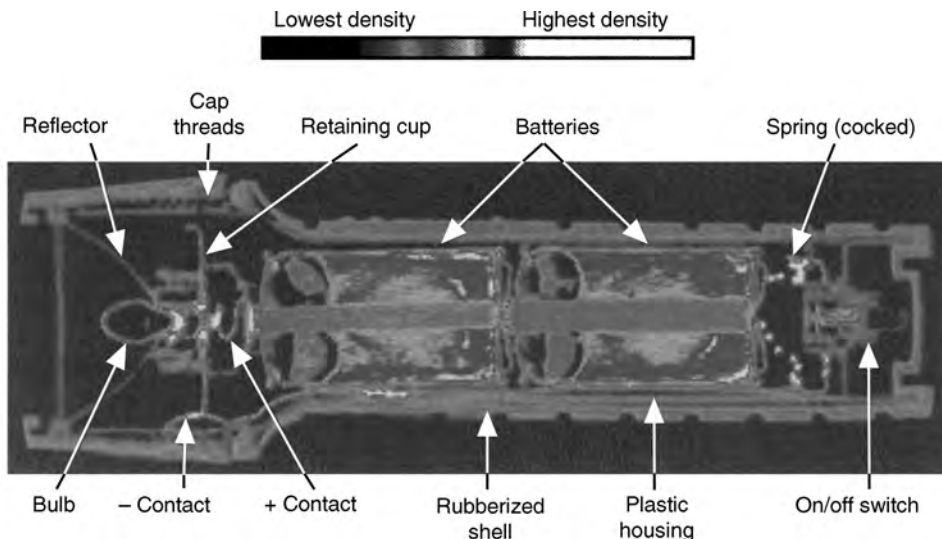


FIGURE 40.15 Computed tomography image of flashlight showing details of internal structure.

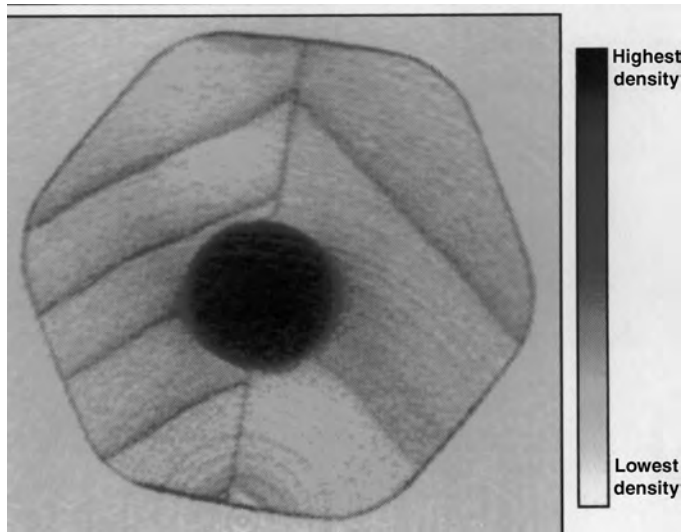


FIGURE 40.16 Computed tomography of pencil. The reader will note the yearly growth rings and even the growth variations within a single growing season.

40.4 ULTRASONIC METHODS

Ultrasonic inspection methods utilize high-frequency sound waves to inspect the interior of solid parts. Sound waves are mechanical or elastic disturbances or waves that propagate in fluid and solid media. Ultrasonic testing (UT) or inspection is similar to the angler who uses sonar to detect fish (Birks and Green, 1991). The government and various technical societies have developed standard practice specifications for UT. These include ASTM specifications 214-68, 428-71, and 494-75 and military specification MIL-1-8950H. Acoustic and ultrasonic testing can take many forms, from simple coin tapping to the transmission and reception of very high frequency or ultrasonic waves into a part to analyze its internal structure.

UT instruments operating in the frequency range between 20 and 500 kHz are referred to as *sonic* instruments, while those that operate above 500 kHz are called *ultrasonic*. To generate and receive ultrasonic waves, a piezoelectric transducer is employed to convert electrical signals to sound waves and back again. The usual form of a transducer is a piezoelectric crystal mounted in a waterproof housing that is electrically connected to a pulsar (transmitter) and a receiver. In the transmit mode a high-voltage, short-duration electrical spike is applied to the crystal, causing it to rapidly change shape and emit an acoustic pulse. In the receive mode, sound waves or returning echoes compress the piezoelectric crystal, producing an electrical signal that is amplified and processed by the receiver. This process is shown schematically in Figure 40.17.

40.4.1 Sound Waves

Ultrasonic waves have physical characteristics such as wavelength (λ), frequency (f), velocity (v), pressure (p), and amplitude (a). The following relationship between

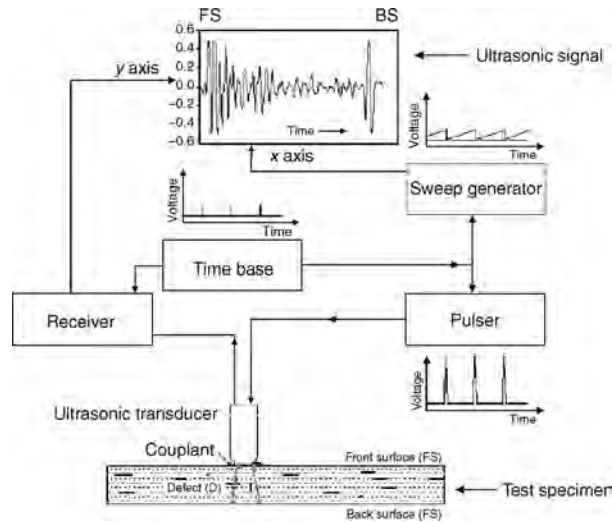


FIGURE 40.17 Schematic of ultrasonic data collection and display in A-scan mode.

wavelength, frequency, and sound velocity is valid for all sound waves:

$$f\lambda = v.$$

For example, the wavelength of longitudinal ultrasonic waves of frequency 2 MHz propagating in steel is 3 mm and the wavelength of shear waves is about half this value, 1.6 mm. The relation between the sound pressure and the particle amplitude is

$$p = 2\pi f \rho v a,$$

where p is density, f is the frequency of the sound wave, v is its velocity, and a the amplitude.

Ultrasonic waves are reflected from boundaries between different materials or media. Each medium has characteristic acoustic impedance and reflections occur in a manner similar to those observed with electrical signals. The acoustic impedance Z of any media capable of supporting sound waves is defined by

$$Z = \rho v$$

where ρ = density of medium, g/cm^3 ; v = velocity of sound along direction of propagation.

Materials with high acoustic impedance are often referred to as sonically hard, in contrast to sonically soft materials with low impedances. For example, steel ($Z = 7.7 \text{ g/cm}^3 \times 5.9 \text{ km/s} = 45.4 \times 10^6 \text{ kg/m}^2\cdot\text{s}$) is sonically harder than aluminum ($Z = 2.7 \text{ g/cm}^3 \times 6.3 \text{ km/s} = 17 \times 10^6 \text{ kg/m}^2\cdot\text{s}$). The Appendix at the end of this chapter lists the acoustic properties of many common materials.

40.4.2 Reflection and Transmission of Sound

Almost all acoustic energy incident on air–solid interfaces is reflected because of the large impedance mismatch between air and most solids. For this reason, a medium with impedance close to that of the part is used to couple the sonic energy from the transducer into the part. A liquid couplant has obvious advantages for parts with a complex geometry, and water is the couplant of choice for most inspection situations. The receiver, in addition to amplifying the returning echoes, also time gates the returning echoes between the front surface and rear surfaces of the component. Thus, any unusually occurring echo is displayed separately or used to set off an alarm, as shown in Figure 40.17. This method of displaying the voltage amplitude of the returning pulse versus time or depth (if acoustic velocity is known) at a single point in the specimen is known as an A scan. In this figure, the first signal corresponds to a reflection from the front surface (FS) of the part and the last signal corresponds to the reflection from its back surface (BS). The signal or echo between the FS and BS is from the defect in the middle of the part.

The portion of sound energy that is either reflected from or transmitted through each interface is a function of the impedances of the medium on each side of that interface. The reflection coefficient R (ratio of the sound pressures or intensities of the reflected and incident waves) and the power reflection coefficient R_{pwr} (ratio of the power in the reflected and incident sound waves) for normally incident waves onto an interface are given as

$$R = \frac{p_r}{p_i} = \frac{Z_1 - Z_2}{Z_1 + Z_2} \quad R_{\text{pwr}} = \frac{I_r}{I_i} = \left(\frac{Z_1 - Z_2}{Z_1 + Z_2} \right)^2.$$

Likewise, the transmission coefficients T and T_{pwr} defined as

$$T = \frac{p_t}{p_i} = \frac{2Z_2}{Z_1 + Z_2} \quad T_{\text{pwr}} = \frac{I_t}{I_i} = \frac{4(Z_2/Z_1)}{[1 + (Z_2/Z_1)]^2},$$

where I_i , I_r , and I_t are the incident, reflected, and transmitted acoustic field intensities, respectively; Z_1 is the acoustic impedance of the medium from which the sound wave is incident; and Z_2 is impedance of the medium into which the wave is transmitted. From these equations, one can calculate the reflection and transmission coefficients for a planar flaw containing air, $Z_1 = 450 \text{ kg/cm}^2\cdot\text{s}$, located in a steel part, $Z_2 = 45.4 \times 10^6 \text{ kg/m}^2\cdot\text{s}$. In this case, the reflection coefficient for the flaw is virtually -1.0 . The minus sign indicates a phase change of 180° for the reflected pulse (note that the defect signal in Figure 40.17 is inverted or phase shifted by 180° from the FS signal). Effectively, no acoustic energy is transmitted across an air gap, necessitating the use of water as a coupling media in ultrasonic testing. Using the acoustic properties of common materials given in the Appendix the reader can make a number of simple, yet informative, calculations.

Thus far, our discussion has involved only longitudinal waves. This is the only wave that travels through fluids such as air and water. The particle motion in this wave, if one could see it, is similar to the motion of a spring, or a Slinky toy, where the displacement and wave motion are collinear (the oscillations occur along the direction of propagation). The wave is called compressional or dilatational as both compressional and dilatational forces are active in it. Audible sound waves are compressional waves. This wave propagates in liquids and gases as well as in solids. However, a solid medium can also support additional types of waves such as shear and Rayleigh or surface waves. Shear or

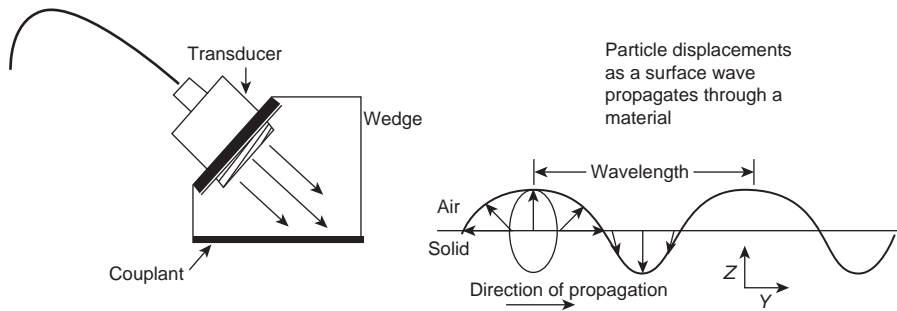


FIGURE 40.18 Generation and propagation of surface waves in a material.

transverse waves have a particle motion that is analogous to what one sees in an oscillating rope. That is, the displacement of the rope is perpendicular to the direction of wave propagation. The velocity of this wave is about half that of compressional waves and is only found in solid media, as indicated in the Appendix. Shear waves are often generated when a longitudinal wave is incident on a fluid-solid interface at angles of incidence other than 90° . Rayleigh or surface waves have elliptical wave motion, as shown in Figure 40.18, and penetrate the surface for about one wavelength; therefore, they can be used to detect surface and very near surface flaws. The velocity of Rayleigh waves is about 90% of the shear wave velocity. Their generation requires a special device, or wedge as shown in Figure 40.18, which enables an incident ultrasonic wave on the sample at a specific angle that is characteristic of the material (Rayleigh angle). The reader can find more details in the scientific literature (Ash and Paige, 1985; Viktorov, 1967; Krautkramer and Krautkramer, 1983).

40.4.3 Refraction of Sound

The direction of propagation of acoustic waves is governed by the acoustic equivalent of Snell's law. Referring to Figure 40.19, the direction of propagation is determined with the equation

$$\frac{\sin\theta_i}{c_I} = \frac{\sin\theta_r}{c_I} = \frac{\sin\gamma_r}{b_I} = \frac{\sin\theta_t}{c_{II}} = \frac{\sin\gamma_t}{b_{II}},$$

where c_I is the velocity of the incident longitudinal wave, c_I and b_I are the velocities of the longitudinal and shear reflected waves, and c_{II} and b_{II} are the velocities of the longitudinal and shear transmitted waves in solid II. In the water-steel interface, there is no reflected shear wave because these waves do not propagate in fluids such as water. In this case, the above relationship is simplified. Since the water has a lower longitudinal wave speed than either the longitudinal or shear wave speeds of the steel, the transmitted acoustic waves are refracted away from the normal. If the incident wave approaches the interface at increasing angles, there will be an angle above which there will be no transmitted acoustic wave in the higher wave speed material. This angle is referred to as a critical angle. At this angle, the refracted wave travels along the interface and does not enter the solid. A computer-generated curve is shown in Figure 40.20 in which the normalized acoustic energy that is reflected and refracted at a water-steel interface is plotted as a function of

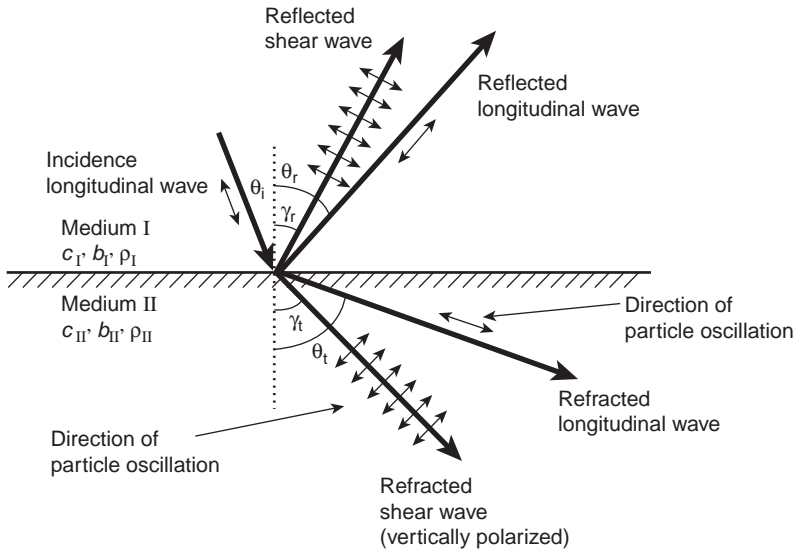


FIGURE 40.19 Representation of Snell's law and mode conversion of longitudinal wave incident on solid–solid interface.

the angle of the incident longitudinal wave. Note that a longitudinal or first critical angle for steel occurs at 14.5° . Likewise, the shear or second critical angle occurs at about 30° . If the angle of incidence is increased above the first critical angle but less than the second critical angle, only the shear wave is generated in the metal and travels at an angle of refraction described by Snell's law. Angles of incidence above the second critical angle produce a complete reflection of the incident acoustic wave; that is, no acoustic energy enters the solid. At a specific angle of incidence (Rayleigh angle) surface, acoustic waves are generating on the material. The Rayleigh angle can be easily calculated from Snell's law by assuming that the refracted angle is 90° . The Rayleigh angle for steel occurs at

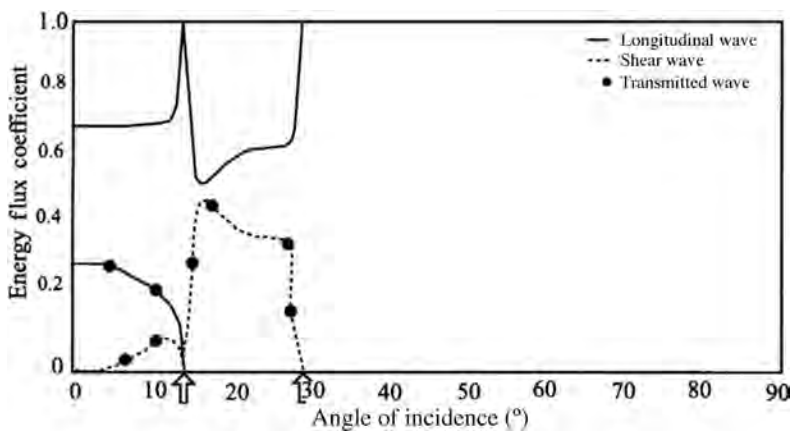


FIGURE 40.20 Amplitude (energy flux) and phase of reflected coefficient and transmitted amplitude versus angle of incidence for longitudinal wave incident on water–steel interface. The arrows indicate the critical angles for the interface.

29.5°. In the region between the two critical angles, only the shear wave is generated and is referred to as *shear wave testing*. There are two distinct advantages to inspecting parts with this type of shear wave. First, with only one type of wave present, the ambiguity that would exist concerning which type of wave is reflected from a defect does not occur. Second, the lower wave speed of the shear wave means that it is easier to resolve distances within the part. For these reasons, shear wave inspection is often chosen for inspection of thin metallic structures such as those in aircraft.

Using the relationships for the reflection and transmission coefficients, a great deal of information can be deduced about any ultrasonic inspection situation when the acoustic wave is incident at 90° to the surface. For other angles of incidence, computer software is often used to analyze the acoustic interactions. Analytic predictions of ultrasonic performance in complex materials such as fiber-reinforced composites require the use of more complex algorithms because more complicated modes of wave propagation can occur. Examples of these include Lamb waves (plate waves), Stoneley waves (interface waves), Love waves (guided in layers of a solid material coated onto another one), and others.

40.4.4 Inspection Process

Once the type of ultrasonic inspection has been chosen and the optimum experimental parameters determined, one must choose the mode of presentation of the data. If the principal dimension of the flaw is less than the diameter of the transducer, then the A-scan method may be chosen, as shown in Figure 40.17. The acquisition of a series of A-scans obtained by scanning the transducer in one direction across the specimen and displaying the data as distance versus depth is referred to as a B scan. This is the mode most often used by medical ultrasound instrumentation. In the A-scan mode, the size of the flaw may be inferred by comparing the amplitude of the defect signal to a set of standard calibration blocks. Each block has a flat bottom hole (FBH) drilled from one end. Calibration blocks have FBH diameters that vary in 1/64-in. increments, for example, a number 5 block has a 5/64-in. FBH. By comparing the amplitude of the signal from a calibration block with one from a defect, the inspector may specify a defect size as equivalent to a certain size FBH. The equivalent size is meaningful only for smooth flaws that are nearly perpendicular to the path of the ultrasonic beam and is used in many industrial situations where a reference size is required by a UT procedure.

If the flaw size is larger than the transducer diameter, then the C-scan mode is usually selected. In this mode, shown in Figure 40.21, the transducer is rastered back and forth across the part. In normal operation, a line is traced on a computer monitor or a piece of paper. When a flaw signal is detected between the front and back surfaces, the line drawing ceases and a blank place appears on the paper or monitor. Using this mode of presentation, a planar projection of each flaw is presented to the viewer and its positional relationship to other flaws and to the component boundaries is easily ascertained. Unfortunately, the C-scan mode does not show depth information, unless an electronic gate is set to capture only information from within a specified time window or time gate in the part. With current computer capability, it is a rather simple matter to store all of the returning A-scan data and display only the data in a C-scan mode for a specific depth.

Depending on the structural complexity and the attenuation of the signal, cracklike flaws as small as 0.015 in. in diameter may be reliably detected and quantified with this method. An example of a typical C-scan printout of an adhesively bonded test panel is shown in Figure 40.22. This panel was fabricated with a void-simulating Teflon implant

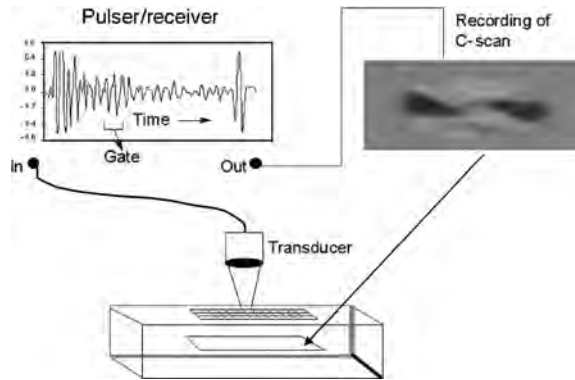


FIGURE 40.21 Representation of ultrasonic data collection. The data are displayed using the C-scan mode. The image shows a defect located at a certain depth in the material.

and the numerous additional white areas indicate the presence of a great deal of porosity in the part.

40.4.4.1 Through Transmission Versus Pulse–Echo Thus far, the discussion of ultrasonic inspection methods has been concerned with the setup that uses a single transducer to send a signal into the part and to receive any returning echoes. This method is variously referred to as pulse–echo or pitch–catch inspection and is shown schematically in Figure 40.23. The other frequently used inspection setup for many structures is called through transmission. With this setup, two transducers are used, one to send ultrasonic pulses and the other placed on the opposite side of the part to receive the transmitted signals, as shown schematically in Figure 40.24. In Figures 40.23 and 40.24 a large number of reflections occur for the many individual layers in a composite part that can obscure subtle reflections from inclusions whose reflectivity is similar to that of the layered materials. An inclusion with an acoustic impedance very close to that of the part, for example, paper or peel-plys in polymer-based composites, is very difficult to detect with the through–transmission mode of inspection. In this case, the pulse–echo inspection mode is often used to detect these flaws. On the other hand, reflections from distributed flaws such as porosity, as shown on the right-hand side of Figures 40.23 and 40.24, can be obscured by the general background noise present in an acoustic signal. Therefore, it is the loss in signal strength

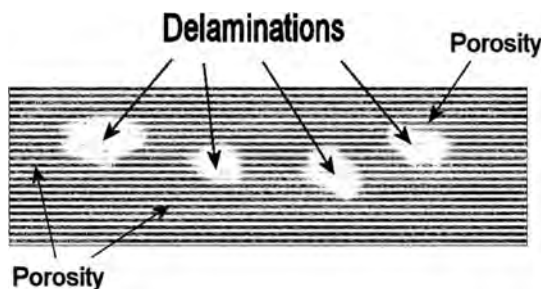


FIGURE 40.22 Typical C-scan image of composite specimen showing delaminations and porosity.

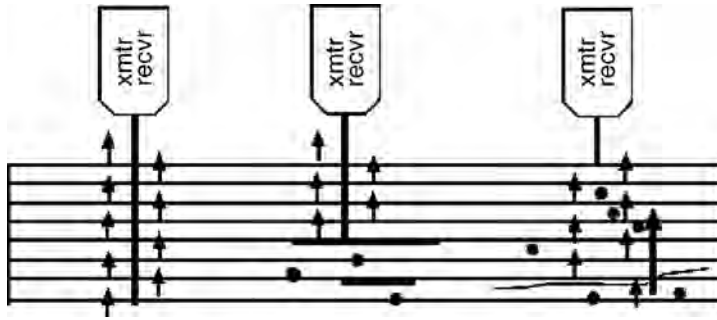


FIGURE 40.23 Representation of pulse-echo mode of ultrasonic inspection.

of the transmitted signal of the through-transmission method that is most often used to detect this type of flaw. While porosity is detectable in this manner, its location may not be determined. In this situation, the pulse-echo mode is required because the distance from the front or back surface to the flaw can be determined by the relative position of the reflections of the scattered porosity with respect to the surface reflection. Because each method supplies important information about potential flaws and its location, modern ultrasonic instrumentation is frequently equipped to perform both types of inspection nearly simultaneously (Jones and Stone, 1976; Jones, 1985). In such a setup, two transducers are used to conduct a through-transmission test and then each is used separately to conduct pulse-echo tests from opposite sides of the part. This method also helps ensure that a large flaw does not shadow a smaller one, as shown in Figures 40.23 and 40.24.

40.4.4.2 Portable Ultrasonic Systems This ability to image defects on specific levels within a layered component, for example, a composite, is so important that C-scan instrumentation has been miniaturized for usage in the field. An example of one such system developed for aircraft inspection is shown in Figure 40.25. The heart of this system is a computer that records the position of a hand-held transducer and the complete A-scan wave train at each point of the scan. Since the equipment tracks the motion of the

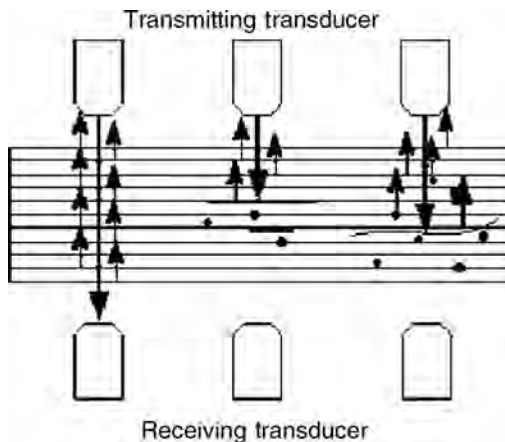


FIGURE 40.24 Representation of through-transmission mode of inspection.

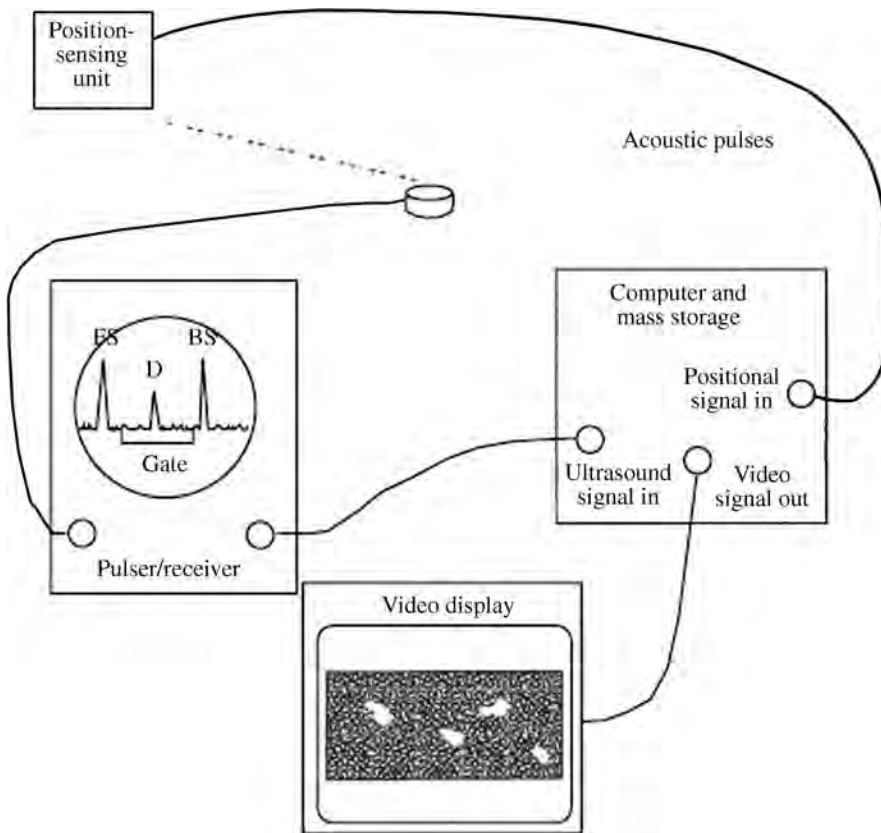


FIGURE 40.25 Field level C-scan instrumentation that is capable of simultaneously tracking the motion of a handheld transducer and recording the ultrasonic information.

transducer as it is scanned manually across a structure, the inspector can see which areas have been scanned. If areas are missed, he or she can return to “color them in,” as shown in the image on the monitor. Additionally, computer manipulation of ultrasonic image data allows the inspector to select either one or a small number of layers for evaluation. In this way, an orderly assessment of the flaws in critical structures can be accomplished. This process of selecting flaws on a layer-by-layer basis for evaluation is shown schematically in Figure 40.26 for the instrument depicted in Figure 40.25.

40.4.5 Bond Testers

A great deal of the ultrasonic inspection literature is devoted to instruments that test adhesive bonds. There has been a recent resurgence of interest in the inspection of adhesive bonds due to concerns about the viability of bonded patches on our aging aircraft (Hsu and Patton, 1993). For an extensive treatment of most of the currently used instruments, the reader is referred to review articles (Swamy and Ali, 1984; Papadakis and Chapman, 1993; Hagemier, 1971, 1972a, 1972b). However, while there may seem to be a large number of instruments, some with exaggerated claims of performance, most operate on the same physical principles.

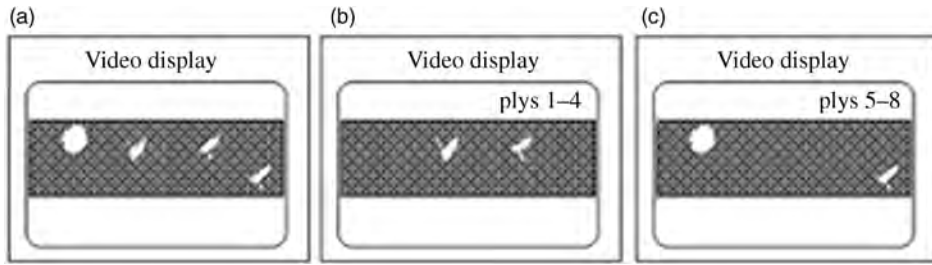


FIGURE 40.26 Three different displays of delaminations in 16-ply composite obtained from field-level C-scan system. (a) Projection of all flaws through specimen. (b, c) Images from selected depths within specimen.

Bond-testing instruments use a variety of means to excite sonic or low-frequency sound waves into the part. In these methods, a low-frequency acoustic transducer is attached to the structure through a couplant.

As the driving frequency of the transducer is varied, the amplitude and phase of the transducer oscillations change dramatically as it passes through a resonance. The phase and amplitude of these vibrations change very rapidly and reach a maximum as the driving frequency passes through the resonance frequency of the transducer. The effect of the structure is to dampen the resonant response of the transducer–block combination because of the transfer of acoustic energy into it. Defects such as delaminations and porosity in the adhesive bond layer increase the stiffness of the structure and lower the resonant frequency of the combination. The amplitude of the resonance is increased since there is less material to adsorb the sound energy. These changes in the sharpness of the resonant response are easily detectable electronically.

An alternate method of detecting flaws in bonded components is with a low-frequency or sonic instrument that senses the change in the time of flight for sound waves in the layered structure due to the presence of planar delamination. In such instruments, the increased time of traveling from a transmitting to a receiving transducer is detected electronically. Several commercially available bond-testing instruments successfully exploit this principle. A clever adaptation of a commercial version of this instrument has recently been used to successfully test the joints of structures made from sheet molding compounds (Papadakis and Chapman, 1993).

Probably, the most often used method of detecting delaminations in laminated structures is with a coin or tap hammer. This simple instrument is surprisingly effective in trained hands at detecting flaws since an exceedingly complex computer interprets the output signal, that is, the human brain. Consider, for a moment, that most parents can easily hear their child playing a musical instrument at a school concert. They can perform this task even though their child may have a minor part to play and all the other instruments are much louder than the one that their child is playing. With this powerful real-time signal-processing capability, inspectors can often detect flaws that cannot be detectable with current instrumentation and computers.

40.5 MAGNETIC PARTICLE METHOD

The magnetic particle method of nondestructive testing is used to locate surface and subsurface discontinuities in ferromagnetic materials (Schmidt and Skeie, 2001). An

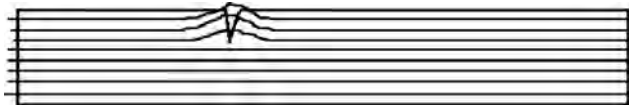


FIGURE 40.27 Representation of magnetic lines of flux in ferromagnetic metal near a flaw. Small magnetic particles are attracted to the leakage field associated with the flaw.

excellent reference for this NDE method is Bray and Stanley (1989), especially Chapters 10–16. Magnetic particle inspection is based on the principle that magnetic lines of force, when present in a ferromagnetic material, are distorted by changes in material continuity, such as cracks or inclusions, as shown schematically in Figure 40.27. If the flaw is open to the surface or close to it, the flux lines escape the surface at the site of the discontinuity. Near-surface flaws, such as nonmagnetic inclusions, cause the same bulging of the lines of force above the surface. These distorted fields, usually referred to as a leakage fields, reveals the presence of the discontinuity when fine magnetic particles are attracted to them during magnetic particle inspection. If these particles are fluorescent, their presence at a flaw will be visible under UV light, much like penetrant indications. Magnetic particle inspection is used for steel components because it is fast and easily implemented and has rather simple flaw indications. The part is usually magnetized with an electric current and then a solution containing fluorescent particles is applied by flowing it over the part. The particles stick to the part, forming the indication of the flaw.

40.5.1 Magnetizing Field

The magnetizing field may be applied to a component by a number of methods. Its function is to generate a residual magnetic field at the surface of the part. The application of a magnetizing force (H) generates a magnetic flux (B) in the component, as shown schematically in Figure 40.28. In this figure, the magnetic flux density B has units of newtons per

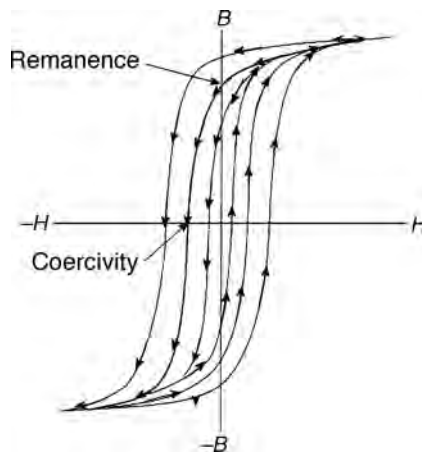


FIGURE 40.28 Magnetic flux intensity H versus magnetic flux density B hysteresis curve for typical steel. Initial magnetization starts at the origin and progresses as shown by the arrows. Demagnetization follows the arrows of the smaller hysteresis loops.

ampere or webers per square meter, and the strength of the magnetic field or magnetic flux intensity H has units of oersteds or amperes per meter. Starting at the origin, a magnetizing force is applied and the magnetic field internal to the part increases in a nonlinear fashion along the path shown by the arrows. If the force is reversed, the magnetic field does not return to zero but follows the arrows around the curve as shown. The reader will note that once the magnetizing force is removed, the flux density does not return to zero but remains at an elevated value called the material's remanence. This is the point at which most magnetic particle inspections are performed. The reader will also note that an appreciable reverse magnetic force H must be applied before the internal field density is again zero. This point is referred to as the coercivity of the material. If the magnetizing force is applied and reversed, the material will respond by continually moving around this hysteresis loop.

Selection of the type of magnetizing current depends primarily on whether the defects are open to the surface or are wholly below it. Alternating-current (ac) magnetization is best for the detection of surface discontinuities because the current is concentrated in the near-surface region of the part. Direct-current (dc) magnetization is best suited for subsurface discontinuities because of its deeper penetration of the part. While dc can be obtained from batteries or dc generators, it is usually produced by half-wave or full-wave rectification of commercial power. Rectified current is classified as half-wave direct current (HWDC) or full-wave direct current (FWDC). The ac fields are usually obtained from conventional power mains, but it is supplied to the part at reduced voltage for reasons of safety and the high-current requirements of the magnetizing process.

Two general types of magnetic particles are available to highlight flaws. One type is low-carbon steel with high-permeability and low-retentivity particles, which are used dry and consist of different sizes and shapes to respond to leakage fields. The other type of is very fine particles of magnetic iron oxide that are suspended in a liquid (either a petroleum distillate or water). These particles are smaller and have a lower permeability than the dry particles. Their small mass permits them to be held by the weak leakage fields at very fine surface cracks. Magnetic particles are available in several colors to increase their contrast against different surfaces or backgrounds. Dry powders are typically gray, red, yellow, and black, while wet particles are usually red, black, or fluorescent.

40.5.2 Continuous Versus Noncontinuous Fields

Because the field is always stronger while the magnetizing current is on, the continuous magnetizing method is generally preferred. Additionally, for specimens with low retentivity, this continuous method is often preferred. In the continuous method, the current can be applied in short pulses, typically 0.5 s. The magnetic particles are applied to the surface during this interval and are free to move to the site of the leakage fields. Liquid suspended fluorescent particles produces the most sensitive indications. For field inspections, the magnetizing current is often continuously applied during the test to give time for the powder to migrate to the defect site. In the residual method, the particles are applied after the magnetizing current is removed. This method is particularly well suited for production inspection of multiple parts.

The choice of direction of the magnetizing field within the part involves the nature of the flaw and its direction with respect to the surface and the major axis of the part. In circular magnetization, the field runs circumferentially around the part. It is induced into

the part by passing current through it between two contacting electrodes. Since flaws perpendicular to the magnetizing lines are readily detectable, circular magnetization is used to detect flaws that are parallel or less than 45° to the surface of the long, circular specimens. Placing the specimen inside a coil to create a field running lengthwise through the part produces longitudinal magnetization. This induction method is used to detect transverse discontinuities to the axis of the part.

40.5.3 Inspection Process

The surface of the part to be examined should be clean, dry, and free of contaminants such as oil, grease, loose rust, loose sand, loose scale, lint, thick paint, welding flux, and weld splatter. Cleaning of the specimen may be accomplished with detergents, organic solvents, or mechanical means, such as scrubbing or grit blasting.

Portable and stationary equipment are available for this inspection process. Selection of the specific type of equipment depends on the nature and location of testing. Portable equipment is available in lightweight units (35–90 lb) that can be readily taken to the inspection site. Generally, these units operate at 115, 230, or 460 V ac and supply current outputs of 750–1500 A in half-wave ac.

40.5.4 Demagnetizing the Part

Once the inspection process is complete, the part must be demagnetized. This is done by one of several ways depending on the subsequent usage of the component. A simple method of demagnetizing to remove residual magnetism from small tools is to draw it through the loop-shaped coil tip of a soldering iron. This has the effect of retracing the hysteresis loop a large number of times, each time with a smaller magnetizing force. When completely withdrawn, the tool will then have a very small remnant magnetic field, which for all practical purposes is zero. This same process is accomplished with an industrial part by slowly reducing and reversing the magnetizing current until it is essentially zero, as shown schematically by the arrows in Figure 40.28. Another method of demagnetizing a part is to heat it above its Curie temperature (about 550°C for iron), at which point all residual magnetism disappears. This last process is the best means of removing all residual magnetism, but it requires the expense and time of an elevated heat treatment.

40.6 THERMAL METHODS

Thermal nondestructive inspection methods involve the detection of infrared (IR) energy emitted from the surface of a test object (Maldague, 2001). This technique is used to detect the flow of thermal energy either into or out of a specimen and the effect of anomalies have on the surface temperature distribution. The material properties that influence this method are heat capacity, density, thermal conductivity, and emissivity. Defects that are usually detected include porosity, cracks, and delaminations that are parallel to the surface. The sensitivity of any thermal method is greatest for near-surface flaws that impede heat flow and degrades rapidly for deeply buried flaws in high-conductivity materials. Materials with lower thermal conductivity yield better resolution because they allow larger thermal gradients.

40.6.1 IR Cameras

All objects emit IR radiation with a temperature above absolute zero. At room temperature, the thermal radiation is predominately IR with a wavelength of approximately 10 μm . IR cameras are available that can produce images from this radiation and are capable of viewing large areas by scanning. Since the IR images are usually captured and stored in digital form, image processing is easily performed and the enhanced images are stored on magnetic or optical media. For many applications, an uncalibrated thermal image of a specimen is sufficient to detect flaws. However, if absolute temperatures are required, the IR instrumentation must be calibrated to account for the surface emissivity of the test subject.

The ability of thermography to detect flaws is often affected by the type of flaw and its orientation with respect to the surface of the object. To have a maximum effect on the surface temperature, the flaw must interrupt heat flow to the surface. Since a flaw can occur at any angle to the surface, the important parameter is its projected area to the camera. Subsurface flaws such as cracks parallel to the surface of the object, porosity, and debonding of a surface layer are easily detected. Cracks that are perpendicular to the surface can be very difficult or impossible to detect using thermography.

Most thermal NDE methods do not have good spatial resolution due to spreading of thermal energy as it diffuses to the surface. The greatest advantage of thermography is that it can be a noncontact, remote-viewing technique requiring only line-of-sight access to one side of a test specimen. Large areas can be viewed rapidly, since scan rates for IR cameras run between 16 and 30 frames per second. Temperature differences of 0.02 $^{\circ}\text{C}$ or less can be detected in a controlled environment.

40.6.2 Thermal Paints

A number of contact thermal methods are available for inspection purposes. These usually involve applying a coating to the sample and observing a color change as the specimen is thermally cycled. Several different types of coatings are available that cover a wide temperature ranges. Temperature-sensitive pigments in the form of paints have been made to cover a temperature range 40–1600 $^{\circ}\text{C}$. Thermal phosphors emit visible light when exposed to UV radiation. (The amount of visible light is inversely proportional to temperature.) Thermochromic compounds and cholesteric liquid crystals change color over large temperature ranges. The advantages of these approaches are the simplicity of application and relatively low cost if only small areas are scanned.

40.6.3 Thermal Testing

Excellent results may be achieved for thermographic inspections performed in dynamic environments where the transient effects of heat flow in the test object can be monitored. This enhances detection of areas where different heat transfer rates are present. Applications involving steady-state conditions are more limited. Thermography has been successfully used in several different areas of testing. In medicine, it is used to detect subsurface tumors. In aircraft manufacturing and maintenance, it may be used to detect debonding in layered materials and structures. In the electronics industry, it is used to detect poor thermal performance of circuit board components. Recently thermography has been used to detect stress-induced thermal gradients around defects in dynamically loaded test samples. For more information on thermal NDE methods, the reader is referred to Refs. Maldague, (2001); Stanley, (1995); Zorc, (1989).

40.7 EDDY CURRENT METHODS

40.7.1 Eddy Current Inspection

Eddy current (EC) methods are used to inspect electrically conducting components for flaws. Flaws that cause a change in electrical conductivity or magnetic permeability such as surface-breaking cracks, subsurface cracks, voids, and errors in heat treatment are detectable using EC methods. Thickness measurements and the thickness of non-conducting coatings on metal substrates can also be determined with EC methods (Udpa, 2004). Quite often, several of these conditions can be monitored simultaneously if instrumentation capable of measuring the phase of the EC signal is used.

This inspection method is based on the principle that ECs are induced in a conducting material when a coil (probe) is excited with an alternating or transient electric current that is placed in close proximity to the surface of a conductor. The induced currents create an electromagnetic field that opposes the field of the inducing coil in accordance with Lenz's law. The ECs circulate in the part in closed, continuous paths, and their magnitude depends on many variables. These include the magnitude and frequency of the current in the inducing coil, the coil's shape and position relative to the surface of the part, electrical conductivity, magnetic permeability, shape of the part, and presence of discontinuities or inhomogeneities within the material. Therefore, EC inspection is useful for measuring the electrical properties of materials and detecting discontinuities or variations in the geometry of components.

40.7.1.1 Skin Effect EC inspections are limited to the near-surface region of the conductor by the skin effect. Within the material, the EC density decreases with the depth. The density of the EC field falls off exponentially with depth and diminishes to a value of about 37% of the surface value at a depth referred to as the standard depth of penetration (SDP). The SDP in meters is calculated with the formula

$$SDP = \frac{1}{\sqrt{\pi f \sigma \mu}},$$

where f = test frequency, Hz; σ = test material's electrical conductivity, mho/m; μ = permeability, H/m.

The latter quantity is the product of the relative permeability of the specimen, 1.0 for nonmagnetic materials, and the permeability of free space, $4\pi \times 10^{-7}$ H/m.

40.7.1.2 Impedance Plane While the SDP is used to give an indication of the depth from which useful information can be obtained, the choice of the independent variables in most test situations is usually made using the impedance plane diagram suggested by Förster (1952). It is theoretically possible to calculate the optimum inspection parameters from numerical codes based on Maxwell's equations, but this is a laborious task that is justified in special situations.

The ECs induced at the surface of a material are time varying and have amplitude and phase. The complex impedance of the coil used in the inspection of a specimen is a function of a number of variables. The effect of changes in these variables can be conveniently displayed with the impedance diagram, which shows the variations in amplitude and phase of the coil impedance as functions of the dependent variables specimen

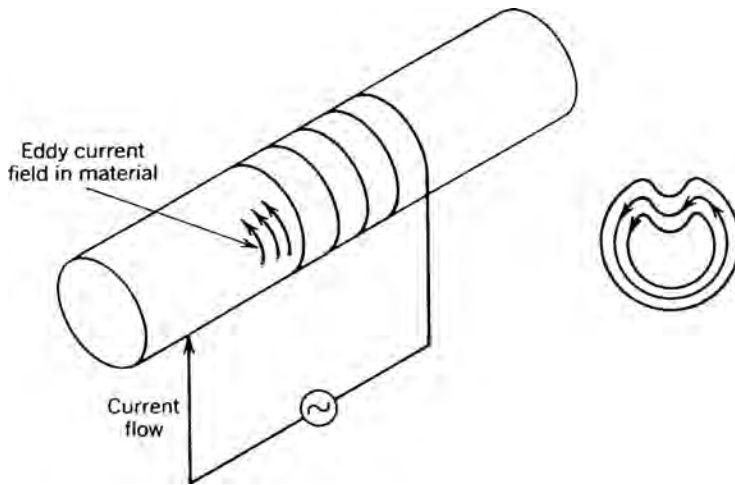


FIGURE 40.29 Representations of EC inspection of solid cylinder. Also shown are the EC paths within the cross section of the cylinder near the crack.

conductivity, thickness, and distance between the coil and specimen, or lift-off. For the case of an encircling coil on a solid cylinder, shown schematically in Figure 40.29, the complex impedance plane is displayed in Figure 40.30. The reader will note that the ordinate and abscissa are normalized by the inductive reactance of the empty coil. This eliminates the effect of the geometry of the coil and specimen. The numerical values shown on the large curve, which are called reference numbers, are used to combine the effects of the conductivity, size of the test specimen, and frequency of the measurement into a single parameter. This yields a diagram that is useful for most test conditions. The reference numbers shown on the outermost curve are obtained with the following relationship for nonmagnetic materials:

$$\text{References number} = r\sqrt{2\pi f\mu\sigma},$$

where r = radius of bar, m; f = frequency of test, Hz; m = magnetic permeability of free space, $4\pi \times 10^{-7}$ H/m; σ = conductivity of specimen, mho/m.

The outer curve in Figures 40.30 and 40.31 is useful only for the case where the coil is the same size as the solid cylinder, which can never happen. For those cases where the coil is larger than the test specimen, which is usually the case, a coil-filling factor is calculated. This is quite easily accomplished with the formula

$$N = \frac{\text{diameter}_{\text{specimen}}}{\text{diameter}_{\text{coil}}}.$$

Figure 40.30 shows the impedance plane with a curve for specimen/coil inspection geometry with a fill factor of 0.75. Note that the reference numbers on the curves representing the different fill factors can be determined by projecting a straight line from point 1.0 on the ordinate to the reference number of interest, as is shown for the reference

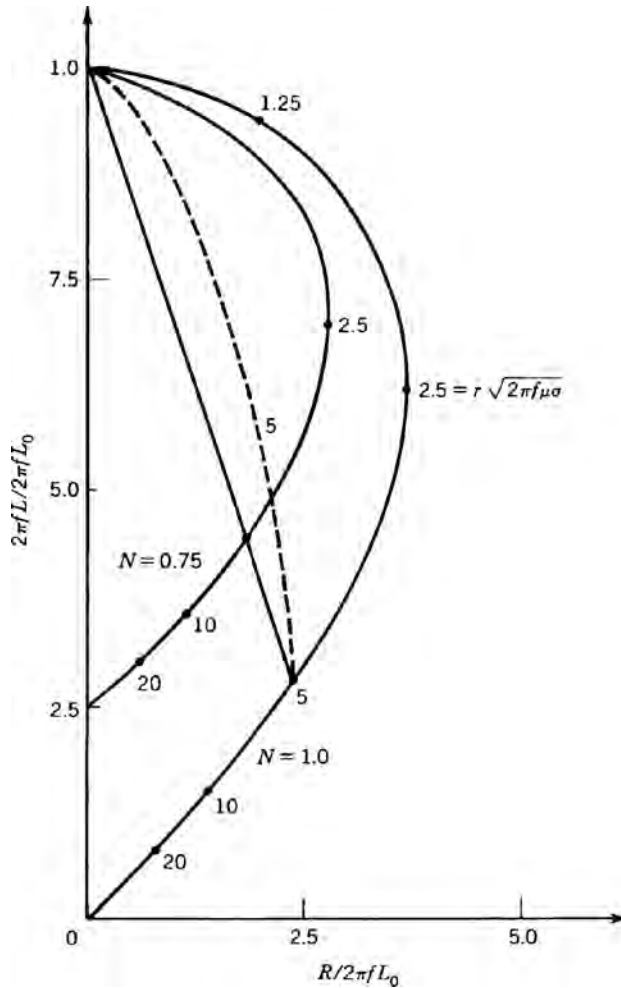


FIGURE 40.30 Normalized impedance diagram for long encircling coil on solid, nonferromagnetic cylinder. For $N=1$, the coil and cylinder have the same diameter, while for $N=0.75$ the coil is approximately 1.155 times larger than the cylinder.

number 5.0. Both the fill factor and the reference number change when the size of either the specimen or coil changes. Assume that a reference number of 5.0 is appropriate to a specific test with $N=1.0$; if the coil diameter is changed so that the fill factor becomes 0.75, then the new reference number will be equal to approximately 7. While the actual change in reference number for this case follows the path indicated by the dashed line in Figure 40.30, we have estimated the change along a straight line. This yields a small error in optimizing the test setup but is sufficient for most purposes. For a more detailed treatment of the impedance plane, the reader is referred to Udpa, (2004). The inspection geometry discussed thus far has been for a solid cylinder. The other geometry of general interest is the thin-walled tube. In this case, the skin effect limits the thickness of the metal that may be effectively inspected.

For an infinitely thin-walled tube, the impedance plane is shown in Figure 40.31, which includes the curve for a solid cylinder. The dashed lines that connect these two cases are for thin-walled cylinders of varying thicknesses. The semicircular curve for the thin cylinder is used in the same manner as described above for the solid cylinder.

40.7.1.3 Lift-off of Inspection Coil from Specimen In most inspection situations, the only independent variables are frequency and lift-off. High-frequency excitations are frequently used for detecting defects such as surface-connected cracks or corrosion, while low frequencies are used to detect subsurface flaws. It is also possible to change the coil shape and measurement configuration to enhance detectability, but the discussion of

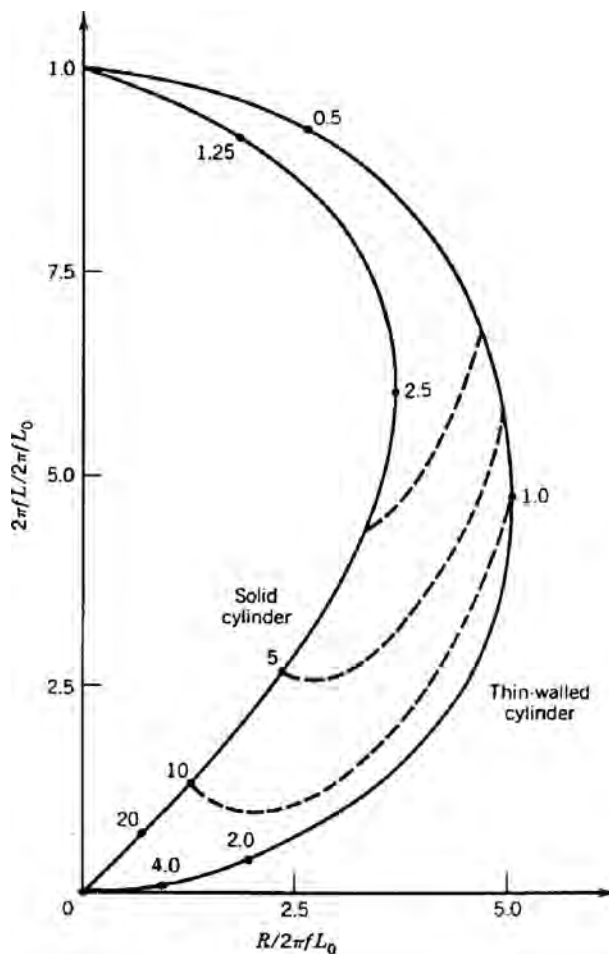


FIGURE 40.31 Normalized impedance diagram for long encircling coil on both solid and thin-walled conductive but nonferromagnetic cylinders. The dashed lines represent the effects of varying wall thicknesses.

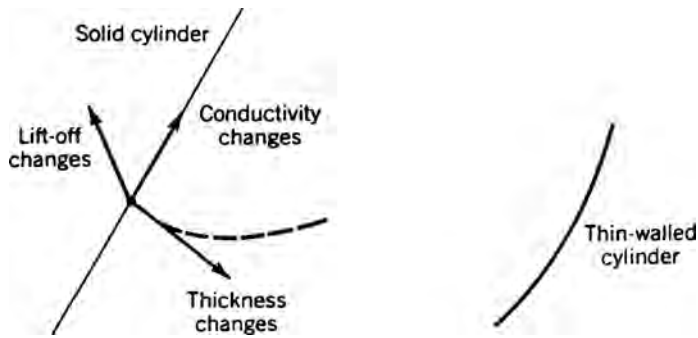


FIGURE 40.32 Effects of various changes in inspection conditions on local signal changes in impedance plane of Figure 40.31. Phase differentiation is relatively easily accomplished with current instrumentation.

these more complex parameters is beyond the scope of this chapter and the reader is referred to the literature. The relationships discussed so far may be applied by examining Figure 40.32, where changes in thickness, lift-off, and conductivity are represented by vectors. These vectors all point in different directions representing the phases of the different possible signals. Instrumentation with phase discrimination circuitry can differentiate between these signals and therefore is often capable of detecting two changes in specimen condition at once. Changes in conductivity can arise from several different conditions. For example, aluminum alloys can have different conductivities depending on their heat treatment. Changes in apparent conductivity are also due to the presence of cracks or voids. A crack decreases the apparent conductivity of the specimen because the ECs must travel a longer distance to complete their circuit within the material. Lift-off and wall thinning are also shown in Figure 40.32. Thus, two different flaw conditions can be rapidly detected. There are situations where changes in wall thickness and lift-off result in signals that are very nearly out of phase and therefore the net change is not detectable. If this situation is suspected, then inspection at two different frequencies is warranted. There are other inspection situations that cannot be covered in this brief description. These include the inspection of ferromagnetic alloys, plate, and sheet stock and the measurement of film thicknesses on metal substrates. For a treatment of these and other special applications of EC inspection, the reader is referred to Udpa (2004).

40.7.2 Probes and Sensors

In some situations, it may be advantageous to have a core with a high magnetic permeability inside the coil. Magnetic fields will pass through the medium with the highest permeability if possible. Therefore, materials with high permeability can be placed in different geometric configurations to enhance the sensitivity of a probe. One example of this is the cup-core probe where the coil has a ferrite core, shield, and cap (Vernon, 1989).

For low frequency or transient tests, using the inductive coil as a sensor is not sufficient since it responds to the time change in magnetic field and not the direct

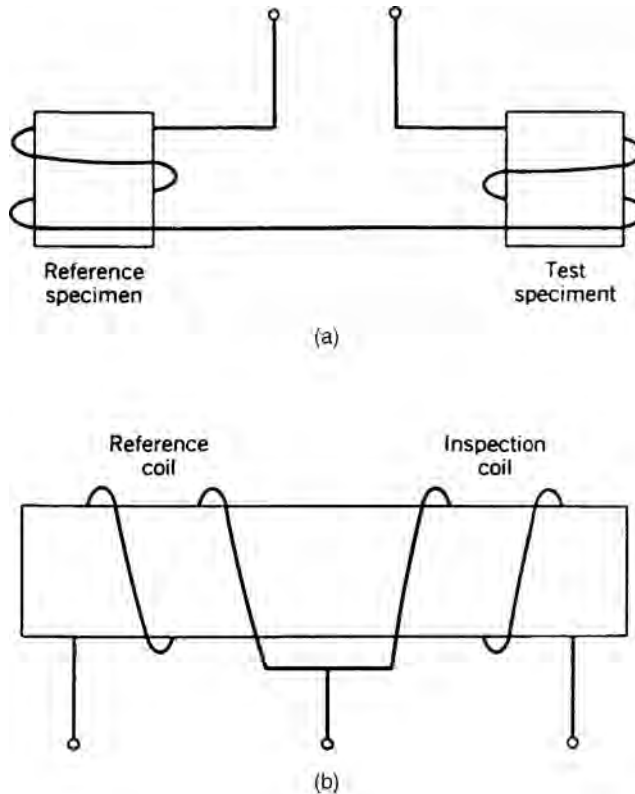


FIGURE 40.33 Representation of (a) absolute versus (b) differential coil configurations used in EC testing.

magnetic field. It is necessary to induce a magnetic field sensor that responds well at the lower frequencies. Sensors such as the Hall effect sensor and giant magneto-resistive (GMR) sensors have been used to accomplish this (Wincheski, 2002).

There are numerous methods of making EC NDE measurements. Two of the more common methods are shown schematically in Figure 40.33. In the absolute coil arrangement, very accurate measurements can be made of the differences between the two samples. In the differential coil method, it is the differences between the two variables at two slightly different locations that may be detected. For this arrangement, slightly varying changes in dimensions and conductivity are not sensed, while singularities such as cracks or voids are highlighted, even in the presence of other slowly changing variables. Since the specific electronic circuitry used to accomplish this task can vary dramatically, depending on the specific inspection situation, the reader is referred to the current NDE and instrumentation Förster, (1952); Blitz, (1997); Libby, (1971).

APPENDIX: Ultrasonic Properties of Common Materials

Material	Comments	Density (g/cm ³)	V _l (km/s)	V _s (km/s)	Impedance (MRayl)	Attenuation (dB/cm/MHz)	Attenuation (dB/cm at 5 MHz)
<i>Gases</i>							
Alcohol vapor			0.23				
Air	25 atm		0.33				
	50 atm		0.34				
	100 atm		0.35				
			0.34				
			0.39				
			0.55				
Ammonia			0.42				
Argon		0.00178	0.32				
Carbon monoxide		0.34					
Carbon dioxide			0.26				
Carbon disulfide			0.19				
Chlorine			0.21				
Ether vapor			0.18				
Ethylene			0.31				
Helium			0.97				
Hydrogen		0.00018	1.28				
Methane		0.00009	0.43				
Neon		0.00074	0.43				
Nitric oxide		0.0009	0.33				
Nitrogen		0.00125	0.33				
		0.00116	0.35				
			0.26				
Nitrous oxide			0.32				
Oxygen		0.00142	0.33				
		0.00132	0.4				
Water vapor			0.41				
			0.42				

(continued)

APPENDIX: Ultrasonic Properties of Common Materials

Material	Comments	Density (g/cm ³)	V _l (km/s)	V _s (km/s)	Impedance (MRayl)	Attenuation (dB/cm/MHz)	Attenuation (dB/cm at 5 MHz)
<i>Gases—cryogenic</i>							
Argon		1.404	0.84			1.18	
		1.424	0.86			1.23	
Helium		0.125	0.18			0.023	
		0.146	0.23			0.034	
Helium-4	Liquid at 2 K	0.15	0.18			0.027	
Hydrogen	Liquid at 20 K	0.07	1.19			0.08	
		0.355	1.13			0.401	
Nitrogen		0.815	0.87			0.708	
		0.843	0.93			0.783	
Oxygen		1.143	0.97			1.04	
		1.272	1.13			1.44	
		1.149	0.95			1.09	
<i>Liquids</i>							
Acetate butyl (n)		0.871	1.17			1.02	
Acetate ethyl		0.9	1.18			1.06	
Acetate methyl		0.928	1.15			1.07	
Acetate propyl		0.891	1.18			1.05	
Acetone		0.79	1.17		0.92	0.92	0.0469
		0.791	1.16	—		—	
Acetonitrile		0.783	1.29			1.01	
Acetonyl acetone		0.729	1.4			1.36	
Acetylene dichloride		1.26	1.02			1.29	
Adiprene	CW-520	0.79	1.68	—	1.33	—	0.0469
Alcohol, butyl		0.81	1.24			1	
Alcohol, ethyl		0.789	1.18			0.93	
Alcohol, furfuryl		1.135	1.45			1.65	
Alcohol, isopropyl		0.79	1.17	—	0.92	—	0.08
Alcohol, methyl		0.792	1.12			0.89	

APPENDIX: Ultrasonic Properties of Common Materials

Material	Comments	Density (g/cm ³)	V _l (km/s)	V _s (km/s)	Impedance (MRayl)	Attenuation (dB/cm/MHz)	Attenuation (dB/cm at 5 MHz)
Alcohol, propyl (<i>i</i>)		0.786	1.17			0.92	
Alcohol, propyl (<i>n</i>)		0.804	1.22			0.98	
Alcohol, <i>t</i> -amyl		0.81	1.2			0.97	
Alkazine 13		0.86	1.32			1.14	
Analine		1.022	1.69			1.68	
A-Spirit	Ethanol >96%	0.79	1.18			0.93	
Benzene		0.87	1.3			1.13	
	C ₆ H ₆	0.88	1.31			1.15	
Benzol		0.878	1.33			1.17	
Benzol ethyl		0.868	1.34			1.16	
Bromo-benzene		1.52	1.17			1.78	
Bromoform	C ₆ H ₅ Br	2.89	0.92			2.66	
Butanol	Butyl	0.71	1.27	—	0.9		
		0.81	1.27			1.03	
Butoxyethanol	(2 <i>n</i> -)		1.31				
<i>tert</i> -Butyl chloride		0.84	0.98			0.82	
Butylene glycol (2.3)		1.019	1.48			1.51	
Butyrate ethyl		0.877	1.17			1.03	
Carbitol		0.988	1.46			1.44	
Carbon disulfide			1.16				
		1.26	1.15			1.45	
Carbon tetrachloride		1.595	0.93			1.48	
Cerechlor 42	CCl ₄	1.26	1.43			1.8	
Chinolin		1.09	1.57			1.71	
Chlorobenzene		1.1	1.3			1.43	
Chloroform	C ₆ H ₅ Cl	1.49	0.99			1.47	
Chlorohexanol	>98%	0.95	1.42			1.35	
Cyclohexanol		0.962	1.45			1.39	
	Freon		1.2				

(continued)

APPENDIX: Ultrasonic Properties of Common Materials

Material	Comments	Density (g/cm ³)	V _l (km/s)	V _s (km/s)	Impedance (MRayl)	Attenuation (dB/cm/MHz)	Attenuation (dB/cm at 5 MHz)
Decahydro-naphthalene	DTE 21 oil		1.39				
	Glycerol	0.948	1.52			1.39	
			1.42				
Diacetyl	C ₁₀ H ₁₈	0.89	1.42			1.27	
	Paraffin		1.41				
			1.24			1.22	
Diamine propane	(1.3) >99%	0.99	1.66			1.47	
		0.89	1.22			1.39	
		1.14	1.13			0.8	
Dichloro isobutane (1.3)	(C ₂ H ₅) ₂ NH	0.7	1.58			1.76	
		1.116	1.31			1.07	
		0.813	1.46			1.75	
Diethylamine		1.2	1.38			1.43	
		1.033	1.5				
			1.41			1.16	
Diethylene glycol	Diphenyl oxide	0.83	1.39				
			1.42				
			1.43				
Diethyl ketone			1.44				
			1.55				
			1.13		0.89	—	0.0421
Dimethyl phthalate			1.72			1.75	
			1.19			1.07	
			1.67				
Dioxane			1.69			1.52	
			1.11			1.88	
Diphenyl							
Dodecanol							
DTE 21							
DTE 24							
DTE 26							
Dubanol							
Ethanol							
Ethanol amide							
Ethyl acetate							
Ethylencycolor							
Ethylene diamine							
Ethylene glycol							

APPENDIX: Ultrasonic Properties of Common Materials

Material	Comments	Density (g/cm ³)	V _l (km/s)	V _s (km/s)	Impedance (MRayl)	Attenuation (dB/cm/MHz)	Attenuation (dB/cm at 5 MHz)
Ethyl ether Fluorinert	1,2-Ethanediol	1.112	1.67			1.86	
		1.113	1.66			1.85	
	H ₂ O 1 : 4		1.6				
	H ₂ O 2 : 3		1.68				
	H ₂ O 3 : 2		1.72				
	H ₂ O 4 : 1		1.72				
		0.713	0.99			0.7	
	H ₂ O-72	1.68	0.51			0.86	
	FC-104	1.76	0.58			1.01	
	FC-75	1.76	0.59			1.02	
Fluoro-benzene Formamide Freon	FC-77		0.6			1.05	
	FC-43	1.85	0.66			1.21	
	FC-40	1.86	0.64			1.19	
	FC-70	1.94	0.69			1.33	
	C ₆ H ₅ F	1.024	1.18			1.21	
		1.134	1.62			1.84	
			0.68				
		1.485	0.8			1.19	
		1.574	0.97			1.52	
		1.157	1.45			1.68	
Furfural Gasoline Glycerine		0.803	1.25			1	
	CH ₃ OHCHO	1.23	1.9			2.34	
	HCH ₂ OH						
	Glycerol <98%	1.26	1.88			2.37	
		1.26	1.92			2.42	
	Water	1.22	1.88			2.29	
	Butanol		1.45				
	Ethanol		1.52				
			1.56				

(continued)

APPENDIX: Ultrasonic Properties of Common Materials

Material	Comments	Density (g/cm ³)	V _l (km/s)	V _s (km/s)	Impedance (MRayl)	Attenuation (dB/cm/MHz)	Attenuation (dB/cm at 5 MHz)
Glycerol trioleate Glycol	Isopropanol	0.91	1.57			1.31	
	Polyethylene	1.06	1.44			1.71	
	Ethylene	1.087	1.62			1.75	
Hexane <i>n</i> -Hexanol Honey		1.108	1.59			1.76	
		1.112	1.67			1.86	
	(<i>n</i> -)C ₆ H ₁₄	0.659	1.1			0.727	
Iodobenzene Isopentane Isopropanol	Sue Bee Orange C ₆ H ₅ I	0.819	1.3			1.06	
		1.42	2.03			2.89	
		1.183	1.1			2.01	
Isopropyl alcohol		0.62	0.99			0.62	
			1.14				
		0.786	1.17			0.92	
Propylene glycol		0.79	1.14			0.9	0
		0.84	1.21			1.01	0.14
		0.88	1.24			1.09	0.4
		0.88	1.28			1.13	0.23
		0.92	1.3			1.2	0.44
		0.94	1.35			1.27	0.33
		0.96	1.36			1.31	0.53
		0.97	1.36			1.32	1.12
		0.99	1.43			1.41	0.47
		1	1.42			1.42	1.08
		1	1.43			1.43	0.67
		1.03	1.48			1.52	1.13
		1.03	1.51			1.56	0.72
		1.04	1.54			1.64	1.2

APPENDIX: Ultrasonic Properties of Common Materials

Material	Comments	Density (g/cm ³)	V _l (km/s)	V _s (km/s)	Impedance (MRayl)	Attenuation (dB/cm/MHz)	Attenuation (dB/cm at 5 MHz)
Jeffox WL-1400		1.07	1.58			1.7	0.93
Kerosene		1.09	1.6			1.75	1.4
Linalool		1.13	1.65			1.86	1.68
Mercury		1.16	1.75			2.03	2.57
Mercury, 20°C		1.2	1.82			2.19	2.12
Mesityloxide		1.25	1.92			2.39	4.55
Methanol		0.81	1.32			1.07	
Methyl acetate		0.884	1.4			1.24	
Methylene iodide	Hg	13.6	1.45			19.7	
Methylethyl ketone			1.42			19.7	
Methyl naphthalene		0.85	1.31			1.11	
Methyl salicylate	CH ₃ OH	0.796	1.09	—	0.87	0.87	0.0262
Modinet P40		0.934	1.21			1.13	
Monochloro- benzene			0.98				
Morpholine		0.805	1.21			0.97	
M-xylol		1.09	1.51			1.65	
NaK		1.16	1.38			1.6	
		1.06	1.38			1.47	
		1.107	1.27			1.41	
		1	1.44			1.44	
		0.864	1.32			1.14	
	Mix	0.64	1.66				
		0.713	1.72				
		0.714	1.84				
		0.73	1.95				
		0.736	1.77				
		0.738	1.89				

(continued)

APPENDIX: Ultrasonic Properties of Common Materials

Material	Comments	Density (g/cm ³)	V _l (km/s)	V _s (km/s)	Impedance (MRayl)	Attenuation (dB/cm/MHz)	Attenuation (dB/cm at 5 MHz)
Nicotine Nitrobenzene Nitrogen Nitromethane Oil, baby Oil, castor Oil, corn Oil, cutting	C ₁₀ H ₁₄ N ₂ N ₂ Jeffox WL-1400 Castor Ricinus oil 64 AS (red)	0.754	2				
		0.759	1.82				
		0.761	1.99				
		0.778	2.05				
		0.781	1.88				
		0.784	1.99				
		0.801	2.1				
		0.804	1.93				
		0.807	2.04				
		0.825	2.15				
		0.826	1.98				
		0.83	2.09				
		0.848	2.2				
		0.849	2.04				
		0.853	2.14				
		0.871	2.25				
		0.876	2.19				
		0.893	2.31				
		1.01	1.49			1.51	
		1.2	1.46			1.75	
		0.8	0.86			0.68	
		1.13	1.33			1.5	
		0.821	1.43			1.17	
		—	1.52				
		0.95	1.54			1.45	
		0.969	1.48			1.43	
		0.922	1.46			1.34	
			1.4				

APPENDIX: Ultrasonic Properties of Common Materials

Material	Comments	Density (g/cm ³)	V _l (km/s)	V _s (km/s)	Impedance (MRayl)	Attenuation (dB/cm/MHz)	Attenuation (dB/cm at 5 MHz)
Oil, diesel			1.25				
Oil, fluorosilicone	Dow FS-1265		0.76				
Oil, grape seed	Cerechlor	0.92	1.43				
	Castor oil	0.936	1.44			1.35	
Oil, gravity fuel AA		0.99	1.49			1.48	
Oil, linseed		0.922	1.77			1.63	
		0.94	1.46			1.34	
Oil, mineral (heavy)		0.843	1.46			1.37	
Oil, mineral (light)		0.825	1.44			1.19	
Oil, motor (2-cycle)			1.43				
Oil, motor (SAE 20)		0.87	1.74			1.51	
Oil, motor (SAE 30)		0.88	1.7			1.5	
Oil, olive			1.43				
		0.918	1.45			1.32	
		0.948	1.43			1.39	
Oil, paraffin			1.28				
			1.43				
		0.835	1.42			1.86	
Oil, peanut		0.914	1.44			1.31	
		0.936	1.46			1.37	
Oil, safflower		0.92	1.45			1.34	
Oil, silicone	Dow 710 fluid		1.35				
	Silicone 200	0.818	0.96			0.74	
		0.94	0.97			0.91	
		0.972	0.99			0.96	
	30 cP	0.993	0.99			0.983	
		1.1	1.37			1.5	
Oil, soybean		0.93	1.43			1.32	

(continued)

APPENDIX: Ultrasonic Properties of Common Materials

Material	Comments	Density (g/cm ³)	V ₁ (km/s)	V _s (km/s)	Impedance (MRayl)	Attenuation (dB/cm/MHz)	Attenuation (dB/cm at 5 MHz)
Oil, sperm		0.88	1.44			1.27	
Oil, sun	Nivea		1.41				
Oil, sunflower		0.92	1.45			1.34	
Oil, synthetic		0.98	1.27			1.33	
Oil, transformer		0.92	1.39			1.28	
Oil, transmission	Dexron (red)		1.42				
Oil, velocite	Mobil		1.3				
Oil, wheat germ		0.94	1.49			1.39	
Paraffin		1.5	1.5			2.3	
<i>d</i> -Penchone		0.94	1.32			1.24	
Pentane		0.621	1.01			0.63	
	(<i>n</i> -)C ₅ H ₁₂	0.626	1.03			0.64	
		0.825	1.29			1.06	
Petroleum			1.3				
Polypropylene glycol	Polyglycol P-400		1.3				
	Polyglycol P-1200		1.37				
	Polyglycol E-200		1.49				
	Ambiflo	0.662	1.55				
Polypropylene oxide		0.707	1.6				
Potassium		0.729	1.65				
		0.751	1.71				
		0.773	1.76				
		0.796	1.81				
		0.818	1.86				
Propane diol	(1.3) >97%	1.05	1.62			1.7	
Pyridine		0.982	1.41			1.38	
Sodium		0.759	2.15				
		0.784	2.21				

APPENDIX: Ultrasonic Properties of Common Materials

Material	Comments	Density (g/cm ³)	V _l (km/s)	V _s (km/s)	Impedance (MRayl)	Attenuation (dB/cm/MHz)	Attenuation (dB/cm at 5 MHz)
		0.809	2.26				
		0.833	2.31				
		0.857	2.37				
		0.881	2.42				
		0.904	2.48				
		0.926	2.53				
Solvesso #3		0.877	1.37			0.201	
Sonotrack	Coupling gel	1.4	1.62			1.68	
Span 20			1.48				
Span 85			1.46				
Tallow			0.39				
Tetraethylene glycol		1.12	1.58			1.77	
Tetrahydro- naphthalene	(1.2.3.4)	0.97	1.47			1.42	
Trichloro-ethylene		1.05	1.05			0.41	
Triethylene glycol		1.12	1.61			1.81	
		1.123	1.61			1.98	
Trithylamine	(C ₂ H ₅) ₃ N	0.73	1.12			0.81	
Turpentine		0.87	1.25			1.11	
		0.893	1.28			1.14	
			1.27				
Ucon 75H450			1.54				
Univis 800		0.87	1.35			1.19	
Water—salt solution	10%		1.47				
	15%		1.53				
	20%		1.6				
Water, sea		1.025	1.53			1.57	
		1.026	1.5			1.54	
Water		1	1.51			1.51	

(continued)

APPENDIX: Ultrasonic Properties of Common Materials

Material	Comments	Density (g/cm ³)	V _l (km/s)	V _s (km/s)	Impedance (MRayl)	Attenuation (dB/cm/MHz)	Attenuation (dB/cm at 5 MHz)
		1	1.55			1.55	
	Propylene glycol	1	1.5			1.5	
		1.01	1.61			1.63	0.021
		1.02	1.69			1.72	0.038
		1.03	1.51			1.56	0.669
		1.03	1.62			1.66	0.213
		1.03	1.69			1.73	0.088
		1.05	1.6			1.69	
		1.06	1.69			1.79	0.059
		1.07	1.58			1.7	1.025
		1.07	1.66			1.78	0.395
		1.07	1.71			1.83	0.174
		1.07	1.73			1.84	0.112
		1.11	1.71			1.89	0.086
		1.11	1.75			1.95	0.321
		1.11	1.76			1.94	0.117
		1.11	1.77			1.97	0.182
		1.12	1.66			1.86	1.744
		1.12	1.71			1.91	0.582
		1.16	1.75			2.03	2.57
		1.16	1.78			2.06	1.242
		1.16	1.8			2.09	0.175
		1.16	1.81			2.09	0.648
		1.16	1.82			2.11	0.241
		1.16	1.82			2.11	0.397
		1.2	1.82			2.19	2.12
		1.2	1.85			2.23	1.469
		1.2	1.85			2.22	2.033
		1.2	1.86			2.24	1.023

APPENDIX: Ultrasonic Properties of Common Materials

Material	Comments	Density (g/cm ³)	V _l (km/s)	V _s (km/s)	Impedance (MRayl)	Attenuation (dB/cm/MHz)	Attenuation (dB/cm at 5 MHz)
Water	UCON 50HB400	1.2	1.87			2.24	0.731
		1.2	1.88			2.25	0.544
		1.25	1.92			2.39	4.55
		0.79	1.16			0.91	0.11
		0.83	1.25			1.04	0.06
		0.83	1.27			1.06	0.06
		0.83	1.28			1.07	0.07
		0.83	1.29			1.07	0.07
		0.84	1.21			1.01	0.15
		0.84	1.24			1.03	0.08
		0.87	1.41			1.23	0.3
		0.88	1.31			1.16	0.15
		0.88	1.34			1.18	0.15
		0.88	1.37			1.2	0.2
		0.88	1.4			1.22	0.24
		0.89	1.26			1.12	0.26
		0.91	1.54			1.4	0.55
		0.92	1.52			1.4	0.74
		0.93	1.44			1.33	0.4
		0.93	1.48			1.38	0.57
		0.94	1.32			1.24	0.35
		0.94	1.4			1.31	0.35
		0.96	1.63			1.57	1.38
		0.96	1.64			1.57	0.03
		0.97	1.54			1.5	1.06
		0.97	1.59			1.55	1.54
		0.99	1.38			1.37	0.52
		0.99	1.49			1.46	0.7
		1	1.48			1.48	

(continued)

APPENDIX: Ultrasonic Properties of Common Materials

Material	Comments	Density (g/cm ³)	V _l (km/s)	V _s (km/s)	Impedance (MRayl)	Attenuation (dB/cm/MHz)	Attenuation (dB/cm at 5 MHz)
		1	1.5			1.5	0
		1	1.5			1.5	0.04
		1.01	1.61			1.63	0.13
		1.01	1.61			1.63	0.13
		1.01	1.63			1.65	0
		1.01	1.69			1.71	0.4
		1.02	1.64			1.69	2.72
		1.02	1.66			1.7	2.11
		1.02	1.66			1.7	2.11
		1.02	1.66			1.7	2.11
		1.02	1.68			1.72	0.84
		1.02	1.69			1.72	1.5
		1.02	1.69			1.73	0.17
		1.02	1.69			1.73	0.29
		1.02	1.7			1.73	0.08
		1.02	1.71			1.74	0.44
		1.03	1.5			1.55	0.92
		1.03	1.5			1.55	0.92
		1.03	1.53			1.58	0.72
		1.03	1.53			1.58	0.72
		1.03	1.54			1.6	0.72
		1.03	1.54			1.6	0.72
		1.03	1.56			1.61	0.73
		1.03	1.56			1.61	0.73
		1.03	1.57			1.63	2.13
		1.03	1.57			1.61	0.92
		1.03	1.57			1.62	0.58
		1.03	1.57			1.63	2.13

APPENDIX: Ultrasonic Properties of Common Materials

Material	Comments	Density (g/cm ³)	V _l (km/s)	V _s (km/s)	Impedance (MRayl)	Attenuation (dB/cm/MHz)	Attenuation (dB/cm at 5 MHz)
		1.03	1.57			1.61	0.92
		1.03	1.57			1.62	0.58
		1.03	1.57			1.63	2.13
		1.03	1.58			1.63	1.25
		1.03	1.58			1.63	1.25
		1.03	1.6			1.65	0.47
		1.03	1.6			1.65	0.47
		1.03	1.61			1.65	0.56
		1.03	1.61			1.65	0.56
		1.03	1.62			1.67	0.38
		1.03	1.62			1.67	0.38
		1.03	1.63			1.68	0.9
		1.03	1.63			1.68	0.9
		1.03	1.64			1.69	2.72
		1.03	1.65			1.7	1.53
		1.03	1.65			1.7	0.46
		1.03	1.65			1.7	0.32
		1.03	1.65			1.7	1.53
		1.03	1.65			1.7	0.46
		1.03	1.65			1.7	0.32
		1.04	1.44			1.5	0.94
		1.04	1.44			1.5	0.94
		1.04	1.46			1.52	1.05
		1.04	1.48			1.53	1.02
		1.04	1.51			1.57	0.85
		1.04	1.4			1.55	
Water, D ₂ O		1.104					
Xylene hexafluoride		1.37	0.88			1.21	

(continued)

APPENDIX: Ultrasonic Properties of Common Materials

Material	Comments	Density (g/cm ³)	V _l (km/s)	V _s (km/s)	Impedance (MRayl)	Attenuation (dB/cm/MHz)	Attenuation (dB/cm at 5 MHz)
<i>Solids (metals and alloys)</i>							
Aluminum		2.7	6.32	3.1		17.1	
	Duraluminum	2.71	6.32	3.1		17.1	
Al 1100-0	2S0	2.71	6.35	3.1	2.9	17.2	
Al 2014	14S	2.8	6.32	3.1	—	17.7	
Al 2024 T4	24ST	2.77	6.37	3.2	2.95	17.6	
Al2117T4	17ST	2.8	6.5	3.1		18.2	
Antimony	Sb	3.4					
Bearing babbitt		10.1	2.3	—	—	23.2	
Beryllium		1.82	12.9	8.9	7.87	23.5	
Bismuth		9.8	2.18	1.1		21.4	
Brass	70% Cu–30% Zn	8.64	4.7	2.1		40.6	
		8.56	4.28	2		36.6	
	Half hard	8.1	3.83	2.1		31.0	
	Naval	8.42	4.43	2.1	1.95	37.3	
Bronze	Phospho	8.86	3.53	2.2	2.01	31.3	
Cadmium	Cd	8.6	2.8	1.5		42.0	
		8.64	2.78	1.5		24.0	
Cesium		1.88	0.97			1.82	
Columbium		8.57	4.92	2.1		42.2	
Constantan		8.88	5.24	2.6		46.5	
Copper		8.93	4.66	2.3	1.93	41.6	
Copper, rolled		8.9	5.01	2.3		44.6	
E-Solder	Cu	2.71	1.9	1		5.14	
Gallium		5.95	2.74	—	—	16.3	
Germanium		5.47	5.41	—	—	29.6	
Gold	Hard drawn	19.32	3.24	1.2	—	62.6	

APPENDIX: Ultrasonic Properties of Common Materials

Material	Comments	Density (g/cm ³)	V _l (km/s)	V _s (km/s)	Impedance (MRayl)	Attenuation (dB/cm/MHz)	Attenuation (dB/cm at 5 MHz)
Hafnium			3.84	—			
Inconel		8.25	5.72	3	2.79	64.5	
Indium		7.3	2.22			16.2	
Iron		7.7	5.9	3.2	2.79	45.4	
	Cast	7.22	4.6	2.6	—	33.2	
Lead		11.4	2.16	0.7	0.63	24.6	
	5% Antimony	10.9	2.17	0.8	0.74	23.7	
Magnesium		1.74	6.31			11.0	
	AM-35	1.74	5.79	3.1	2.87	10.1	
	FS-1	1.69	5.47	3		9.2	
	J-1	1.7	5.67	3		9.6	
	M	1.75	5.76	3.1		10.1	
	O-1	1.82	5.8	3		10.6	
	ZK-60A-TS	1.83	5.71	3.1		10.4	
		1.72	5.8	3		10.0	
Manganese		7.39	4.66	2.4		34.4	
Molybdenum		10.2	6.29	3.4	3.11	64.2	
Monel		8.83	6.02	2.7	1.96	53.2	
Nickel		8.88	5.63	3	2.64	50.0	
Nickel–silver		11.2	3.58	2.2		40.0	
Platinum		21.4	3.96	1.7		84.7	
Plutonium			1.79			28.2	
	1% gallium		1.82			28.6	
Potassium		0.83	1.82			1.51	
Rubidium		1.53	1.26			1.93	
Silver		10.5	3.6	1.6		37.8	
	Nickel	8.75	4.62	2.3	1.69	40.4	
	Germanium	8.7	4.76			41.4	

(continued)

APPENDIX: Ultrasonic Properties of Common Materials

Material	Comments	Density (g/cm ³)	V _l (km/s)	V _s (km/s)	Impedance (MRayl)	Attenuation (dB/cm/MHz)	Attenuation (dB/cm at 5 MHz)
Steel	302 Cres 347 Cres 410 Cres 1020 1095 4150	8.03 7.91 7.67 7.71 7.8 7.84	5.66 5.74 7.39 5.89 5.9 5.86	3.1 3.1 3 3.2 3.2 2.8	3.12 — 2.16	45.4 45.4 56.7 45.4 51.0 45.9	
		7.82 7.81 7.8 7.8 7.89	5.89 5.87 5.82 5.85 5.9	3.2 3.2 2.8 3.2 3.1		46.1 45.8 45.4 51.0 46.00	
	4340 Mild Stainless 347	16.6 11.9 11.3	5.79 4.1 1.62	2.9		45.70 54.8 19.3	
Tantalum				1.6		33.2	
Thallium			2.4				
Thorium			3.3	1.7		24.1	
Tin			6.07	3.1		27.3	
Titanium			6.1	3.1		27.3	
			5.18	2.9	2.65	99.7	
Tungsten			3.4	2		63.0	
Uranium			6	2.8		36.2	
Vanadium			4.17	2.4		29.6	
Zinc			4.72	2.4		44.2	
Zircalloy			4.65	2.3		30.1	
Zirconium							
<i>Solids (ceramics)</i>							
Ammonium	502/118.9 : 1	1.35	2.73	—	3.69		
dihydrogen							
phosphate (ADP)	502/118.5 : 1	1.35	2.67 3.28	—	3.60		

APPENDIX: Ultrasonic Properties of Common Materials

Material	Comments	Density (g/cm ³)	V _l (km/s)	V _s (km/s)	Impedance (MRayl)	Attenuation (dB/cm/MHz)	Attenuation (dB/cm at 5 MHz)
Arsenic trisulfide		3.2	2.58	1.4			8.25
Barium titanate		5.55	5.64	2.9			33.5
Boron carbide		2.4	11				26.4
Brick		1.7	4.3				7.40
		3.6	3.65	2.6			15.3
Calcium fluoride	CaFl. X-cut		6.74				
Clay rock		2.5	3.48	3.4			14.2
Concrete		2.6	3.1				8.00
Flint		3.6	4.26	3			18.9
Glass	Crown	2.24	5.1	2.8			11.4
	205 Sheet	2.49	5.66				14.1
	FK3	2.26	4.91	2.9			11.1
	FK6	2.28	4.43	2.5			10.1
	Flint	3.6	4.5				16.0
	Macor	2.54	5.51				14.0
	Plate	2.75	5.71				10.7
	Pyrex	2.24	5.64	3.3			13.1
	Quartz	2.2	5.57	3.4			14.5
	Silica	2.2	5.9				13.0
	Soda lime	2.24	6				13.4
	T1K	2.38	4.38				10.5
	Window		6.79	3.4			
Glass crown	Reg.	2.6	5.66	3.5			14.5
Granite		4.1	6.5				26.8
Graphite	Pyrolytic	1.46	4.6				6.60
	Pressed	1.8	2.4				4.10
Hydrogen	Solid at 4.2 K	0.089	2.19				0.19
Ice		0.92	3.6				3.20
		2.65	3.99	3.3			16.4

(continued)

APPENDIX: Ultrasonic Properties of Common Materials

Material	Comments	Density (g/cm ³)	V _l (km/s)	V _s (km/s)	Impedance (MRayl)	Attenuation (dB/cm/MHz)	Attenuation (dB/cm at 5 MHz)
Ivory		2.17	3.01				10.4
Leadmeta niobate	PbNbO ₃	6.2	3.3				20.5
	K-81	6.2	3.3				20.5
	K-83	4.3	5.33				22.9
	K-85	5.5	3.35				18.4
Lead zirconate titanate	PbZrTiO ₃	7.75	3.28				29.3
		7.5	4				30.0
		7.45	4.2				31.3
Lithium niobate		7.43	4.44				33.0
		7.95	4.72				37.5
	46 Rot. Y-cut	4.7	7.08				33.0
	Z-cut	4.64	7.33				34.0
	Y-cut		6.88				
Lithium sulfate	Y-cut	2.06	5.46				11.2
Marble		2.8	3.8				10.5
Porcelain		2.3	5.9				13.50
Potassium bromide			3.38				
Potassium chloride			4.14				
Potassium sodium niobate		4.46	6.94				31.0
PZT-2		7.6	4.41	1.7			31.3
PZT-4		7.5	4.6	1.9			34.5
PZT-5A		7.75	4.35	1.7			33.7
PZT-5H		7.5	4.56	1.8			34.2
Quartz		6.82	5.66				15.2
Salt	X-cut	2.65	5.75				15.3
	NaCl	2.17	4.85				10.5

APPENDIX: Ultrasonic Properties of Common Materials

Material	Comments	Density (g/cm ³)	V _l (km/s)	V _s (km/s)	Impedance (MRayl)	Attenuation (dB/cm/MHz)	Attenuation (dB/cm at 5 MHz)
Salt, rochelle		2.2	5.36	3.8		13.1	
Salt, rock	KNaC ₄ H ₄ O ₆		2.47				
Sapphire	Xdir	2.6	4.78			11.7	
	Al ₂ O ₃	3.98	9.8			44.5	
Silica, fused		2.2	11.2			13.1	
Silicon	Anisotropic	2.33	5.96	3.8		21.0	
Silicon carbide		13.8	6.66			91.8	
Silicon nitride		3.27	11	6.3		36.0	
Slate			4.5				
		3	4.5			13.5	
Sodium bismuth titanate		6.5	4.06			26.4	
Sodium bromide			2.79				
Sulfur	NaBr		1.35				
Titanium carbide		5.15	8.27	5.2		42.6	
Tourmaline	Z-cut	3.1	7.54			23.4	
Uranium oxide	UO ₂		5.18			56.7	
Zinc oxide		5.68	6.4	3		36.4	
<i>Solids (polymer)</i>							
ABS	Acrylonitrile	1.04	2.11	—	2.20		
Acrylic		1.2	2.7	—	3.24		
Acrylic resin		1.18	2.67	1.1		3.15	
Araldite	502/956	1.16	2.62	—	4.04		
Bakelite		1.4	2.59			3.63	
		1.9	1.9			4.80	
Butyl rubber		1.11	1.8			2.00	
Carbon, pyrolytic	Soft	2.21	3.31			7.31	

(continued)

APPENDIX: Ultrasonic Properties of Common Materials

Material	Comments	Density (g/cm ³)	V _l (km/s)	V _s (km/s)	Impedance (MRayl)	Attenuation (dB/cm/MHz)	Attenuation (dB/cm at 5 MHz)
Carbon, vitreous		1.47	4.26	2.7		6.26	
Celcon	Acetal copolymer	1.41	2.51			3.54	
Cellulose acetate		1.3	2.45			3.19	
Cyclac	Acrylonitrile- butadiene-styrene	2.27	2.27			2.49	
Delrin	Acetal	1.36	2.47			3.36	
	homopolymer	1.42	2.52			3.57	
DER317	10.5PHR DEH20	1.18	2.75			3.25	
	13.5PHR MPDA	2.23	2.07			4.61	
		1.6	2.4			3.84	
		2.03	2.19			4.44	
		3.4	1.86	0.9		6.40	
	9PHR DEH20	7.27	1.5	—		10.9	
		2.23	2.03	1		4.53	
		2.37	1.93	—		4.58	
DER332	10PHR DEH20	1.76	3.18	1.6		5.58	
		1.2	2.6			3.11	
	10.5PHR DEH20	1.29	2.65			3.41	
		1.26	2.61			3.29	
		1.37	2.75			3.78	
	11PHR DEH20	1.72	2.35			4.05	
		1.29	2.71			3.49	
	14PHR MPDA	1.25	2.59			3.24	
	15PHR MPDA	1.54	2.78	1.5		4.27	

APPENDIX: Ultrasonic Properties of Common Materials

Material	Comments	Density (g/cm ³)	V _l (km/s)	V _s (km/s)	Impedance (MRayl)	Attenuation (dB/cm/MHz)	Attenuation (dB/cm at 5 MHz)
		1.49	2.8	1.4		4.18	
		1.24	2.66	—		3.30	
		1.24	2.55	1.2		3.16	
		2.15	3.75			8.06	
		2.24	3.9			8.74	
		6.45	1.75			11.3	
	64PHR V140	1.13	2.36			2.65	
	75PHR V140	1.12	2.35			2.62	
	100PHR V140	1.1	2.32			2.55	
		1.13	2.27			2.55	
		1.16	2.36			2.74	
ECHOGEL 1265	100PHA of B	9.19	1.32			12.2	
		1.4	1.7			2.38	
		1.1	1.71			1.90	
EPON 828	MPDA	1.21	2.83	1.2		3.40	
EPOTEK 301		1.08	2.64			2.85	
EPOTEK 330		1.14	2.57			2.94	
EPOTEK H70S		1.68	2.91			4.88	
EPOTEK V6	10PHA of B	1.23	2.55			3.14	
		1.23	2.61			3.21	
		1.26	2.55			3.22	
		1.25	2.6			3.25	
Epoxxy	Silver	3.098	1.89			5.85	
		3.383	1.87			6.31	
EPX-1 or EPX-2	100PHA of B	1.1	2.44			2.68	
Ethyl vinyl acetate		0.94	1.8			1.69	

(continued)

APPENDIX: Ultrasonic Properties of Common Materials

Material	Comments	Density (g/cm ³)	V _l (km/s)	V _s (km/s)	Impedance (MRayl)	Attenuation (dB/cm/MHz)	Attenuation (dB/cm at 5 MHz)
Glucose		0.95	1.68			1.60	
		0.93	1.86			1.72	
Hysol	C8-4143/3404	1.56	3.2			5.00	
		1.58	2.85			4.52	
	C9-4183/3561	3.17	2.16			7.04	
		2.14	2.49			5.33	
		1.8	2.62			4.70	
		1.48	2.92			4.30	
		2.66	2.3			6.10	
		1.68	2.02			3.39	
		1.5	2.32			3.49	
Hysol							
Ivory	R9-2039/3404		3.01				
Kel-F			1.79				
Kydex		1.35	2.22			2.99	
Lucite	Polymethyl- acrylate	1.29	2.72			3.50	
Marlex 5003	High-density polyethylene	1.18	2.68	1.3		3.16	
		1.15	2.7	1.1		3.10	
		0.95	2.56			2.43	
Melopas		1.7	2.9			4.93	
Micarta	Linen base		3				
Mylar		1.18	2.54			3.00	
Neoprene		1.31	1.6			2.10	
Noryl	Polyphenyl- ene oxide	1.08	2.27			2.45	
Nylon 6-6		1.12	2.6	1.1		2.90	

APPENDIX: Ultrasonic Properties of Common Materials

Material	Comments	Density (g/cm ³)	V _l (km/s)	V _s (km/s)	Impedance (MRayl)	Attenuation (dB/cm/MHz)	Attenuation (dB/cm at 5 MHz)
Penton	Chlorinated polyether Syntactic foam (33 lb/ft ³)	1.4	2.57			3.60	
Phenolic							
Plexiglas	UVA UVAII	1.34 1.27 1.18	1.42 2.76 2.73	1.4		1.90 3.51 3.22	
Polyamide			2.6			2.90	
Polycarbonate	Lexan	1.18	2.3			2.71	
Polyester	Casting resin	1.07	2.29			2.86	
Polyethylene	Low density	0.92	2.06	—	1.90	22	26.5
		1.1	2.67			2.80	
	TCI		1.6				
Polyisobutylene	HD, LB-861	0.96	2.43			2.33	
			1.49				
Polypropylene	Mol. wt. 200 Profax 6423	0.901 0.88	1.85 2.49 2.74			2.24 2.40	
Polysulfone		1.24	2.24			2.78	
Polystyrene		1.1	2.67			2.80	
		1.05	2.4			2.52	
Polyurethane	Styron 666 RP-6400 RP-6401	1.04 1.07	1.5 1.71	—	1.83	1.56 35	73
		1.07	1.63			1.74	
	RP-6402	1.08	1.77			1.91	
	RP-6403	1.1	1.87			2.05	
	RP-6405	1.3	2.09			2.36	
	RP-6410	1.04	1.71	—	1.78	36	73

(continued)

APPENDIX: Ultrasonic Properties of Common Materials

Material	Comments	Density (g/cm ³)	V _l (km/s)	V _s (km/s)	Impedance (MRayl)	Attenuation (dB/cm/MHz)	Attenuation (dB/cm at 5 MHz)
Polyvinyl chloride (PVC) Polyvinylbutyral Polyvinylidene difluoride Profax Refrasil Rubber Scotchcast Scotchply XP241	RP-6413	1.04	1.33			1.38	
		1.04	1.71		1.78	21	35.2
	RP-6414	1.04	1.65			1.66	
		1.05	1.78			1.86	
	RP-6422	1.05	1.85		1.94	18	35.2
	EN-9	1.04	1.6			1.66	
	REN plastic	1.01	1.68			1.70	
		1.07	1.71			1.83	35
		1.04	1.49			1.55	36
		1.04	1.62			1.69	15
		1.04	1.71			1.78	21
		1.05	1.85			1.92	18
	RP6422	1.04	1.62		1.69	14	27.6
		1.45	2.27			3.31	
	Butracite	1.11	2.35			2.60	
		1.79	2.3			4.20	
Polypropylene BFG#6063-19-71 BFG#35080 Hard Rho-C Soft XR2535 Syntactic foam (42 lb/ft ³)	Polypropylene	0.9	2.79			2.51	
		1.73	3.75			6.49	
		0.97	1.53			1.56	
	BFG#6063-19-71						
	BFG#35080						
	Hard	1.1	1.45			2.64	
	Rho-C	1	1.55			1.55	
	Soft	0.95	0.07			1.00	
	XR2535	1.49	2.48			3.70	
	Syntactic foam	0.65	2.84			1.84	
	(42 lb/ft ³)						
		0.61	2.81			1.71	

APPENDIX: Ultrasonic Properties of Common Materials

Material	Comments	Density (g/cm ³)	V _l (km/s)	V _s (km/s)	Impedance (MRayl)	Attenuation (dB/cm/MHz)	Attenuation (dB/cm at 5 MHz)
Scotchply	Syntactic foam (38 lb/ft ³)						
	XP241	0.65	2.84				1.84
Scotch tape	SP1002	1.94	3.25				6.24
	2.5 mils thick	1.16	1.9				2.08
Silicon rubber	Sylgard 170	1.38	0.97				1.34
	Sylgard 182	1.05	1.03				1.07
		1.12	1.03				1.15
	Sylgard 184	1.03	1.03				1.04
	RTV-11	1.18	1.05				1.24
	RTV-21	1.31	1.01				1.32
	RTV-30	1.45	0.97				1.41
	RTV-41	1.31	1.01				1.32
	RTV-60	1.47	0.96				1.41
	RTV-77	1.33	1.02				1.36
	RTV-90	1.5	0.96				1.44
	RTV-112	1.05	0.94				0.99
	RTV-511	1.18	1.11				1.31
	RTV-116	1.1	1.02				1.12
	RTV-118	1.04	1.03				1.07
	RTV-577	1.35	1.08				1.46
	RTV-560	1.42	1.03				1.46
	RTV-602	1.02	1.16				1.18
	RTV-615	1.02	1.08				1.10
	RTV-616	1.22	1.06				1.29
	RTV-630	1.24	1.05				1.30
	PRC 1933-2	1.48	0.95				1.40

(continued)

APPENDIX: Ultrasonic Properties of Common Materials

Material	Comments	Density (g/cm ³)	V _l (km/s)	V _s (km/s)	Impedance (MRayl)	Attenuation (dB/cm/MHz)	Attenuation (dB/cm at 5 MHz)
Silly Putty		1	1			1.00	
Sty cast	1251-40	1.67	2.9	1.5		4.83	
		1.63	2.95			4.82	
		1.57	2.88			4.53	
		1.5	2.77			4.16	
	1264	1.19	2.22			2.64	
	2741	1.17	2.29			2.68	
	CPC-41	1.01	1.52			1.54	
	CPC-39	1.06	1.53			1.63	
Styrene 50D	Polystyrene	1.04	2.33			2.43	
Styron	Modified	1.03	2.24			2.31	
	polystyrene						
Surlyn	1555 Ionomer	0.95	1.91			1.81	
Tapox	Epoxy	1.11	2.48			2.76	
Techform	EA700	1.2	2.63			3.14	
Teflon		2.14	1.39			2.97	
		2.2	1.35			2.97	
TPX	DX845	0.83	2.22		1.84	4.2	5.8
Tracon	2135 D	1.03	2.45			1.52	
	2143 D	1.05	2.37			2.50	
	2162 D	1.19	2.02			2.41	
	3011	1.2	2.12			2.54	
	401 ST	1.62	2.97			4.82	
Uvex			2.11				
WR 106-1	Fluoro elastomer		0.87				
Zytel-101	Nylon-101	1.14	2.71			3.08	

APPENDIX: Ultrasonic Properties of Common Materials

Material	Comments	Density (g/cm ³)	V _l (km/s)	V _s (km/s)	Impedance (MRayl)	Attenuation (dB/cm/MHz)	Attenuation (dB/cm at 5 MHz)
<i>Solids (natural)</i>							
Ash	Along fiber		4.67				
Beech	Along fiber		3.34				
Beef			1.55			1.68	
Brain			1.49			1.55	
Cork			0.5				
Douglas Fir	Cross grain		1.4				
	With grain		4.8				
Elm			1.4			0.798	
Human			1.47			1.58	
Kidney			1.54			1.62	
Liver			1.54			1.65	
Maple			4.11				
Oak	Along fiber		4.47			3.60	
Pine	Along fiber		3.32				
Poplar	Along fiber		4.28				
Spleen			1.5			1.60	
Sycamore	Along fiber		4.46				
Water		0.88	4	2		3.50	
Wood	Cork	0.24	0.5			0.12	
	Elm		4.1				
	Oak	0.72	4			1.57	
	Pine	0.45	3.5			1.57	

REFERENCES

- Altergott W, Henneke E, editors. *Characterization of Advanced Materials*. New York: Plenum; 1990.
- Ash EA, Paige EGS. *Rayleigh Wave Theory and Application*, Springer Series on Wave Phenomena, Vols. 1 and 2. Berlin: Springer-Verlag; 1985.
- ASM International. *Metals Handbook*, 3rd ed., Vol. 17: *Nondestructive Evaluation and Quality Control*, Metals Park, OH: ASM International; 1989.
- Birks AS, Green J. *Ultrasonic Testing*, 2nd ed., Vol. 7 of *Nondestructive Testing Handbook*, P. Intire, editor Columbus, OH: American Society for Nondestructive Testing; 1991.
- Blitz J. *Electrical and Magnetic Method of Non-Destructive Testing*. London: Chapman & Hall; 1997.
- Boogaard J, van Dijk GM, editor. *Nondestructive Testing*. Proceedings of the 12th World Conference on Nondestructive Testing; Elsevier Science, New York; 1989.
- Bossi RH, Iddings FA, Wheeler GC. *Radiographic Testing*, 3rd ed., Vol. 4 of *Nondestructive Testing Handbook*, P. Moore (Ed.), Columbus, OH: American Society for Nondestructive Testing; 2002.
- Bray DE, Stanley RK. *Nondestructive Evaluation, A Tool for Design, Manufacturing, and Service*. New York: McGraw-Hill; 1989.
- Burger H. *Neutron Radiography; Methods, Capabilities and Applications*. New York: Elsevier Science; 1965.
- Fassbender RH, Hagmaier DJ. Low-Kilovoltage Radiography of Composites. *Materials Evaluation* 1983; 41(7):381–838.
- Förster, F., Theoretische und experimentelle Grundlagen der zerstörungsfreien Werkstoffprüfung mit Wirbelstromverfahren, I. Das Tastpulverfahren. *Zeitschrift Fur Metallkunde* 1952;43:163–171.
- Geier MH. *Quality Handbook for Composite Materials*. London: Chapman & Hall; 1994.
- Green RE. Jr., editor. *Nondestructive characterization of materials*. Vol. 8, International Symposium on Nondestructive Characterization of Materials, Plenum, New York, 1998.
- Hagemier DJ. Bonded Joints and Nondestructive Testing–1. *Nondestructive Testing* 1971;4(12): 401–406.
- Hagemier DJ. Bonded Joints and Nondestructive Testing–2. *Nondestructive Testing* 1972a;5(2): 38–47.
- Hagemier DJ. Nondestructive Testing of Bonded Metal-to-Metal Joints–2. *Nondestructive Testing* 1972b;5(6):144–153.
- Halmshaw R. *Nondestructive Testing Handbook*. London: Chapman & Hall; 1991.
- Hsu DK, Patton TC. Development of ultrasonic inspection for adhesive bonds in aging aircraft. *Materials Evaluation* 1993;51(12):1390–1397.
- Jones TS. Inspection of composites using the automated ultrasonic scanning system (AUSS). *Materials Evaluation* 1985;43(5):746–753.
- Jones RB, Stone DEW. Toward an ultrasonic-attenuation technique to measure void content in carbon-fibre composites. *Nondestructive Testing* 1976;9(3):71–79.
- Kline RA. *Nondestructive Characterization of Materials*. Lancaster (PA): Technomic Publishing; 1992.
- Krautkramer J, Krautkramer H. *Ultrasonic Testing of Materials*. 3rd ed. New York: Springer-Verlag; 1983.
- Krishnadas Nair, C. G. editor. *Trends in NDE Science and Technology, Proceedings of the 14th World Conferences on Nondestructive Testing*. Brookfield (VT): Ashgate Publishing; 1997.

- Libby HL. *Introduction to Electromagnetic Nondestructive Test Methods*. New York: Wiley-Interscience; 1971.
- Maldague XPV. *Infrared and Thermal Testing*, 3rd ed., Vol. 3 of *Nondestructive Testing Handbook*, P. Moore (Ed.), Columbus (OH): American Society for Nondestructive Testing; 2001.
- Mallick PK. Nondestructive Tests, in *Composites Engineering Handbook*, P. K. Mallick (Ed.), New York: Marcel Dekker; 1997.
- McGonnagle W. *Nondestructive Testing*. New York: Gordon Breach; 1961.
- Mitchell MR, Buck O. editors. *Cyclic Deformation, Fracture, and Nondestructive Evaluation of Advanced Materials*. Philadelphia (PA): American Society for Testing and Materials; 1992.
- Papadakis EP, Chapman II GB. Modification of a Commercial Ultrasonic Bond Tester for Quantitative Measurements in Sheet-Molding Compound Lap Joints. *Materials Evaluation* 1993;51(4):496–500.
- Quinn RA. *Industrial Radiography—Theory and Practice*. Rochester (NY): Eastman Kodak; 1980.
- Rose JL, Tseng AA. editors. *New Directions in Nondestructive Evaluation of Advanced Materials*. New York: American Society of Mechanical Engineers; 1988.
- Ruud CO, et al. editors. *Nondestructive Characterization of Materials*. Vols. I–IV. New York: Plenum; 1986.
- Schmidt JT, Skeie K. *Magnetic Particle Testing*, 2nd ed., Vol. 6 of *Nondestructive Testing Handbook*, P. McIntire (Ed.), Columbus (OH): American Society for Nondestructive Testing; 2001.
- Shapuk HJ. editor. *Annual Book of ASTM Standards: E-7, Nondestructive Testing*. West Conshohocken (PA): American Society for Testing and Materials; 1997.
- Sharpe RS. *Research Techniques in Nondestructive Testing*. New York: Academic; 1984.
- Stanley RK. editor. *Special Nondestructive Testing Methods*, 2nd ed., Vol. 9 of *Nondestructive Testing Handbook*, P. O. Moore and P. McIntire (Eds.), Columbus (OH): American Society for Nondestructive Testing; 1995.
- Summerscales, J. Manufacturing defects in fibre-reinforced plastic composites. *Insight* 1994;36(12):936–942.
- Swamy RN, Ali AMAH. Assessment of in situ concrete strength by various non-destructive tests. *NDT International* 1984;17(3):139–146.
- Thompson, DO, Chimenti, DE. editors. *Review of Progress in Quantitative Nondestructive Evaluation*. New York: Plenum; 1982–2000.
- Tracy N. editor. *Liquid Penetrant Testing*, 3rd ed., Vol. 2 of *Nondestructive Testing Handbook*, P. Moore (Ed.), Columbus (OH) American Society for Nondestructive Testing 1999.
- Udpa SS. editor. *Electromagnetic Testing*, 3rd ed., Vol. 5 of *Nondestructive Testing Handbook*, P. Moore (Ed.), Columbus (OH): American Society for Nondestructive Testing; 2004.
- Vernon, S. N., Parametric eddy current defect depth model and its application to graphite epoxy. *NDT International* 1989;22(3):139–148.
- Viktorov IA. *Rayleigh and Lamb Waves*. New York: Plenum; 1967.
- Wincheski RA. et al., Development of Giant Magnetoresistive Inspection System for Detection of Deep Fatigue Cracks under Airframe Fasteners. in *Review of Progress in Quantitative Nondestructive Evaluation* 2002.
- Zorc TB. editor. *Nondestructive Evaluation and Quality Control*, 9th ed., Vol. 17 of *Metals Handbook*, Metals Park (OH): ASM International; 1989.

TESTING OF METALLIC MATERIALS

PETER C. McKEIGHAN

- 41.1 Mechanical test laboratory
 - 41.1.1 Test machines
 - 41.1.2 Sensors and instrumentation
- 41.2 Tensile and compressive property testing
- 41.3 Creep and stress relaxation testing
- 41.4 Hardness and impact testing
- 41.5 Fracture toughness testing
- 41.6 Fatigue testing
- 41.7 Other mechanical testing
- 41.8 Environmental considerations
- Acknowledgments
- References

One of the daunting challenges that a designer faces when making a material selection is knowing what properties are the most critical for a given application. Assuming that the appropriate properties can be chosen, the designer next needs to be able to find the properties for each metallic material of interest (a challenge in itself) and then interpret the values relative to the design requirements. This requires that the designer has rudimentary knowledge of both the properties and test methods as well as an understanding of how to interpret the properties in the design context.

The purpose of this chapter is to provide some of the basic background for how metallic materials are mechanically tested. The goal is not to provide enormous detail for each different type of property; other sources, for instance Kuhn (2000), can provide this. Rather, the intent is to provide sufficient basic information so that the reader understands what the purpose of the mechanical test is, what mechanical property typically results, and what the corresponding design implication is.

The complexity of metallic materials testing and the specific technical language required can be confounding, especially to engineers not intimately familiar with the field. Those of us in the field also can confuse each other as new methods, terminologies, and approaches develop and are adopted. Luckily, there exists an organization whose mission in part is to minimize this type of misunderstanding.

The American Society for Testing and Materials (ASTM) is one of the world's largest standards development organizations with over 34,000 members responsible for more than 10,000 standards. The breadth of these standards is enormous (74 volumes), covering not only test/evaluation procedures but also standardization approaches for designating metals, paints, plastics, and the like. Originally developed in 1898, this independent, not-for-profit organization serves as a forum for producers, users, consumers, and others in developing voluntary, consensus standards. In practice, a hierarchical committee organization develops the standards initially with a group of experts providing a draft standard. This draft is subsequently balloted by ASTM and does not become a standard until all concerns raised during the voting process that have been fully addressed.

The test-related standards that ASTM publishes help to ensure that mechanical properties generated in one laboratory are consistent with those generated in another laboratory. This consistency is critical from a design point of view as specific property values are typically extracted from numerous sources, and knowing that consistent methods were used in generating these properties is subsequently crucial. The majority of the ASTM standards related to metals test methods and discussed in this chapter are contained within ASTM (2000).

Knowing where to look for the standard does not necessarily imply that it is easy to know which property is the most germane for a given design. For instance, under the classification "fracture testing," there are 16 separate standards in ASTM (2000), all very different and evaluating disparate quantities such as crack-tip opening displacement (CTOD), plane-strain fracture toughness, surface crack toughness, dynamic tear properties, and the like. The fundamental characteristics of the design scenario involved and the candidate metallic materials considered will dictate which fracture property is suitable. In a case such as this where many different test methods are available, selection of the appropriate test standard is complex and beyond the scope of this elementary treatise.

Finally, ASTM also provides other resources to ensure that the properties required are available. In particular, the Directory of Testing Labs (ASTM, 2001) provides a state-by-state listing of laboratories that are available to perform materials testing. The laboratories are identified by the type of test performed, with an index cross-referenced to specific test methods. Moreover, a brief synopsis of the laboratory, including scope of testing, facilities, and staff, is also provided to allow differentiating between the different available choices.

41.1 MECHANICAL TEST LABORATORY

During mechanical testing, a component or material is loaded with either displacement or force, causing a deformation and associated mechanical response to occur. The loading occurs at a design relevant rate (impact, static, or cyclic), and the test is typically performed in the most suitable environment for the end use of the component. The two key components in this process that make up the primary constituents in a mechanical test laboratory (Figure 41.1) are the test machine required to apply the specified loading as well as the sensors and instrumentation used to measure the behavior of the test article.



FIGURE 41.1 Typical material test laboratory containing numerous servohydraulic test frames (Courtesy of Fracture Technology Associates).

41.1.1 Test Machines

The test machine, or loading frame, is used to apply stress (or strain) resulting in axial tension or compression, bending, shear, torsion, or pressure. Although a variety of different loading systems are available, the most common is the universal testing machine, which allows all of these different types of loading types by use of different test fixtures and load train components. Load is applied through a hydraulic piston and cylinder (servohydraulic machine) or with precision machine screws (screw machine) driven by an electric motor and a gear box system. Although screw machines have lower load capacities than servohydraulic systems (generally 125 kips or less), their displacement stability, accuracy, and precision is better than typically observed with servohydraulic systems. Nevertheless, the advantage of the more common servohydraulic systems is the rapid rate of response necessary for many types of tests (e.g., any type of cyclic testing, including fatigue and fracture applications).

For the case of servohydraulic machines, the actuators are typically double-ended allowing tension and compression of the load train. The displacement rate of these machines range from nearly static to approximately 350–1000 in./s. The actual highest rate possible for a given machine/specimen combination is difficult to assess as it depends upon the actuator size, flow rating of the servovalve, as well as the supply pump, plumbing configuration, and control system parameters. Servohydraulic test machines range in size but are typically available from 1–1000 kips or more.

The control of screw and servohydraulic machines is accomplished through dedicated microprocessor-based electronics. Controllers generally fall into two categories: the older generation (but stable and cost effective) analog controller and the newer digital controller (Figure 41.2). In the strictest sense, an analog controller needs a signal source to drive it, typically either a built-in function generator or an external computer for complex waveforms or more stringent control requirements. Whereas machine control in this case is with analog electronics, in the fully digital system closed-loop control is achieved with a computer running software designed to perform the test of interest. The digital systems are typically provided with a suite of programs designed to perform a wide variety of standard test protocols.



FIGURE 41.2 Servohydraulic test frame and digital controller (Courtesy of MTS Corporation).

A test machine is incomplete without a range of grips and fixtures suitable for inserting a test specimen into the load train of the frame. Although the majority of the ASTM test standards address fixtures and grips in some manner, inevitably the hardware available in a given laboratory is diverse and varied, consisting of both standard and hybrid fixtures whose functionality has been proven through repeated use. Versatile pneumatically or hydraulically actuated mechanical clamp grips constitute one of the more popular types available.

41.1.2 Sensors and Instrumentation

The sensors and instrumentation in a laboratory are designed to measure mechanical behavior. Perhaps, the most commonly used sensor is the load-measuring device, also called a load cell. Load cells are placed in series with the specimen in the loading train and can come in a wide variety of sizes and configurations. Normally, threaded couplings are available on either side of the load cell to attach fixtures. Inside the load cell is an instrumented transducer, typically using strain gauge technology to relate applied load to a resistance change suitable for sensing.

Although numerous nontraditional methods and systems exist for measuring displacement or strain, (Lucas, 1997, 2001) the most common method is based on a sensor that provides output based upon strain gauge technology. An example of this is the highly reliable modern material test extensometer (Figure 41.3). This device, fastened to the specimen by mechanical clamping, is used to accurately measure displacement between a

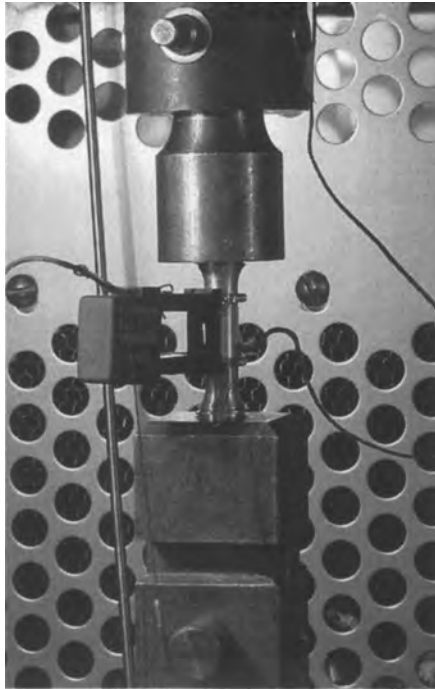


FIGURE 41.3 Extensometer (displacement measurement device) mounted on a tensile specimen.

pair of short knife edges affixed to the gauge length of the specimen. Another type of displacement gauge is the clip gauge, which is basically a spring-loaded extensometer with longer arms. These devices are extremely accurate and robust, but they do have environmental limitations to some extent due to their reliance on strain gauge technology.

Sometimes, it is also useful to mount strain gauges directly on the surface of the test specimen to infer localized behavior. Furthermore, linear variable differential transformers (LVDTs) and capacitive coupling-based displacement devices are commercially available for use in the most aggressive environments. In addition, noncontact optical systems (laser or charge coupled device, CCD, camera based) can also be used to obtain displacement measurements. Recent improvements in computer processor speed, CCD camera size, as well as more affordable random-access memory (RAM), and disk storage make these vision-based technology methods more desirable now more than ever before. The preponderance of optical systems and lack of any standardization in the core technology has resulted in recent ASTM standards activity and a draft standard that defines the relevant terminology.

Data acquisition tools have also benefited greatly from the recent advances in personal computers. Newer data acquisition systems tend to be more flexible than the older, device-dependent applications. The most popular platform for application building currently appears to be the LabVIEW (National Instruments, Austin, TX) programming environment. The LabVIEW platform provides a graphical, icon-based environment to build and modify applications that are more device/board independent and well suited to typical complex process and data acquisition requirements.

Some of the most complex tests performed in a mechanical test laboratory are fatigue and fracture evaluations of metallic components and materials. These tests usually require highly specialized, custom software to control the test machine, take the data, and analyze the results. It is always difficult to know *a priori*, without extensive experience performing these tests, what data acquisition parameters are the most critical. For this reason, a recent guide, ASTM E1942, has been developed that assists a user to understand the limitations of instrumentation and data acquisition systems, especially as related to cyclic fatigue and fracture mechanics testing. Furthermore, this standard is also useful for the segment of the test community not familiar with fatigue and fracture tests, as it provides a practical guide for understanding data acquisition performance.

41.2 TENSILE AND COMPRESSIVE PROPERTY TESTING

The tension test, defined for metallic materials in ASTM E8, is perhaps one of the simplest and most common tests of mechanical behavior. The test is conducted by gripping each end of a reduced section specimen and slowly pulling it until catastrophic failure occurs. Several sample specimen geometries (round and flat, threaded and smooth shank), conventionally described as “dogbone” specimens by virtue of their reduced section geometry, are shown in Figure 41.4. Moreover, tension test specimens are sometimes notched in the center of the gauge length as described in ASTM E338 and E602. These tests are useful for determining how the given metallic material behaves in the presence of a stress concentration. Furthermore, results from these tests have also been related to fracture mechanics parameters for a given material.

In a tensile (or compressive) test, the applied load and displacement (or strain) in the gauge section is used to develop a plot of the stress–strain behavior using the specimen geometry and calibration constants of the transducers used during the test. One challenge during tensile testing is measuring elongation to high strain levels while still having sufficient sensitivity in the elastic region. It is therefore not uncommon to utilize two strain

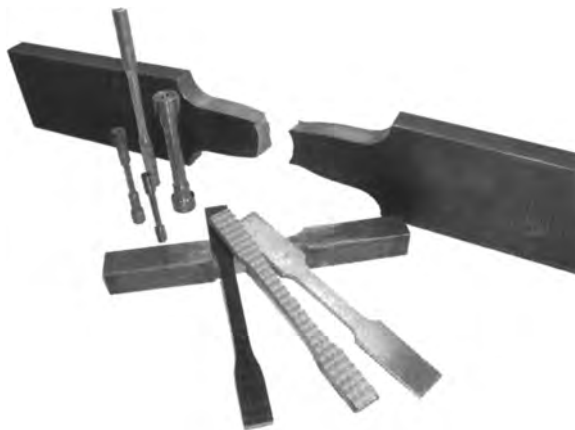


FIGURE 41.4 Several sample specimen geometries for tensile testing in accordance with ASTM E8 (Courtesy of the Instron Corporation).

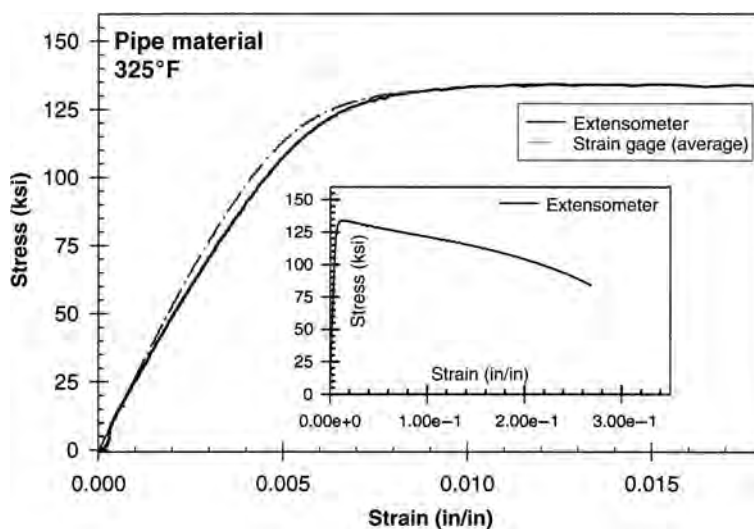


FIGURE 41.5 Example stress–strain data from a tensile test from a pipe material tested at 325°F.

transducers as shown in Figure 41.5. Special high-gain strain gauges or transducers are typically required to provide sufficient sensitivity and resolution in the elastic region to derive a Young's modulus value. This is contrasted with conventional extensometry, which is usually used to measure higher levels of strain, even 20% or higher. Alternately, two extensometers, one measuring behavior over a small range while the other configured for a large range, can be mounted on a single specimen. One advantage with this approach is that strain differences as a consequence of different gauge lengths, due to issues such as strain localization, is greatly reduced since each extensometer can have the same basic gauge length but be calibrated to different ranges.

During tensile testing, the ultimate strength is the largest applied stress measured in the specimen section. The proportional limit, often called the yield strength, indicates the onset of plasticity where the assumption of linear stress–strain behavior is violated. As there are many definitions of proportional limit (e.g., the conventional 0.2% strain offset method or the American Petroleum Institute (API) approach (API, 1998) using the stress corresponding to a specific strain level), it is imperative that the physical meaning of the limit be clearly defined when presenting the results. Moreover, post-test measurements made on the gauge length and fracture surface of the specimen are used to quantify the ductility of the specimen. Ductility is the ability of a material to deform plastically without failure occurring. Ductility parameters include percent elongation and percent reduction of area.

The mode of failure can also sometimes influence the strain results obtained. For instance, if the metallic material tested tends to exhibit highly localized plastic deformation, the region of confined necking results in either higher or lower effective strain if the gauge length is short or long, respectively. On the other hand, if deformation is equally distributed along the gauge section, the strain data exhibits no gauge length dependence. Obviously, the potential exists for misinterpretation when comparing stress–strain behavior from one sized specimen to another. Therefore, it is important to ensure that specimen

size is constant when performing a relative comparison of properties from one condition to another or between metallic materials.

Compression testing is typically performed on short cylindrical specimens (ASTM E9). Care must be taken during testing to minimize the influence of both buckling and barreling, two modes of behavior that can corrupt compression test results. Buckling is prevented by minimizing the ratio of specimen length to diameter (typically in the range of 1–2). Load train alignment is also a critical feature of these tests. Alignment is often practically achieved by loading the specimen through a spherical joint so that slight, nearly unmeasurable misalignment will not cause buckling. Barreling results from the friction between the end faces of the specimen and the loading platens. Friction is minimized by using lubricant and machining concentric rings on the end of the specimen to trap the lubricant and keep it from squeezing out.

41.3 CREEP AND STRESS RELAXATION TESTING

Creep and stress relaxation properties are critically important for high-temperature applications. Without knowledge of these properties, both air- and land-based jet turbines, internal combustion engines, and even some electronic equipment could not be adequately designed. Furthermore, there is always increasing pressure to develop new metallic materials with better performance at high temperature so that operating temperature in these applications can be increased to yield higher efficiencies.

Creep deformation is examined by applying either a constant load or a constant true stress to a material held at a specified temperature. When the material is first loaded, a small permanent loading strain occurs and the creep strain rate gradually decreases. This portion of the creep strain versus time plot is called primary creep (stage I). This is subsequently followed by secondary (or also steady-state or stage II) creep with the highest rates observed in the final tertiary creep (stage III) immediately before failure. These three regions of the creep curve are shown schematically in Figure 41.6. Whereas the test methods are described in detail in ASTM E139, the primary components of the test include a static load frame, furnace, and high-temperature extensometry.

The test specimens are similar to those used in a quasi-static tensile test, although the actual geometry used is often controlled by the metallic material form available. Static

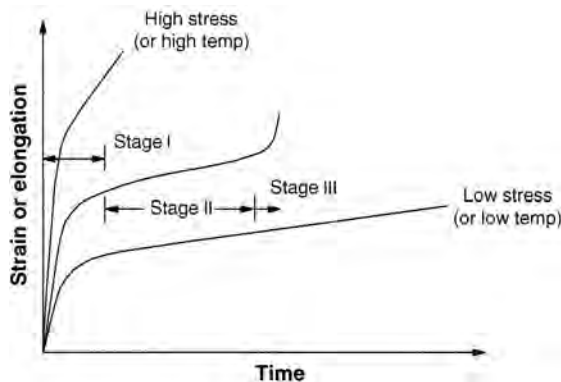


FIGURE 41.6 Schematic of the different regions of a classic creep curve.

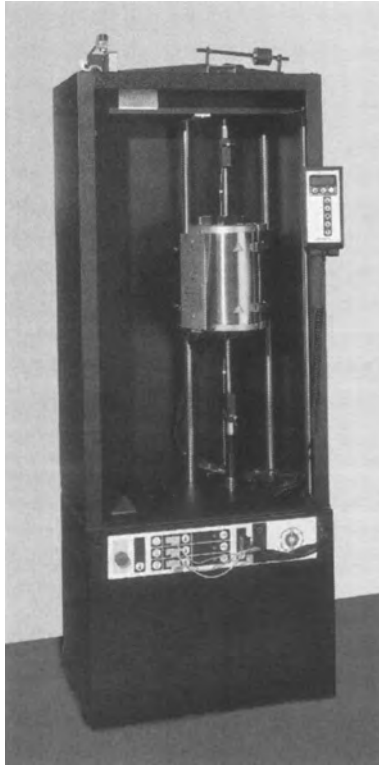


FIGURE 41.7 Loading frame to apply static (or extremely slow cyclic) load conditions (Courtesy of the Instron Corporation).

deadweight machines are typically used for this type of test, as shown in Figure 41.7. The frames usually have a balance beam to magnify the load applied to the specimen by a factor of between 2 and 25. All of the pull-rod and coupling fixtures used are manufactured from high-temperature alloys impervious to the environment used during the test. Tubular, electric resistance furnaces are usually used with the number of zones typically dependent upon the size of the specimen. Special attention must be made to ensure minimal thermal gradients and uniform temperatures applied to the specimen.

The measurement of gauge length deformation is typically critical during a creep experiment. If temperatures are sufficiently low ($<400^{\circ}\text{F}$), conventional extensometry as used in tensile tests can be used on the specimen. However, higher temperatures are usually involved for most metals, which necessitates using LVDT, capacitive, or other types of displacement measurement transducers. The most common transducer is the LVDT coupled with linkages that reach up through the furnace to attach to the specimen. Special care must subsequently be taken to ensure that the displacements measured actually correspond to gauge length deformation and are not biased by linkage or transducer effects.

Stress relaxation testing (ASTM E328) is similar in many ways to creep testing except that the applied boundary condition is a constant displacement (as opposed to a constant load). This type of test is more suitable for metallic materials that have significant constraint and hence exhibit little strain increase as a function of time. One example of this is

a high-temperature bolted joint where a displacement boundary condition is first applied and then the stress relaxes with time. Testing times include both long term (months or more) at the service temperature and accelerated tests (days) at more aggressive thermal conditions.

Creep and stress relaxation testing can also be performed on specimens with a more complex stress state than simple tension. Examples include compression testing (typically using a carefully machined cylindrical specimen similar to that described earlier for quasi-static compression testing), bend testing (e.g., with a cantilever beam specimen), or even torsion testing. Testing has also been performed on tubular specimens that are subject to complex, multiaxial stress states.

41.4 HARDNESS AND IMPACT TESTING

This section addresses two mechanical test methods that are used to indirectly determine the tensile and toughness properties of a metallic material. As with any indirect test method, there are well-known limitations to these techniques, but they are clearly offset by the reduced cost and increased efficiency of the indirect methods. The test methods, as well as their limitations, will be explained in more detail in this section.

Hardness refers to a material's ability to resist deformation when loaded by an indenter. The indenter can differ in geometry; for instance, it can be conical (the Rockwell test), spherical (Brinell), or pyramidal (Vickers and Knoop). For metallic materials, hardness is directly proportional to the yield stress of the material. Nevertheless, conversions are readily available to relate hardness to ultimate tensile strength (most commonly for steel although other conversions for different metallic materials do exist). Hardness testing (ASTM E6, E10, E18, E92, E1842, and others) is industrially attractive since the testing is rapid (<20 s), low cost, nondestructive, and generally does not require a machined specimen. As a consequence of these issues, hardness is often used as an industrial screening tool to ensure metallic material processing control. An ASTM standard (E140) exists to convert between the different superficial hardness and conventional hardness scales.

The majority of hardness testers (e.g., those shown in Figure 41.8) utilize deadweight, springs, or (infrequently) closed-loop computerized systems. Although many testers provide output on an analog meter, bench-top units with digital readouts (sometimes with an interface for a PC) are generally more popular today. When the component being hardness tested is too large to practically position onto the anvil of the tester, portable test units are available to locally measure hardness. A typical portable hardness measurer is shown in Figure 41.8b. Regardless of the type of hardness tester used, care must be taken to ensure that the system is in proper working order by using commercially available calibration blocks.

Although hardness testing is relatively simple, accurate measurements depend greatly on careful attention to detail. The anvil should minimize the contact area with the component. Surface finish can sometimes be of paramount concern; for 100–150 kgf, a finish ground surface is sufficient whereas for a 15-kgf load, a polished or lapped surface usually is required. Surfaces that are ridged due to machining or that exhibit loose, flaking scale can yield incorrect hardness results. The test surface should be perpendicular to the indenter since even slight ($<5^\circ$) angular deviation can cause hardness to shift by several points. Indenter size should be chosen with the underlying grain structure of the metallic material in mind. In general, the indenter should be larger than the grain size if a

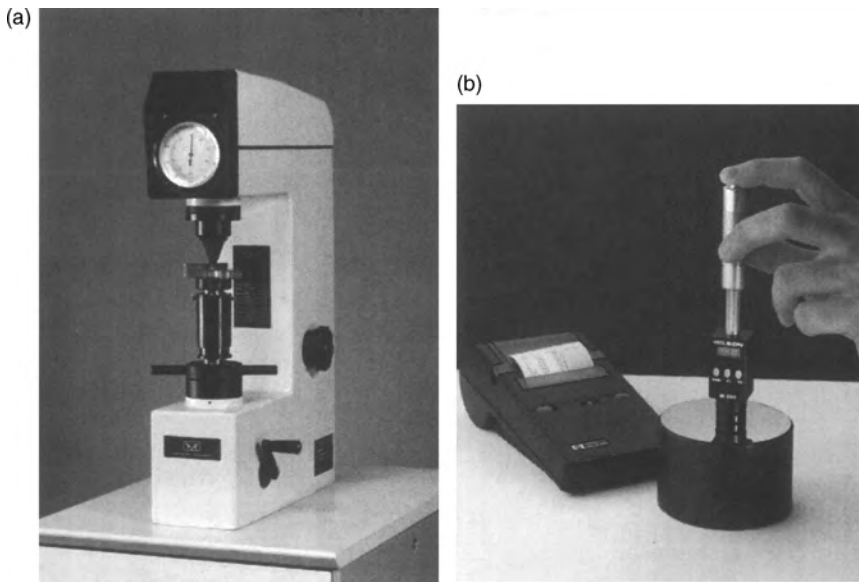


FIGURE 41.8 Hardness testers including (a) a benchtop and (b) portable unit (Courtesy of the Instron Corporation).

macroscopic measurement of hardness is desired. Finally, metallic material homogeneity is also important since the indenter effectively samples a region approximately 10 times deeper than the depth of the indenter penetration.

Instrumented indentation testing is a more recent method developed to derive more extensive data from effectively a hardness test. During this type of testing, the instantaneous load–displacement data is recorded continuously. The measured data can be related to such things as elastic modulus, yield strength, and strain hardening exponent. Highly empirical punch test methods are also continually evolving to relate local deformation behavior to properties such as fracture toughness.

Since some percentage of loading is inevitably dynamic and components often have to be designed to survive high loading rates typical during an accident, dynamic fracture testing can ensure structural integrity under this type of loading. During impact testing, a pendulum machine (example shown in Figure 41.9) with a striker is typically used to impart high energy to a notched sample. The notch toughness of the specimen, typically a flat bar with a carefully machined notch of various geometries, is measured in terms of the absorbed impact energy that causes fracture in the specimen. Some other parameters that are commonly examined include fracture appearance and the deformation noted on the surface of the specimen. The independent variable controlled during testing is most often temperature. A typical energy–temperature curve for A36 steel is shown in Figure 41.10 contrasting the typical difference between slow bend tests and rapid impact results.

The Charpy and Izod impact tests both utilize pendulum loading albeit with specimens of slightly different geometry, especially as related to the specimen notch dimensions. The primary difference between the two tests though is that the Charpy V-notch (CVN



FIGURE 41.9 Example of a pendulum impact machine (Courtesy of the Instron Corporation).

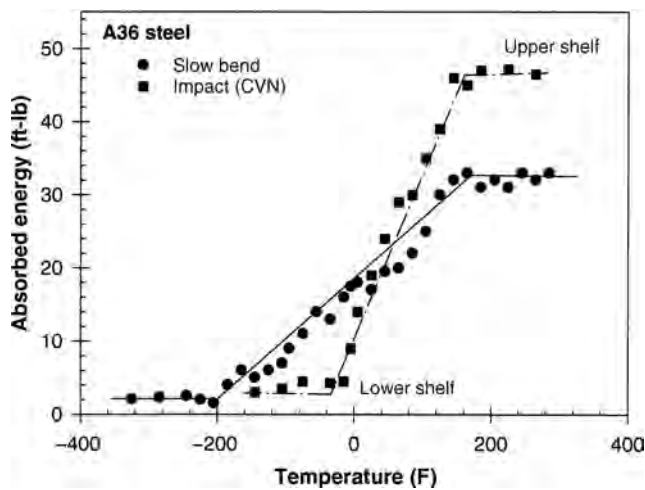


FIGURE 41.10 Impact test results illustrating CVN energy versus temperature curves for A36 steel.

test as defined in ASTM E23) specimen is three-point loaded whereas the Izod specimen is cantilever loaded.¹ Of the two methods, the CVN test remains the most popular method available, and manufacturing lot release requirements using CVN data are actually written into many metallic material specifications. The CVN test is rapid and low cost; the most expensive part is machining specimens although commercial machining templates to ease fabrication are readily available.

Two other impact tests include the drop-weight (ASTM E208) and the drop-weight tear test (DWTT as per E436). During drop-weight tests, a series of experiments are performed where different specimens at a range of temperatures are impacted with a guided, free-falling weight dropped from a fixed height. The key finding is the nil-ductility transition temperature (NDTT), which is the maximum temperature where fracture was observed. The DWTT test is similar to the drop-weight test although the specimen used is slightly larger and the test fixturing is very different. The fracture surfaces are evaluated and the percentage of shear quantified. The temperature at which 50% shear occurs is loosely defined as the ductile-to-brittle transition temperature.

CVN testing can also be performed with a fatigue precracked specimen to eliminate the notch acuity and depth restrictions inherent to the standard CVN specimen geometry. Considerable effort has also been recently expended developing instrumented impact testers to record more data. Nevertheless, the primary shortfall with most of the impact test methods is (a) the relatively small specimen size (relative to component size) and (b) the high degree of empiricism used to relate the results to metallic material properties. These disadvantages are offset by the simplicity and low cost of the test, particularly for production environments where the test can be used effectively as a quality control tool.

41.5 FRACTURE TOUGHNESS TESTING

The foregoing discussion introduced the concept of notch toughness, defined loosely as the energy required to initiate a running fracture from a machined notch. This is an important design parameter as nearly all structures contain notchlike stress concentrations. However, in the strictest sense according to the science of fracture mechanics (the study of cracks), fracture is broadly classified as either brittle [governed by linear elastic fracture mechanics concepts (LEFM)] or ductile [elastic plastic fracture mechanics (EPFM)]. Fracture toughness is the ability of a material to withstand unstable crack propagation (e.g., fracture) assuming of course the presence of a crack. Fracture mechanics provides the theoretical underpinning that relates the following quantities:

- magnitude and range of the remote applied stress,
- geometry of the component,
- morphology (size and shape) of the crack

to what is loosely termed the “crack driving force,” or stress–intensity factor K . Detailed stress analyses of cracks have shown that the magnitude of the local crack-tip stress field

¹ Both machines were developed in the beginning of the twentieth century although Charpy was a Frenchman and Izod was an Englishman.

is proportional to this K parameter. Although K applies for LEFM, in the case of EPFM, there exists another similitude parameter that is termed J .

Fracture toughness testing requires closed-loop, computer-controlled servohydraulic test equipment, and a suite of accurate displacement extensometers. Virtually, all tests are computer controlled due to the complexity of the test and the requirements for recorded data. As a minimum, most fracture tests require measuring the applied load and some displacement measurement across the crack. Most tests are analyzed by custom software developed specifically for the particular test. The most common fracture toughness test is the plane-strain fracture toughness measurement, which yields the K_{Ic} parameter as described in detail in ASTM E399. For an ASTM valid K_{Ic} value, the loading must be such that there is no significant plastic deformation (specifically, the plastic zone at the tip of the crack is less than 2% of the thickness and total crack length). The plane-strain requirement implies that the fracture toughness specimens tend to be quite thick, most are commonly greater than 0.5 in.

Although E399 allows a variety of specimens, the most commonly used is the compact-tension specimen, a pin-loaded, edge-notched geometry. The specimen is first pre-cracked under fatigue loading to ensure that the crack tip is sufficiently sharp for the fracture test. The specimen is then loaded to failure by a constant displacement ramp applied between the pins. During the test, the load and crack mouth opening displacement (CMOD) is measured. Two sample load-displacement plots are shown in Figure 41.11. In Figure 41.11a, the behavior is more brittle with catastrophic failure at peak load. Conversely, in Figure 41.11b, the trend is more suggestive of a ductile and somewhat more stable material with slow crack advance gradually relieving load as the test progresses past the point of maximum load.

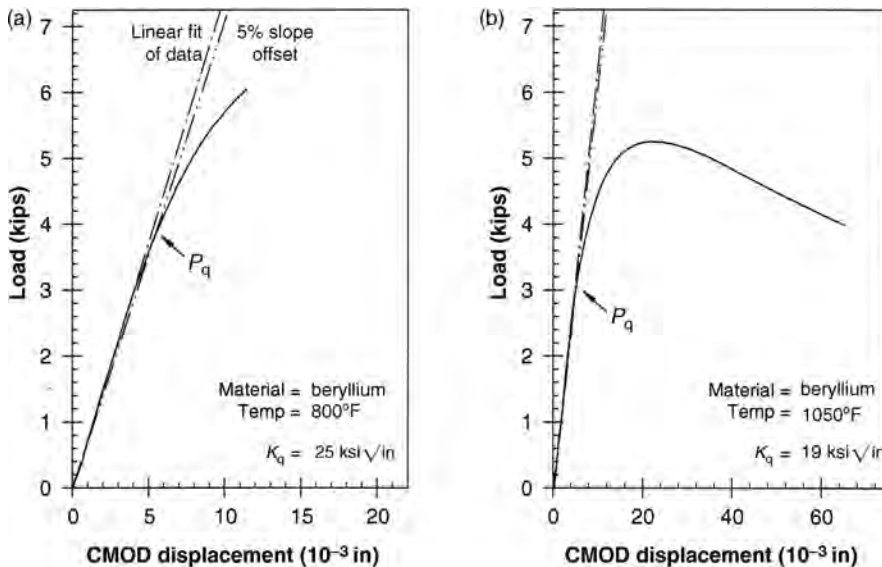


FIGURE 41.11 Two load-displacement plots from K_{Ic} fracture toughness tests conducted on beryllium at elevated temperature.

Determining a K_{Ic} toughness from the load–CMOD data requires fitting the initial linear portion of the data and plotting a 5% secant offset to that linear region. The intersection between the measured load–CMOD data and this secant line defines the critical load. This graphical construction is evident in Figure 41.11a. However, there are stringent limits for qualifying the conditional K_q toughness as an ASTM-valid K_{Ic} . Most importantly, the crack length and thickness are required to be sufficiently long to satisfy the small-scale yielding criteria described earlier. Furthermore, nonlinearity is minimized by requiring that the critical load (corresponding to K_q and denoted by P_q) must be no less than 90% of the maximum measured load. The two sets (Kuhn, 2000) of data indicated in Figure 41.11 are clearly invalid since the conditional P_q load is too low relative to the maximum measured load.

Whereas K_{Ic} applies for thicker, more constrained sections where catastrophic fracture occurs rapidly, a different K_C fracture test (ASTM E561, also termed a K – R curve) applies for thinner structure where slow stable crack extension precedes unstable fracture. The test specimen geometry typically used is a wide, center cracked panel precracked at low loads to generate a sharp fatigue crack suitable for fracture. The specimen is monotonically displaced similar to a K_{Ic} specimen while recording load, CMOD, and often some parameter such as electric potential drop that can be used to instantaneously provide crack length measurement.

The resistance curve, also called an R curve, is then generated by plotting crack driving force K against crack extension, termed Δa . Crack extension is defined as the sum of the physical crack length change, usually derived from either compliance (stiffness) or some other nonvisual crack length indicator, plus the half plastic zone size. Because the stress state is typically something more like plane stress, the plastic zones as observed on the surface of the specimen tend to be quite large. An example R curve is shown in Figure 41.12. The K_C value is defined as the K magnitude at the tangent intersection between the R curve and lines of constant load plotted in K versus crack length (or crack length extension) space. This load represents the maximum level that can be achieved before the crack extends (often unstably) and structural integrity is lost. A careful examination of the data in Figure 41.12 clearly illustrates that some crack growth was apparent in this metallic material since catastrophic failure did not occur at the point of instability.

Fracture toughness testing in the elastic–plastic regime is time consuming, not trivial, and requires advanced instrumentation to yield excellent results (Joyce, 1996). After conducting and analyzing the test, the goal is to generate an R curve, similar to that described earlier for the K_C testing, as shown in Figure 41.13. However, instead of K on the abscissa, the J parameter is included. The J parameter is calculated by including an elastic portion related to K combined with a nonlinear portion related to the area under the load–displacement curve (with respect to load-line displacement). The plane-strain, elastic plastic J_{Ic} parameter measures the onset of instability, in this case corresponding to the point where the crack has advanced a certain amount (after blunting) on the R curve (note this instability criteria is very different than that described for K_C). It is often the case in this type of test that slow, stable cracking will continue as the specimen is displaced and the load gradually drops to a small fraction of the peak obtained.

The J_{Ic} test procedure is usually performed using compact-tension specimens, although three-point bend specimens are also possible provided care is taken to ensure that accurate instrumentation is available to measure load-line displacement. The primary challenge with J_{Ic} testing is the high accuracy required from measurements of the load-line displacement. High-resolution measurement is required so as to infer the extremely small

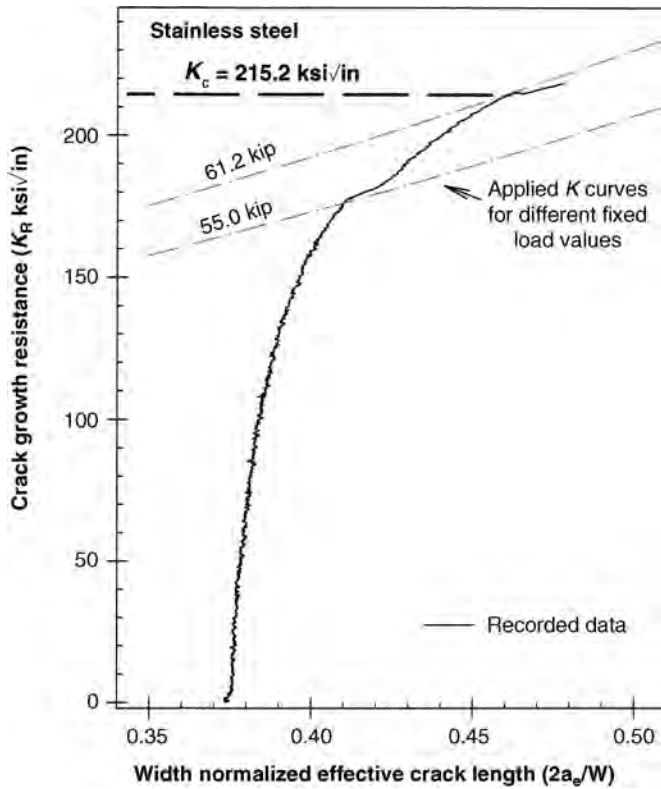


FIGURE 41.12 Example R curve from a K_{Ic} fracture toughness test performed on stainless steel.

amount of crack tearing necessary for the J_{Ic} fracture toughness definition. This high accuracy is practically achieved through higher accuracy (minimum of 16 bit), low noise data acquisition boards combined with extreme care and control over the critical test details, especially as related to loading pin interfaces and clip gauge attachment.

In editions of the ASTM test standards before 1998, the test method for the elastic-plastic J_{Ic} parameter has always been described in ASTM E813. However, the E08 committee of ASTM has recently published a combined approach under the designation E1820. This combined approach outlines a procedure to follow that can result in any one of the primary toughness parameters, namely K_{Ic} , K -, and J -resistance curves. Although these three parameters constitute the most common fracture test standards, other ASTM methods are also available:

- CTOD fracture toughness (E1290),
- plane-strain crack arrest fracture toughness (E1221),
- fracture testing with surface-crack tension specimens (E740),
- Chevron notch fracture toughness of metallic materials (E1304).

However, these methods are typically neither as wide-ranging nor as commonly encountered in design as the previously noted fracture toughness test methods.

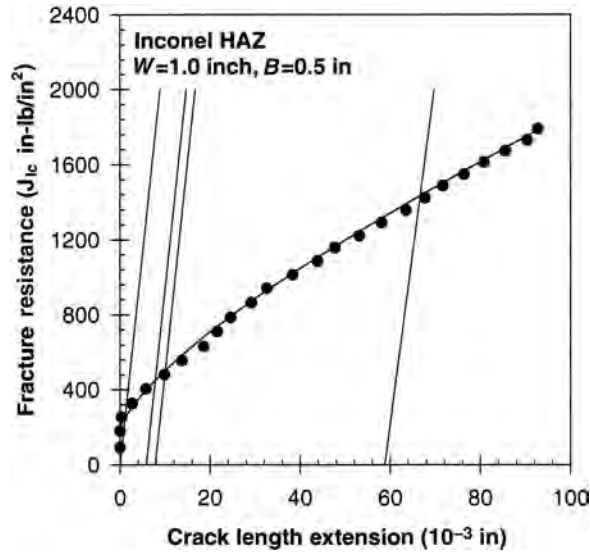


FIGURE 41.13 Example of an R curve from a J_{Ic} fracture toughness test performed on a high-strength weld.

41.6 FATIGUE TESTING

The previous section described basic fracture mechanics concepts as well as the test methods used to quantify fracture behavior given a pre-existing crack. In this section, the process of crack nucleation (e.g., crack initiation) and stable propagation (e.g., crack growth) under cyclic loading is addressed. Understanding crack propagation behavior is critically important since some structures, most notably the current fleet of commercial airliners, adopt a damage tolerance design philosophy whereby cracks are assumed in the structure, and a safe structural life requires detecting and managing these cracks. Quantifying crack nucleation is also important as much of a crack's cyclic life is spent initiating or growing when it is small in size.

The test specimens used for fatigue crack initiation testing are similar geometrically to tensile specimens. The two basic test methods for initiation testing are ASTM E606 and E466 for strain- and stress-controlled testing, respectively. With both of these methods, the goal is to cycle a specimen until failure occurs. Strain control is typically used if (a) the loading is with a constant displacement or (b) the focus is on the short life regime (estimated to be less than 10^3 to 10^5 cycles although this is fairly controversial) where applied stress levels often approach the yield strength of the metallic material. Conversely, load control testing is performed for the mid- to high-cycle life regime. Stress (effectively load) control testing is relatively straightforward since no displacement monitoring or measurement is required. The specimen can be gripped, cycling can commence, and continue unmonitored (provided the servocontrol loop does not require any attention) until failure occurs.

One of the most critical details when performing this type of fatigue testing is controlling bending in the specimen. Poor control of bending typically increases the apparent scatter in the fatigue data generated. Commercial alignment fixtures are available to

mitigate bending. These fixtures utilize a strain gauged specimen and a fixture that allows minute adjustments of the loading train. However, these fixtures are not always available, which means less high tech and pragmatic methods are often required to minimize the effect.

Although guidelines on how to measure and verify specimen alignment are provided in ASTM E1012, minimizing bending practically requires control of three issues. First, the specimen must be carefully machined with excellent dimensional control, parallelism, and axisymmetry. Second, the fixtures used to connect the specimen to the loading train must not introduce any asymmetry. Threaded specimen ends, although commonly used during fatigue testing, should be avoided as the tolerances required to engage the thread make accurate alignment problematic (however, it must be stated that it is not always practical or possible to avoid threaded ends). If the loading cycle is tensile only (e.g., a positive R ratio where R is the ratio of minimum to maximum load), a spherical joint, or more commonly known as a U-joint, in the loading train can minimize bending. Third, the fixtures in the load train should be designed well enough to grip repeatably in the same manner so that once a load train has been carefully aligned with an instrumented specimen it does not vary.

Some typical stress-life ($S-N$) data are shown in Figure 41.14 for two materials, an aluminum and titanium alloy. $S-N$ data are usually generated at a fixed R ratio with the results represented as a plot with stress amplitude or maximum stress on the abscissa and cyclic life on the ordinate. Data are sometimes generated on specimens that have a groove or hole to simulate stress concentrations typically found in structure. Furthermore, it is often of practical interest to know the endurance limit of a metallic material; in theory the endurance limit is defined as the applied stress level that corresponds to infinite life, in practical terms infinite life is approximated in the laboratory as a million cycles. The

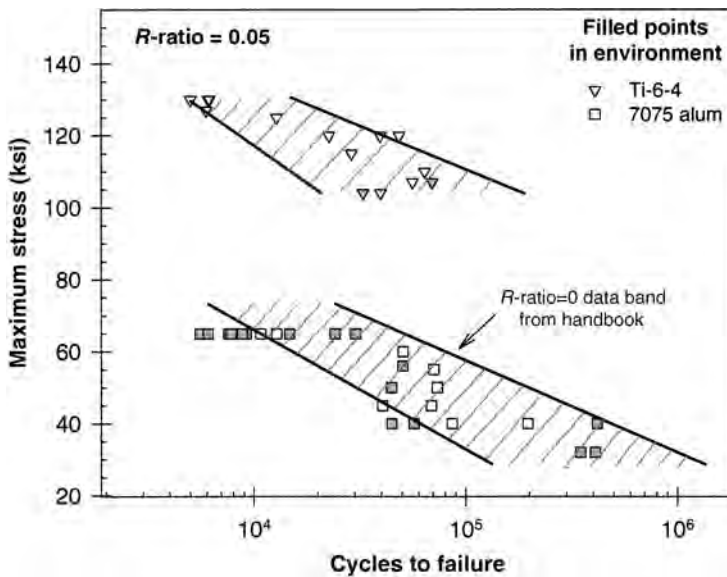


FIGURE 41.14 Conventional $S-N$ fatigue test data presented for example titanium and aluminum alloy materials.

usual approach (ASTM, 1963) is to test 30 specimens using an iterative, staircase method to converge on the stress level corresponding to the endurance limit. The nominal stress step chosen is approximately 2% of the ultimate tensile strength of the metallic material. Although this is a test-intensive methodology, the technique is able to clearly define the endurance limit to the accuracy given by the stress step.

The specimens used for fatigue crack propagation testing are similar to that described previously when discussing fracture. The most common geometries are the compact tension, C(T), and the middle cracked tension, M(T), specimens. However, the pin-loaded C(T) is most suitable for positive R ratio (e.g., tension) as the cracked geometry is unstable in compression. Fatigue crack growth (FCG) testing is described in detail in one of the longest and most complex ASTM test standards, ASTM E647. The standard has grown in length as the understanding of the FCG process has deepened. During FCG testing, the goal is to determine crack growth rate as a function of stress intensity factor range, ΔK . A typical series of FCG curves are shown in Figure 41.15 for an aluminum alloy subject to a number of different stress ratios. The central regions of the crack growth curves can be observed to be relatively linear as expected in the moderate growth rate Paris regime.

One of the most critical aspects of FCG testing is measurement of the crack length. The most common method is to use a traveling telemicroscope mounted on a vernier scale for accurate measurement. Recent advances in both crack length measurement and test automation (Braun, 2002; Ruschau, 1995) use elastic compliance and electric potential

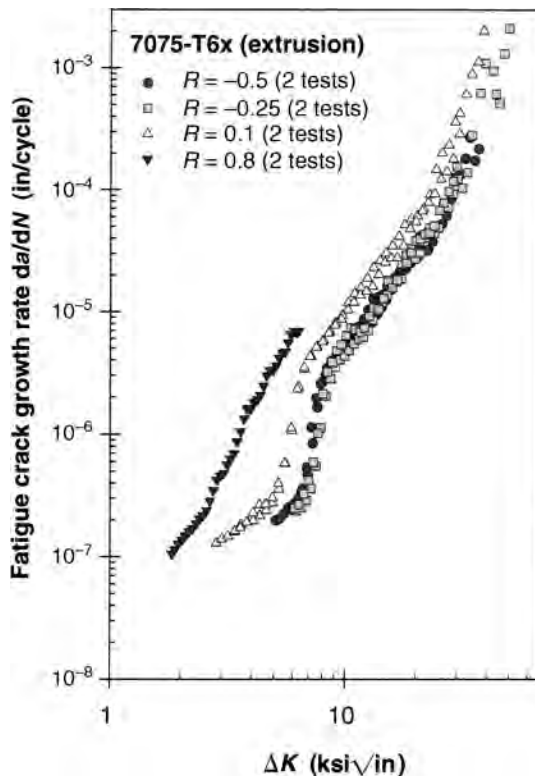


FIGURE 41.15 Series of FCG curves for aluminum at various R ratios.

drop methods to measure crack length. However, it is critical to note that these indirect, nonabsolute measurement techniques both require periodic visual crack length measurement for each test to ensure proper calibration. The quality of a laboratory's nonvisual crack length measurement technique and equipment is critical and often directly related to the quality of the FCG data generated.

The conventional method for performing crack growth testing is to record growth rate data with a fixed loading cycle as the crack grows. More efficiently, newer methods are now available that critically depend upon nonvisual crack length measurement methods to control the applied loading. For instance, K gradient techniques can be used where the quantity $C = dK/K da$ is fixed at different levels for different portions of the test. This increases efficiency since (a) it allows the generation of multiple crack growth data sets from the same specimen and (b) data can be generated in short growth intervals resulting in much faster data generation than with constant amplitude loading. However, these advanced FCG test techniques must be carefully applied by experienced testing professionals who fully understand the limitations of this accelerated approach and have the instrumentation available to accurately measure crack length nonvisually.

Most FCG rate testing is performed at a fixed load ratio as this variable does not change for many structures. However, another method gaining wide popularity is generating FCG data from the highest rates down to near threshold (e.g., crack growth rate at a very low magnitude, a quantity especially useful for design purposes) using a constant K_{\max} , increasing K_{\min} loading condition. Consequently, when the test starts, the R ratio is low (0.1 or so) and as the test progresses R increases until approximately 0.9 or more in the near-threshold regime. Whereas conventional FCG threshold testing may take weeks or months to complete, this method results in a highly accelerated test and can be performed in a matter of days. Again, care must be taken to ensure that the choice of the variable test control parameters does not affect the magnitudes or trends of the FCG rate data obtained.

For design purposes, it is sometimes necessary to understand the potential impact of external variables on FCG behavior. For this case, constant ΔK testing can be used to determine effects difficult and time consuming to detect if a complete crack growth curve were generated. As observed in the data in Figure 41.16, several regions of nominally constant ΔK are sequentially applied with the external environmental variable perturbed for each segment. This allows a better assessment of the effect of the variable since a statistically relevant number of crack growth rates are available for each segment. Although this is a highly efficient method to determine effects, the disadvantage is that the measurement strictly applies to only one constant ΔK level. Nevertheless, a similar approach can be used to indirectly determine, for instance, the magnitude of internal residual stresses in a component by examining small changes in FCG rate.

The challenge in applying FCG rate data is accounting for the variable amplitude loading environment typically applied to a component in service. Most of the analytical tools available are empirically based and require some testing under the variable amplitude, spectrum load condition. In these cases, it is not uncommon to perform a crack growth test using a repeated variable amplitude spectrum derived from either service recorded or design loading. The test specimens in these cases are typically more complex, reflecting some design detail such as a fastener hole or weld. Rather than generating crack growth rate data with these tests, the goal is typically to provide crack length as a function of spectrum passes (or cycle count). Although this type of test is widely performed in the aerospace and ground vehicle industries, (McKeighan, 2003) the myriad of approaches

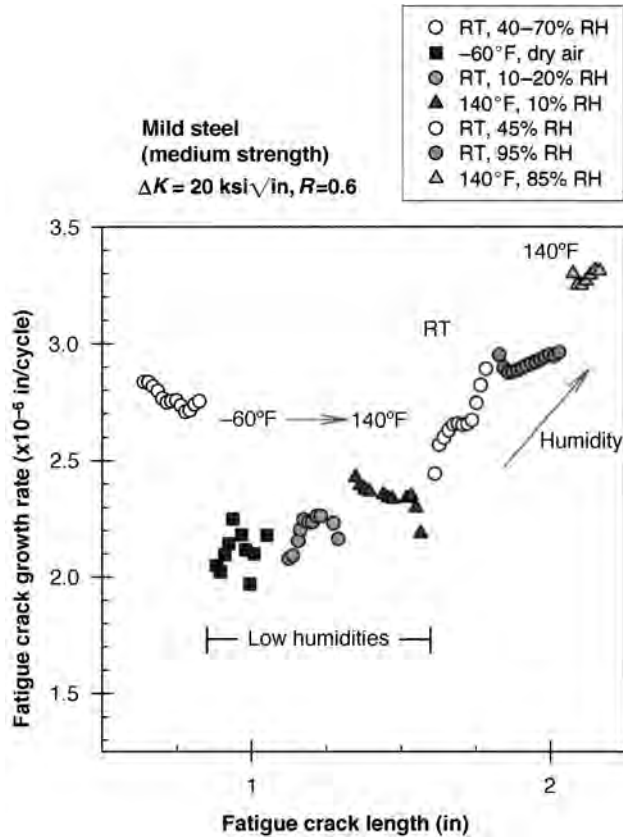


FIGURE 41.16 Influence of environment on FCG rate behavior using a constant ΔK testing approach.

used has stifled standardization, and it is only now underway within ASTM to develop a standard practice applicable to this highly complex test.

41.7 OTHER MECHANICAL TESTING

There are a number of test methods and procedures that are related to the areas described earlier but, in the spirit of brevity, have been omitted. Nevertheless, several are sufficiently popular and warrant some mention, albeit brief.

Some highly specialized assessments other than tensile tests exist for determining the strength and ductility of a metallic material. One common application for the bending strength test is the use of flat metallic materials for spring applications (ASTM E855). The two primary properties that this test provides are the modulus of elasticity in bending and the bending yield strength (proportional limit). Moreover, bend tests used to assess ductility (ASTM E190 and E290) do not provide quantitative results, but rather provide a deformed sample suitable for close visual inspection to determine basic metallic material behavior. The region of highest strain is typically inspected for possible cracking, rumples, or orange peeling as a quality control assessment suitable for the shop floor.

High strain rate testing is another area where significant effort has been expended to generate the constituent properties applicable in the high rate regime. For instance, the split Hopkinson pressure bar deforms a sample at a high strain rate ($>100\%$) while applying uniaxial stress to the sample. Similar methods have been developed that also allow high rate torsional loading. These methods have been successfully applied to a wide range of metallic materials.

Finally, numerous test techniques are also available for testing adhesion, friction, and wear of surfaces. These are critical issues from a design viewpoint with new standardized methods being developed fairly regularly.

41.8 ENVIRONMENTAL CONSIDERATIONS

It is often necessary to perform mechanical testing under environmental conditions other than ambient air. For instance, FCG testing is often performed in either dry, high humidity, or aqueous conditions. For aqueous conditions, a setup as shown in Figure 41.17 can be constructed to perform crack growth rate measurements. Sealing against a flat surface is relatively straightforward provided the materials used in the chamber are impervious to the environment contained.

Other strategies such as that shown in Figure 41.18 can be used to seal the loading pins in a compact tension specimen. This technique has proven very suitable in the past for containing aqueous environments. Clear Teflon (PTFE) bags are commercially available that allow the setup to also be subjected to temperatures in the range of -400 to 300°F . Simpler methods using supermarket polypropylene bags can also be used to contain conditioned air by introducing either desiccant or a hydrated sponge for the dry and humid conditions, respectively. Although the engineer's inclination might be to build a more elaborate environmental chamber, a simple plastic bag with duct tape seals works sufficiently well to contain dry [$<5\%$ relative humidity (RH)] and humid [$>95\%$ RH] environments.

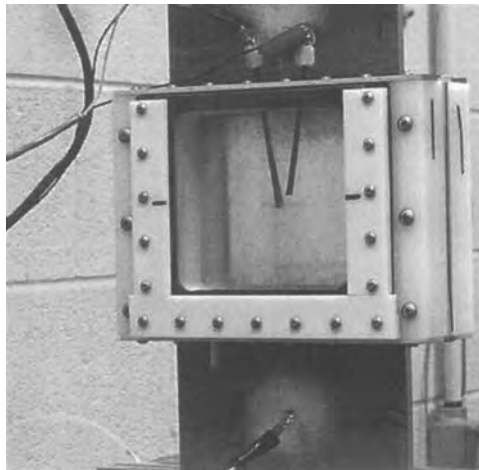


FIGURE 41.17 Chamber around a flat panel M(T) specimen for suspending jet fuel or other aqueous environments while measuring FCG rate behavior.

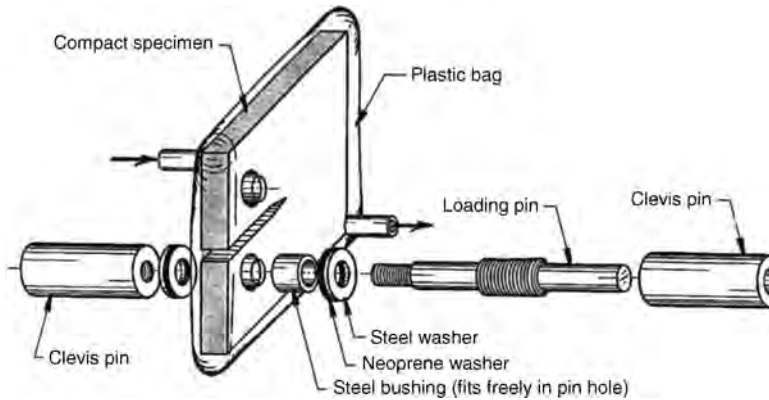


FIGURE 41.18 Schematic of a compact tension specimen sealed with a plastic bag.

Finally, it is sometimes not simple to superimpose two rate-dependent processes without special attention. For instance, care must be taken to ensure that the time-dependent effects that often occur in an environment are properly applied in the materials test. As an example, the deleterious effect of seawater is gradually lost during a FCG test as the frequency exceeds 1 Hz.

If a geometry is somewhat unusual and a setup does not fit into existing furnaces, it is sometimes necessary to construct a custom furnace from insulating fiberboard so long as the temperature required is not too high. In this case, a circulating fan is included along with a heater element and an inexpensive temperature controller. However, an inexpensive approach such as the one just described will not perform well given an aggressive environment, higher pressure applications or more highly elevated temperatures. For these cases, more expensive and custom containments are typically required as shown in Figure 41.19.

Thermomechanical fatigue (TMF) differs from the fatigue testing described earlier as the temperature of the specimen is programmed to vary in a precise manner relative to the mechanical loading. The challenge with this type of testing includes (a) achieving and (b) controlling the thermal environment required. The limiting factor is not usually with

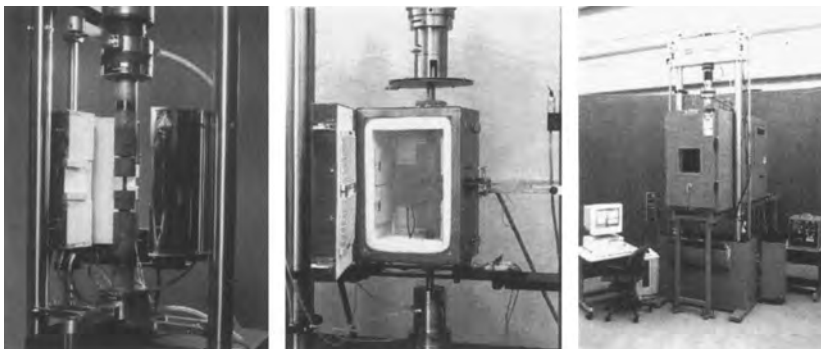


FIGURE 41.19 Several examples of enclosures for both high- and low-temperature materials testing (Photos courtesy of the Instron Corporation and MTS Corporation).

heating as induction furnaces, for instance, can heat a sample very rapidly in a uniform manner. The cycling rate is typically controlled by cooling and concerns regarding large thermal gradients exist if cooling occurs too quickly. Recent work within ASTM has generated a draft standard for TMF testing that is currently undergoing balloting.

ACKNOWLEDGMENTS

The author would like to graciously acknowledge the wisdom and lessons gained from working closely with two highly experienced colleagues and mentors at Southwest Research Institute. Jack FitzGerald taught me that often the simplest approach, the one that seemed too elementary and often at odds with engineering education, was always the best approach when faced with something unusual or unknown in the laboratory. Furthermore, Andy Nagy exhibited incredible creativity by taking half-baked, malformed ideas and transforming them into real devices that functioned flawlessly in the test laboratory. Both of these consummate professionals, now deceased, provided excellent role models and mentors for the development of an inexperienced test engineer. Finally, gratitude is extended to FTA Inc, Instron, and MTS for supplying some of the photographs used herein.

REFERENCES

- Specification for Casing and Tubing, API. Specification 5CT*, 6th ed. Washington, DC: American Petroleum Institute; October 1998.
- A Guide for Fatigue Testing and the Statistical Analysis of Fatigue Data*, ASTM. STP 91 A, 2nd ed. Philadelphia, (PA): American Society for Testing and Materials; 1963.
- 2000 Annual Book of ASTM. Standards, Section Three: Metal Test Methods and Analytical Procedures*. Vol. 3.01. West Conshohocken (PA): American Society for Testing and Materials; 2000.
- International Directory of Testing Laboratories: 2001 Edition*. West Conshohocken (PA): American Society for Testing and Materials; 2001.
- Braun AA, McKeighan PC, Nicolson MA, Lohr RP, editors. *Applications of Automation Technology to Fatigue and Fracture Testing and Analysis: Fourth Volume*. ASTM STP 1411. West Conshohocken, (PA): American Society for Testing and Materials; 2002.
- Joyce JA. *Manual on Elastic-Plastic Fracture Laboratory Test Procedures*, ASTM Manual Series MNL. 27, West Conshohocken, (PA): American Society for Testing and Materials; 1996.
- Kuhn H, Medlin D, editors. *ASM Handbook on Mechanical Testing and Evaluation*. Vol. 8. Materials Park (OH): ASM International; 2000.
- Lucas GF, Stubbs DA, editors. *Nontraditional Methods of Sensing Stress, Strain and Damage in Materials and Structures*, ASTM STP 1318. West Conshohocken (PA): American Society for Testing and Materials; 1997.
- Lucas GF, McKeighan PC, Ransom JS, editors. *Nontraditional Methods of Sensing Stress, Strain and Damage in Materials and Structures: Second Volume*. ASTM STP 1323. West Conshohocken (PA): American Society for Testing and Materials; 2001.
- McKeighan PC, Ranganathan N, editors. *Fatigue Testing and Analysis Under Variable Amplitude Loading Conditions*, ASTM STP 14xx. West Conshohocken (PA): American Society for Testing and Materials; 2003.
- Ruschau JJ, Donald JK, editors. *Special Applications and Advanced Techniques for Crack Size Determination*, ASTM STP 1251. Philadelphia: American Society for Testing and Materials; 1995.

CERAMICS TESTING

SHAWN K. MCGUIRE AND MICHAEL G. JENKINS

- 42.1 Introduction
- 42.2 Mechanical testing
 - 42.2.1 Strength
 - 42.2.2 Creep
 - 42.2.3 Hardness
 - 42.2.4 Fracture toughness
 - 42.2.5 High strain rate
 - 42.2.6 Fatigue
- 42.3 Thermal testing
 - 42.3.1 Thermal expansion
 - 42.3.2 Thermal conductivity
 - 42.3.3 Heat capacity
- 42.4 Nondestructive evaluation testing
 - 42.4.1 Ultrasonography
 - 42.4.2 Radiography
- 42.5 Electrical testing
 - 42.5.1 Electrical resistance at elevated temperatures
 - 42.5.2 Flexural strength of electronic-grade ceramics
- 42.6 Summary
- References

42.1 INTRODUCTION

Ceramics are being used in an ever-increasing capacity in numerous industrial and scientific applications. Advanced ceramics, in particular, have great future possibilities for use in a wide variety of applications where their unique combination of properties can achieve better results than other materials such as metals and polymers. Advanced ceramics

are typically wear and corrosion resistant, lightweight, and thermodynamically stable. In addition, many advanced ceramics have electrical properties that make them advantageous for use in electronic applications such as electronic packaging. Ceramics are finding increasing use in such areas as biomaterials, electronics, microelectromechanical systems (MEMS), and heat transfer applications such as gas turbines in the aerospace industry. The market value of advanced ceramics in the United States is estimated to increase to \$11 billion by the year 2003 from a value of \$7.5 billion in 1998 (Richerson and Associates and Energetics, Inc., 2000). Similarly, the world demand for advanced ceramics has increased from an estimated \$16.7 billion in 1994 to \$25.3 billion in the year 2000 (ISO/TC, 2000).

With the steadily increasing use of ceramics in industry, there exists an increasing demand to characterize and quantify the properties of ceramics. This leads to higher demand for improved testing techniques to yield more exact data used for endeavors such as design, safety analysis, quality control, and scientific understanding. In particular, the brittle nature of ceramics makes their fracture characteristics especially important for the engineer that needs to design ceramic components with life-cycle and safety considerations. In addition to the predominant mechanical testing being performed, there is demand for improved testing of thermal, electrical, and environmental properties as well as others.

Of equal importance to the driving forces behind ceramic testing is the value of the data being obtained from these tests. Do the potential benefits of the data outweigh the cost of performing the tests? What are the predictive abilities of the quantified test results?

To aid an engineer or a scientist who is starting their research into potential ceramic test methods, this chapter gives a brief overview of the major test methods currently being used, relative advantages and disadvantages to each other, and some common sources of error. In addition, a listing of standards that are both relevant and directly applicable to ceramics is given for various testing methods. These standards can be further investigated by the reader who wants a more in-depth look into the exact setup and requirements of a given test. Standardization is an important process in the evolution of ceramic testing to improve accuracy and precision of measured results. New standards are continuously being developed and while not all the standards listed for each test are specifically targeted for ceramics, they still have relevance in terms of improving the exactness of the test setup and results.

42.2 MECHANICAL TESTING

With the increasing role of ceramics in technology, further understanding of mechanical properties has become increasingly more important. This has resulted in the use and standardization of various test methods to better understand and quantify mechanical properties. Test methods reveal such properties as strength, fatigue, fracture resistance, creep, and slow crack growth, which contribute to design, scientific understanding, and estimations of service life among others.

42.2.1 Strength

One of the most important properties for characterizing a material is strength. Strength (yield, ultimate, etc.) is often used as a measure of success in materials development (Jenkins et al., 1998). The characterization of the fracture strength (approximately equal to

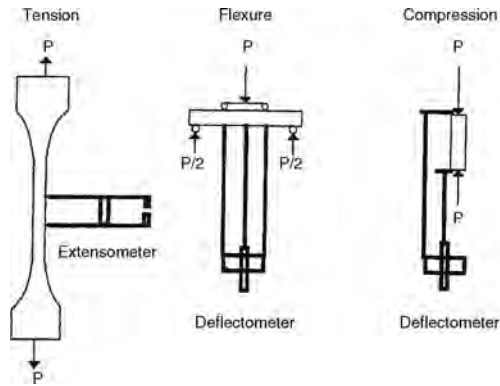


FIGURE 42.1 Schematic of tensile, flexure, and compressive setups (from Jenkins et al. (1998) by courtesy of Marcel Dekker Inc.).

ultimate strength for brittle materials) distribution is needed when ceramic design for structural applications involves failure probability as a criteria (Khandelwhal and Vaccari, 1992). Various test methods are used to determine the fracture strength of a given ceramic material. Usually, the fracture strength is equated to the maximum stress at the point of fracture, which requires that the stress distribution in the test specimen be known. A common source of error in tests measuring fracture strength is that the strength of ceramic materials is strongly influenced by the test specimen's size, geometry, and surface finish. The more common methods for measuring fracture strength include uniform uniaxial stress loading, nonuniform uniaxial stress loading, and biaxial stress loading.

Uniform uniaxial stress load testing is performed by putting a given test specimen into a state of tension or compression (Figure 42.1). Testing in a uniform, uniaxial stress field controls the stress-state variable that allows determination of the mechanical behavior at given loads. Tension and compression tests are not as common as the nonuniform flexure test because the brittle nature of many ceramics results in the need for specialized testing equipment, extensive preparation of test specimen, and an adequately uniform stress state, which can be difficult to achieve (Jenkins et al., 1998).

42.2.1.1 Tension The tensile test equipment consists of two main parts, the test specimen grip holding the actual test specimen and the interface attachment that connects the test specimen grip to the test machine. The grip should be designed to reduce any eccentricity to maintain a uniaxial stress state. Ceramics, in particular, are not able to use a threaded head as an interface because of the difficult nature of machining such a brittle material. The tapered head test specimen is also difficult to machine to proper tolerances. The button head test specimen has emerged as the most reliable interface method (Richerson and Associates and Energetics, Inc., 2000). The attachment interfacing between the test specimen grip and the test machine is usually one of two designs: flexible, self-aligning and fixed, alignable (Jenkins et al., 1998).

Ideally, the stress state will be characterized by the simple equation:

$$\sigma = \frac{P}{A}$$

where σ is the normal stress, P is the applied force, and A is the cross sectional area.

However, the actual stress state in the gauge section, σ_{gs} , will include error from eccentricity in the testing equipment and/or the test specimen and can be characterized as:

$$\sigma_{gs} = \sigma_a + \sigma_b$$

where

$$\text{axial stress, } \sigma_a = \frac{P}{A} \quad \text{and} \quad \text{bending stress, } \sigma_b = \frac{Per}{I}$$

with e being the eccentricity distance, r being the distance from the point in the gauge section at which the stress is being measured to the centroid, and I being the moment of inertia in the gauge section cross section.

Some of the tensile testing standards that apply directly to ceramics are:

Japanese Industrial Standards Committee (JISC),¹ “Testing Methods for Tensile Strength of Monolithic Advanced Ceramics at Room and Elevated Temperature,” JIS R1606-1995;

American Society for Testing and Materials (ASTM),² “Standard Test Method for Tensile Strength of Monolithic Advanced Ceramics at Ambient Temperatures,” C1273-95;

ASTM, “Standard Test Method for Tensile Strength of Monolithic Advanced Ceramics at Elevated Temperatures.” C1366-97.

The following standards are directly applicable to metals but can have applicability to ceramics:

ASTM, “Standard Test Methods for Tension Testing of Metallic Materials,” E8-96a;

ASTM, “Standard Test Methods for Elevated Temperature Tension Testing of Metallic Materials.” E21-92;

ASTM, “Standard Practice for Verification of Specimen Alignment under Tensile Loading,” E1012-93A.

42.2.1.2 Compression The compression strength of ceramic materials is usually much greater than that in tension. Consequently, tensile strength is usually the critical factor in terms of design. The compression test usually consists of two load blocks exerting a compressive force on a cylindrical test specimen. The test specimen should be uniform to avoid buckling of individual layers aligned with the applied load.

Common sources of error are size mismatches between the load block and specimen, surface irregularities, and eccentric loading as explained in the tensile testing section (Jenkins et al., 1998). The first three errors bring about excessive stresses in the ends of the test specimen, which can cause failures in the nongauge section of the test specimen. Eccentricity, as in tensile testing, reduces uniformity in the gauge section stress state.

The following standards are directly applicable to compressive testing of ceramics:

JISC, “Testing Method for Compressive Strength of High Performance Ceramics.” R1608-1990;

¹ Japan Standards Association, Tokyo, Japan.

² ASTM International, West Conshohocken, Pennsylvania, USA.

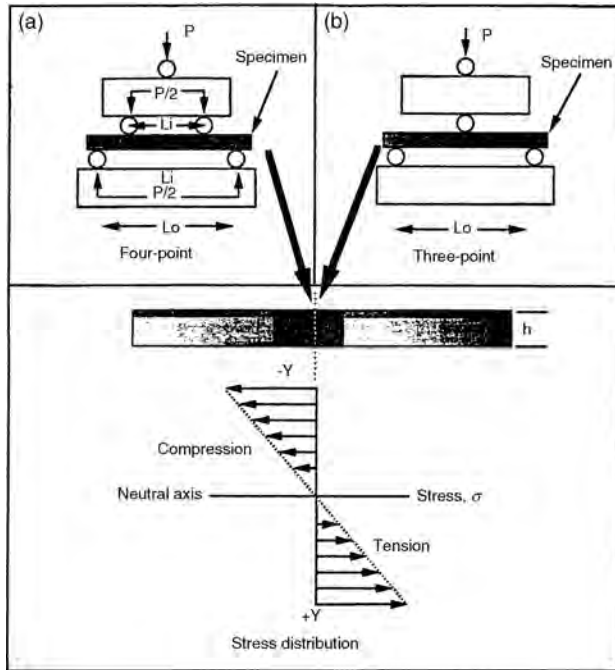


FIGURE 42.2 Three- and four-point flexure testing and resulting stress state (from Jenkins et al. (1998) by courtesy of Marcel Dekker Inc.).

ASTM, “Standard Test Method for Compressive (Crushing) Strength of Fired White-ware Materials,” C773-88.

42.2.1.3 Flexure Testing in flexure typically involves a three- or four-point loading of a test specimen as shown in Figure 42.2. Compared to tension and compression tests, flexure tests are less expensive, simpler in setup, and easier to adapt to elevated temperature testing (Jenkins et al., 1998). The applied moment, M , yields the equation for uniaxial normal stress, σ :

$$\sigma = \frac{MC}{I}$$

where the moment of inertia I is $(bh^3)/12$ (b = width and h = height) and C is the distance from the neutral axis to the outer surface of the test specimen. Given that the moment $M = P(L_o - L_i)/4$ (L_o = outer span, L_i = inner span) and $y = h/2$ we can substitute into the equation to find the fracture strength, S_f , which is the maximum tensile stress obtained at the fracture force, P_f :

$$S_f = \frac{1.5P_f(L_o - L_i)}{bh^2}$$

Various sources of internal and external errors in the testing process can affect the measured results of flexure tests. Errors classified as “internal” involve deviations from simple beam theory and involve test specimen geometry and properties (Quinn and Baratta, 1985). A test specimen with an initial curvature resulting from residual stresses generated

during machining would cause an internal error. Another example of an internal error would be excessive specimen deflection during testing. Support point frictional forces at large deflections will have a component aligned with the applied force that will increase the applied moment. "External" errors are those classified as being caused by incorrect test fixture geometry. Improper location of the inner load points is an example that causes external error. Another example is the generation of torque on the test specimen. This can be caused by an initially twisted test specimen, unparallel line loads, or nonuniform line loads at the contact points (Quinn and Baratta, 1985; Baratta, 1984). External error can also occur from compressive contact stresses at the support pins, which can result in localized crushing. This error can be reduced by using support pins above a critical radius r_c (Jenkins et al., 1998).

To minimize the errors associated with flexure testing, the test specimen and test fixtures must adhere to certain restraints and standards. Test specimen geometry has been standardized based on error considerations. A common geometry in the United States is 3 by 4 by 50 mm for the test specimen and inner and outer spans of 20 and 40 mm, respectively, for the fixtures (Quinn, 1990). The test specimen must be isotropic and homogeneous to apply the maximum tensile stress equation previously given. In addition, the specimen should be as free as possible of surface defects as the maximum tensile stress occurs on the surface.

In comparing the three- and four-point flexure test methods, it is found that the four-point method is more appropriate for determining fracture strength because no shear stresses are generated as in the three-point test method. The three-point test method with its simpler geometry, however, may be more attractive for tests in which stable crack growth must be induced into the test specimen (Jenkins et al., 1998).

Some of the flexure test standards that are applicable to advanced ceramics are listed here:

JTSC, "Testing Method for Flexural Strength (Modulus or Rupture) of Fine Ceramics," R1601-1995;

JISC, "Testing Method for Flexural Strength of Fin Ceramics at Elevated Temperature," R1604-1995;

ASTM, "Standard Test Method for Flexural Strength of Advanced Ceramics at Ambient Temperatures," C1 161-94;

ASTM, "Standard Test Method for Flexural Strength of Advanced Ceramics at Elevated Temperatures," C1211-92;

Comité Européen de Normalisation (CEN),³ "Advanced Technical Ceramics—Monolithic Ceramics—Mechanical Properties at Room Temperature—Part 1: Determination of Flexural Strength," EN843—1:1995;

CEN, "Advanced Technical Ceramics—Monolithic Ceramics—Thermomechanical Properties—Part 1: Determination of Flexural Strength at Elevated Temperature," ENV820—1:1993.

42.2.2 Creep

Similar to the fracture testing methods previously discussed, the three test methods used to determine creep characteristics in ceramics are the tensile, compressive, and flexural

³ Comité Européen de Normalisation, Brussels, Belgium.

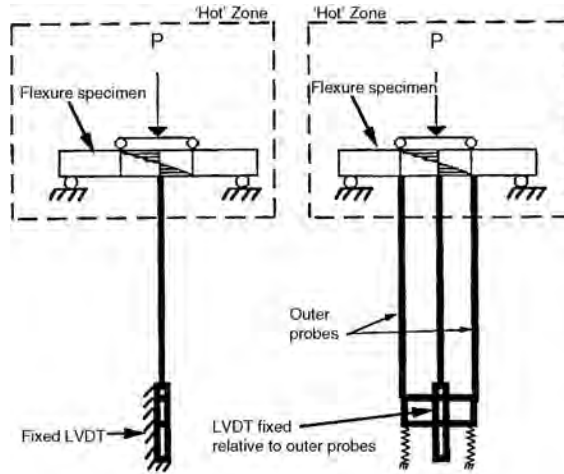


FIGURE 42.3 Schematic of typical test setup for creep testing in flexure (from Jenkins et al. (1998) by courtesy of Marcel Dekker Inc.).

tests. The primary differences in the test fixtures are that the gauge section must be heated to the desired elevated temperature and an alternate method of measuring deflection such as extensometers must be employed because of the increased temperature. Figure 42.3 shows a typical test setup for a flexure test that utilizes extensometers.

42.2.2.1 Flexure Creep Flexure testing using four-point loading yields the maximum tensile stress of a beam in bending given in the fracture strength section:

$$S_f = \frac{1.5P_f(L_o - L_i)}{bh^2}$$

In addition, the maximum tensile strain equation is given as:

$$\varepsilon_{\max} = \frac{4h\delta_{\text{center}}}{(L_i)^2}$$

Compared to the tensile and compressive creep tests, flexure testing is less expensive, easier to design, and requires less test specimen preparation. However, the nonuniform stress state induced in the flexure testing results in difficulty in determining the creep characteristics of a material if it deforms differently in tension than in compression. This difference in tensile and compressive creep behavior is a result of cavitation among other factors. Conducting an additional test in either compression or tension allows for the determination of uniform stress-state creep characteristics from the nonuniform flexure test results. Additional errors from twisting, friction between test specimen and load points, and excessive contact stress at support points are also considerations.

42.2.2.2 Tensile Creep Tensile testing to determine creep characteristics yields a uniform stress state characterized by the equation:

$$\sigma_{\text{gs}} = \frac{P}{A}$$

The uniform stress state eliminates the problem of different deformations in tension and compression. This results in the much simpler calculation of stress and strain than with flexure testing. The test setup, however, is more expensive and difficult to implement. The test specimens required for creep tensile testing are also more difficult and time intensive to design and manufacture. In addition to stress gradients from the test specimen geometry, temperature differences induce an unwanted temperature gradient in the specimen from the grip contact portion of the specimen to the heated gauge section (Jenkins et al., 1998). This temperature gradient becomes larger and poses more of a problem when cold grips are used. Unwanted failure in the grip interface portion of the specimen can also be a problem. Various test standards are applicable to tensile testing of advanced ceramics. ASTM C1291-95 directly applies to tensile testing for creep behaviors:

ASTM, "Standard Test Method for Elevated Temperature Tensile Creep Strain, Creep Strain Rate, and Creep Time to Failure for Advanced Monolithic Ceramics," C1291-95;

ASTM, "Standard Test Method for Tensile Strength of Monolithic Advanced Ceramics at Elevated Temperatures," C1366-95;

ASTM, "Standard Test Method for Monotonic Tensile Strength Testing of Continuous Fibre-Reinforced Advanced Ceramics with Solid Rectangular Cross-Sections at Elevated Temperatures," C1359-96.

42.2.2.3 Compressive Creep Compressive testing to determine creep behavior, like tensile testing, results in a uniform stress state that allows for easy determination of creep characteristics. It has been shown that creep in compression can vary significantly from creep in tension so it is of value to conduct both tests to "get the best picture" of a material's creep properties. Some advantages to the compressive test are the low cost in preparing the material as well as the small geometry of the specimen, which allows for more testing per sample material size.

There are some drawbacks to the compressive test as well. The test specimen must be aligned to a greater accuracy than tension or flexure tests to prevent unwanted bending and buckling. End constraints on the test specimen can result in shear stresses if the specimen ends are constrained, which results in a nonuniform stress state. In addition, temperature gradients can have an adverse effect on test result accuracy.

42.2.3 Hardness

Hardness is an important property to quantify in ceramics. Measured hardness indicates the ability of the ceramic to resist deformation by a hard object. Usually, Knoop or Vickers diamond indenters are used in conjunction with a microindentation hardness machine (Quinn, 2000). Rarely are the popular Rockwell and Brinell indenters used for ceramics research. Vickers indenters are used to characterize roughly 60% of the ceramic hardness values that are published (Quinn, 2000).

The indentation force should always be included with the hardness value. For the most accurate results, the entire force versus hardness curve should be measured as shown in Figure 42.4.

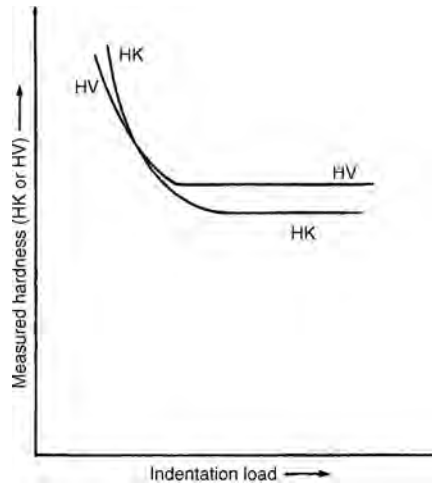


FIGURE 42.4 Typical hardness versus load (force) plot (from Quinn (2000), Figure 3, p. 245).

Discrepancies can arise at different indentation forces. At higher forces, cracking can complicate the measuring process or make measuring impossible (Quinn, 2000). Measuring the hardness from the indentation, especially at small forces, is also a significant source of error in hardness testing. The hardness value can change based on the force value applied to the test specimen at small forces. Volume 8 of the *ASM Handbook* (Quinn, 2000) recommends forces greater than or equal to 9.8 N for Vickers and Knoop indentations. Errors in the measurement of the indentation diagonal length essentially double the hardness error as the hardness value is proportional to the square of the diagonal length. A Versailles Advanced Materials and Standards (VAMS) round-robin test project conducted on alumina ceramic samples resulted in uncertainty in the hardness values given by the laboratories involved (Butterfield et al., 1989). Although using numerous indentations can reduce some of this uncertainty, engineers and scientists conducting hardness tests should nevertheless keep this uncertainty in mind when considering their data.

Standards for measuring hardness for ceramics are listed below:

Vickers hardness;

ASTM, “Standard Test Method for Vickers Indentation Hardness of Advanced Ceramics,” C1427-97;

CEN, “Advanced Technical Ceramics—Monolithic Ceramics—Mechanical Properties at Room Temperature—Part 4: Determination of Vickers, Knoop and Rockwell Superficial Hardness Tests,” prEN834-4;

JISC, “Testing Method for Vickers Hardness of High Performance Ceramics,” R1610:1991.

ISO, “Fine Ceramics (Advanced Ceramics, Advanced Technical Ceramics)—Test Method for Hardness of Monolithic Ceramics at Room Temperature,” ISO 14705:2000

Knoop hardness;

ASTM, "Standard Test Method for Knoop Indentation Hardness of Advanced Ceramics," C1426-99;

CEN, "Advanced Technical Ceramics—Monolithic Ceramics—Mechanical Properties at Room Temperature—Part 4: Determination of Vickers, Knoop and Rockwell Superficial Hardness Tests," prEN834-4;

ISO, "Fine Ceramics (Advanced Ceramics, Advanced Technical Ceramics)—Test Method for Hardness of Monolithic Ceramics at Room Temperature," ISO 14705:2000;

ISO, "Glass and Glass-Ceramics—Knoop Hardness Test," ISO 9385-1990.

42.2.4 Fracture Toughness

The brittle nature of ceramics results in low resistance to fracture, quantified as fracture toughness, which is an important factor in many applications of ceramics. Fracture toughness is a measure of a specimen's ability to resist further growth of a crack. Low fracture toughness values increase the risk of catastrophic failure of a ceramic component. Ceramic matrix composites (CMCs) have better fracture toughness compared to monolithic ceramics as the additional reinforcing elements help to deter crack growth. The concepts of the linear-elastic fracture mechanics (LEFM) method commonly used for other materials can be applied to monolithic ceramics for the purpose of analysis. For ceramic matrix composites, ongoing research is being conducted regarding the application of the concepts of elastic-plastic fracture mechanics (EPFM) methods (Miller and Liaw, 2000).

There are various testing methods and setups for fracture toughness testing. Some methods included in the testing standard, ASTM C1421, are the single-edge precracked beam (SEPB) method, the surface crack in flexure (SCF) method, and the chevron notched beam (CNB) method.

The SEPB, SCF, and CNB methods all consist of a flexural test of a beam in which a short straight crack is induced into the tensile side. The primary differences are the methods by which the beam is deformed to introduce the preliminary short crack. All three methods have good calibration characteristics but have some drawbacks as well. With the SEPB method, it can be difficult to obtain straight-fronted cracks. Crack initiation in CNB testing can be difficult as a result of residual stresses from machining. SCF testing can only be performed on materials that produce high-quality cracks from the indentation (Miller and Liaw, 2000).

Standards organizations have introduced fracture toughness testing standards utilizing various testing methods. For example, ASTM C1421 generally uses a flexure test of a cracked bend bar while JIS R1607 utilizes only two test methods, SEPB and IF (indentation fracture). Standards that apply directly to fracture toughness testing of ceramics are given below.

Japanese Industrial Standards Committee, "Testing Methods for the Fracture Toughness of High Performance Ceramics," JIS R1607-1990;

American Society for Testing Materials, "Standard Test Methods for the Determination of Fracture Toughness of Advanced Ceramics at Ambient Temperature" ASTM C1421-99;

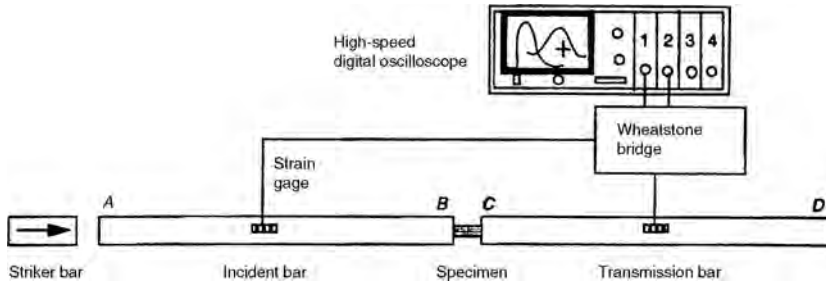


FIGURE 42.5 Test setup schematic for traditional split Hopkinson pressure bar testing (from Subhash and Ravichandran (2000), Figure 1, p. 497).

International Organization for Standardization, “Fine Ceramics (Advanced Ceramics, Advanced Technical Ceramics)—Test Method for Fracture Toughness of Monolithic Ceramics at Room Temperature by Single Edge Pre-Cracked Beam (SEPB) Method,” ISO DIS15732 (in 1999).

42.2.5 High Strain Rate

Split Hopkinson pressure bar (SHPB) testing has primarily been used in the past to measure the plastic properties of metals. The fact that ceramics are brittle and many ceramic compounds and alloys show only elastic strains before fracture makes accurate testing and measurement more difficult than with more ductile materials such as metals.

A traditional SHPB test configuration is shown in Figure 42.5 (Subhash and Ravichandran, 2000). The test setup consists of an incident bar, a transmission bar, and a striker bar. The ceramic test specimen is placed in between the incident and transmission bars as shown. The striker bar strikes the incident bar at a predetermined velocity by being launched from a gas gun. When it impacts the incident bar, it generates a compressive pulse that continues to travel to the test specimen. Part of the incident pulse travels through the test specimen, and the remaining part is reflected back into the incident bar. The stress pulses are measured by strain gauges placed at midpoints on the transmission and incident bars. An oscilloscope displays the measured pulses. The stress, strain, and strain rate equations are given as (Kolsky, 1949):

$$\sigma_s(t) = \frac{A_o E_o}{A_s} \varepsilon_T(t)$$

$$\dot{\varepsilon}_s = -\frac{2C_o}{l_s} \varepsilon_R(t)$$

$$\varepsilon_s(t) = \int_0^t \dot{\varepsilon}_s(t) dt$$

where E is Young's modulus, l is length, A is cross-sectional area, σ is stress, ε is strain, $\dot{\varepsilon}$ is strain rate, and t is time. Subscripts T , s , o , and R refer to the transmitted pulse, specimen, bar, and reflected pulse.

Some of the inherent properties of ceramics such as brittleness and high strength, conflict with assumptions that are made in deriving the stress, strain, and strain rate equations

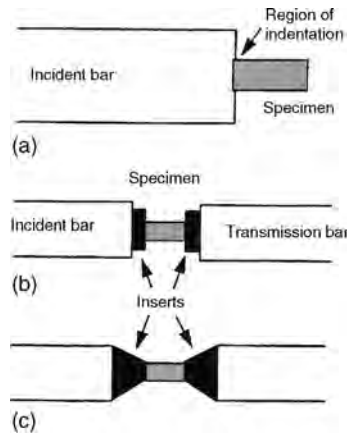


FIGURE 42.6 Incident bar and transmission bar protection using tungsten-carbide inserts (from Subhash and Ravichandran (2000), Figure 3, p. 499).

(Subhash and Ravichandran, 2000). An example is the assumption that the bar end surfaces remain flat and parallel during the deformation of the test specimen. The hard nature of ceramics can cause the test specimen to indent into the steel bar ends, resulting in stress concentration around the test specimen end faces. These stress concentrations result in nonuniform, nonuniaxial stress state in the test specimen, which violates another assumption in using the equations, which is that the stress state is uniform and uniaxial. The stress concentrations can consequently lead to chipping and failure of the ceramic test specimen through microcrack initiation.

One way to minimize the problem of test specimen damage is to place tungsten-carbide inserts in between the test specimen and bar ends as shown in Figure 42.6 (Subhash and Ravichandran, 2000). The high strength of the inserts prevents the ceramic test specimen from indenting into the incident and transmission bars, thereby reducing stress concentrations. However, the tungsten-carbide inserts can have the adverse effect of altering the incident, transmitted, and reflected stress-wave properties that can result in inaccurate strain gauge measurements. This can be prevented by selecting tungsten-carbide inserts with the same impedance as the bar material. Selection of the diameter of the inserts can also help solve this problem. Typically, the insert lengths are one fourth of the test specimen length (Subhash and Ravichandran, 2000).

42.2.6 Fatigue

Fatigue testing is an important design tool for the designer of ceramic components where reliability and lifetime estimates need to be made. Fatigue tests for ceramics generally cover the three situations of cyclic fatigue, static fatigue, and dynamic fatigue (Breder and Wereszczak, 1998). Cyclic fatigue is the periodic loading of a component under various load ratios usually denoted as R :

$$R = \frac{\sigma_{\min}}{\sigma_{\max}}$$

Common scenarios for the stress ratio include $R > 0$ (both minimum and maximum stresses in tension or compression) and $R < 0$ (minimum stress in compression and

maximum stress in tension). The simple loading scenario would involve a sinusoidal loading that varied continuously with a minimum and maximum stress. In real-world examples, however, the loading can be much more complicated.

Static fatigue testing involves slow crack growth of a test specimen with a constant tensile force under the desired conditions. The length of time for the test specimen to fail and the applied stress are measured and used to obtain the fatigue characteristics by using various fatigue equations. Dynamic fatigue testing involves applying a constant, increasing stress (i.e., constant nonzero stress rate) to a test specimen under desired conditions. The stress rate and maximum applied stress at failure are measured and used in conjunction with various fatigue equations to determine the fatigue characteristics.

Fatigue testing can further be divided into the two categories of “direct” and “indirect.” Direct methods, also known as fracture mechanics methods, involve running tests with test specimens with previously induced cracks and directly observing crack growth. Indirect methods, also known as strength measurement techniques, involve measuring the strength of the test specimen over time intervals and using the data to estimate fatigue properties.

42.2.6.1 Indirect Methods Static loading utilizing the indirect method involves subjecting a tensile or flexure test specimen to a constant load under the desired environment conditions. The applied force stress and the amount of time for specimen failure are measured and used to estimate the fatigue properties of the ceramic material. A typical plot of failure stress versus time to failure is shown in Figure 42.7.

Uniaxial flexural tests in three- or four-point loading situations are performed under the directions given in JIS R1601, “Test Method for Flexural Strength (modulus of rupture) of Fine Ceramics.” At least three levels of stress are recommended. The fatigue strength can be related to the failure time with the following equation (Salem and Jenkins, 2000):

$$t_f = \frac{B}{\sigma_f^2} \left[\left(\frac{\sigma_i}{\sigma_f} \right)^{n-2} - 1 \right]$$

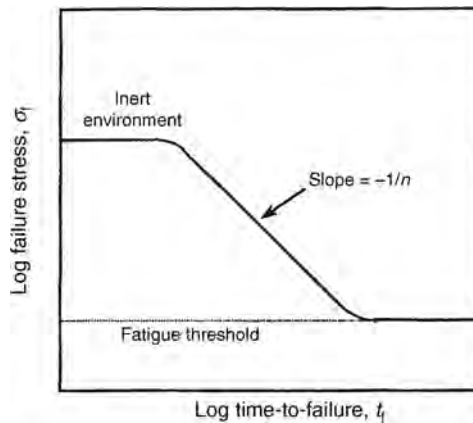


FIGURE 42.7 Failure stress versus time to failure (from Salem and Jenkins (2000)).

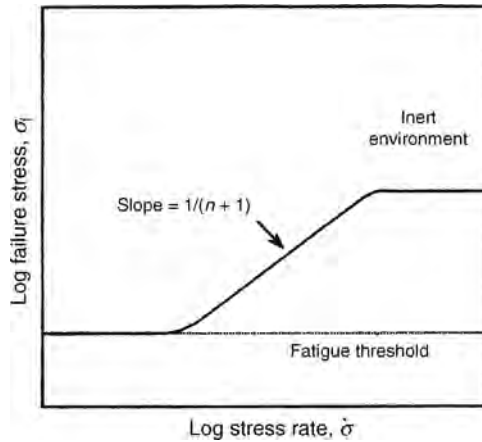


FIGURE 42.8 Failure stress versus stress rate (from Salem and Jenkins (2000)).

where B is a variable determined from crack geometry, fracture toughness, loading situation, and K and n , which are material and environment constants; σ_i is the inert strength at which no crack extension occurs for noncorrosive conditions.

Indirect dynamic loading usually involves tensile or flexure test specimens subjected to a constant stress rate. The applied stress rate and the failure stress are measured and used to estimate fatigue properties. Figure 42.8 shows a typical plot of failure stress versus the stress rate.

The standard, ASTM C1368, gives guidelines for fatigue strength estimating via dynamic loading. It involves four-point flexure testing of a ceramic specimen in accordance with ASTM C1161 “Standard Test Method for Flexural Strength of Advanced Ceramics at Ambient Temperature.” The fatigue strength and applied stress rate are related by the equation (Salem and Jenkins, 2000):

$$\sigma_f = [B(n+1)\sigma_i^{n-2}\dot{\sigma}]^{1/n+1}$$

where the same variable definitions apply as in the static loading equation.

There are various methods for applying periodic loads to ceramics. With bending fatigue tests, a periodic force with a specified frequency and stress ratio is applied to three- or four-point flexure test setups. Cyclic tensile fatigue tests involve cyclically loading a test specimen with a testing apparatus with specified load ratio and frequency. Various test specimen shapes can be used and tension-compression cyclic fatigue (stress ratio < 0) has been performed utilizing a button-head specimen and clamping fixture. Rotary bending is also used for fatigue testing ceramics and is especially useful for analyzing components such as shafts that undergo similar stress states while in use.

42.2.6.2 Direct Methods Direct methods, also known as fracture mechanics methods, use fatigue tests that use test specimens with an induced crack (Salem and Jenkins, 2000). Crack growth is observed directly or via strain gauges or other measuring devices.

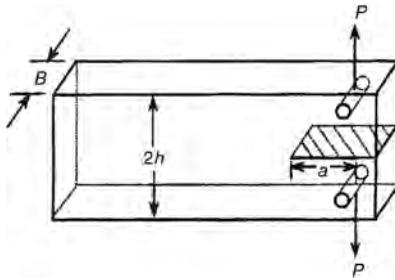


FIGURE 42.9 Double-cantilever beam test method (from Salem and Jenkins (2000)).

One testing technique is the double-cantilever beam method, shown schematically in Figure 42.9. For the configuration shown, the stress-intensity factor and fracture toughness can be found utilizing the equation (Salem and Jenkins, 2000)

$$K_I = \frac{Pa}{h^{3/2}\sqrt{Bb}} \left(3.47 + 2.32 \frac{h}{a} \right)$$

where B is the thickness of the test specimen, b is the web thickness, a is the crack length, h is half the specimen height, and P is the applied force. Some standards that directly apply to fatigue testing of ceramics are given below.

- Japanese Industrial Standards Committee, “Test Methods for Static Bending Fatigue of Fine Ceramics,” JIS R1632—1998;
- Japanese Industrial Standards Committee, “Test Method for Flexural Strength,” JIS R1601—1995;
- American Society for Testing Materials, “Standard Test Method for Determination of Slow Crack Growth Parameters of Advanced Ceramics by Constant Stress-Rate Flexural Testing at Ambient Temperature,” ASTM C1368-99;
- American Society for Testing Materials, “Standard Test Method for Flexural Strength of Advanced Ceramics at Ambient Temperature,” ASTM C1161-99;
- Japanese Industrial Standards Committee, “Test Method for Bending Fatigue of Fine Ceramics,” JIS R1621—1996;
- American Society for Testing Materials, “Standard Test Method for Flexural Strength of Advanced Ceramics at Elevated Temperatures,” ASTM C1211-95.

42.3 THERMAL TESTING

42.3.1 Thermal Expansion

Thermal expansion is an important property that quantifies the volume change a material undergoes when it is subjected to temperature changes. Typically, materials expand when heated and contract when cooled. This characteristic is valuable to the engineer or scientist involved in ceramic design or research. The coefficient of thermal expansion (COTE),

α , is strongly related to the strength of the atomic bonds. Energy must be put into the material for the atoms to move from their equilibrium positions. Typically, ceramics, which have strong ionic or covalent bonds, have lower thermal expansion coefficients than metals. The average COTE is simply the change in the length of the material per unit length per unit temperature:

$$\alpha = \frac{\Delta L}{L_0 \Delta T}$$

The volume COTE, α_v , is the change in volume of the material for a given temperature. Consequently, for an isotropic material the volume COT is given as:

$$\alpha_v = 3\alpha$$

Thermal expansion is a tensor property that is different along individual crystallographic axes (Hammetter, 1991). This varying expansion can cause residual stresses in some test specimens that can actually cause cracks between the crystal faces in some rare situations. This thermal cycling cracking will manifest itself in the form of a hysteresis curve from expansion data during thermal cycling tests.

Typically, the COTE is lower for materials having a high melting point. The linear COTE is a more exact definition and continuously changes with regard to temperature. The COTE for materials is usually listed in handbooks as a complicated temperature-dependent function or is shown as a constant that is valid only for a specified temperature range. Taking into account, the temperature dependence of the linear COTE, experimental measuring gives the more exact equation (Hammetter, 1991)

$$\alpha(T_i) = \left(\frac{\partial L(T)}{\partial T} \right)_{T=T_i} \frac{1}{L(T)}$$

where $L(T_i)$ is the experimentally measured length of the specimen at a given temperature T_i . The difference in the linear COTE and the average COTE is shown graphically in Figure 42.10.

Understanding the thermal expansion of a given ceramic is especially important in design applications such as composites that combine structurally distinct constituents (fiber, matrix, etc.) of different materials (ceramic, metal, etc.) with different COTE. For example, a ceramic with a low COTE combined with a metal possessing a high COTE can result in critical stresses for a given temperature change.

42.3.1.1 Dilatometry The most popular method for measuring thermal expansion is dilatometry (Hammetter, 1991). The test setup is simple, consisting of a cylindrical test specimen placed between a fixed base and a movable push rod. The ceramic test specimen is heated at a fixed rate, which displaces the push rod a given distance that is recorded by a sensing device. Software is readily available to record the data and determine the temperature-dependent COTE and the average COTE record.

Dilatometers can be of the single or double push rod variety. The single push rod method involves first calibrating the test apparatus in terms of the change in length versus temperature-sensing device with a test specimen possessing a known COTE. Next, the ceramic test specimen is tested compared to the previously measured standard to determine its COTE. Calibrating the test apparatus with a standard reference specimen first in

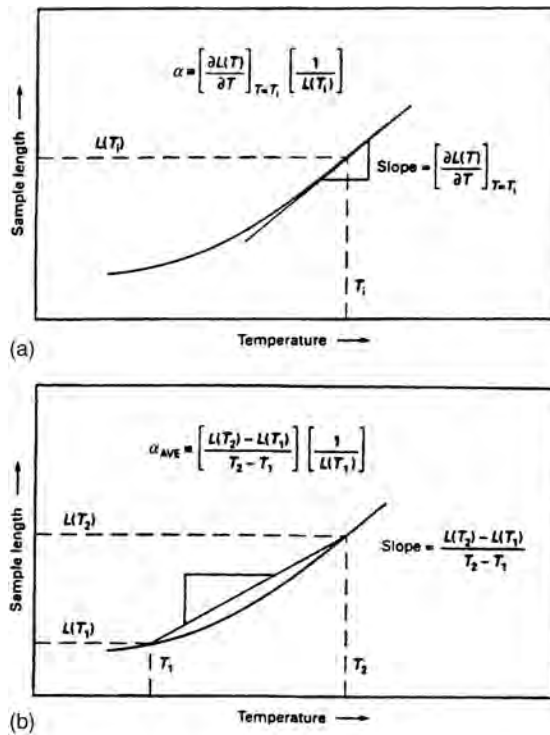


FIGURE 42.10 Difference in linear COTE and average COTE (from Hammett (1991), Figure 1, p. 612).

this manner can correct for problems such as dilatometer material expansion, thermal gradients, and nonlinearity in the heating rate (Hammett, 1991). The double push rod design tests the ceramic test specimen and the standard reference material side by side. The thermal expansion data is then calibrated by comparing it with the standard reference material with a known COTE.

42.3.1.2 Other Techniques X-ray diffraction can give insight into the structure of ceramics. The tensor components of thermal expansion can be determined with X-ray diffraction through the measuring of changes of the interplanar spacing with relation to the temperature. One distinct advantage with this method is that only a very small sample of the ceramic is required. Another advantage is that measuring the tensor components of thermal expansion of a specimen is possible with X-ray diffraction. This can be important when trying to estimate residual stresses that will occur during thermal cycling.

Interferometry is a technique that involves placing the test specimen between two reflecting surfaces and measuring the displacement through the movement of the surfaces. The basis for this being that parallel reflective surfaces a short distance apart will show interference fringes when illuminated by monochromatic light. As the reflective surfaces move apart from the specimen expansion. The interference fringes move past a reference point on one of the reflecting planes. The expansion of the sample in terms of length can

be expressed as (Hammetter, 1991):

$$\frac{\Delta L}{L} = \frac{\lambda N}{2L} + \frac{A}{L}$$

where N is the number of interference fringes that pass the reference point, L is the specimen length, ΔL is the change in the specimen length, λ is the wavelength of the light source, and A is the correction for the light source based on the atmosphere in the measuring environment. Restrictions on the test specimen geometry make this technique less popular than dilatometry methods. One advantage to this technique, however, is that ceramics with volatile components can be tested by interferometry inside a closed, heated test chamber where the volatile component's vapor pressure can be controlled.

42.3.2 Thermal Conductivity

Thermal conductivity is a measure of the rate of heat transfer in a given material by conduction. The heat flowrate through a material is proportional to the heated area of the material and the temperature gradient across the specimen. This proportionality yields the coefficient of thermal conductivity, κ , which is shown in the following equation (Hammetter, 1991):

$$\frac{dQ}{dt} = -\kappa A \frac{dT}{dx}$$

The negative sign on the right side indicates that heat flows from a higher to a lower temperature. The rate of heat flow is dQ/dt , A is the cross-sectional area of the material, and dT/dx is the temperature gradient.

Two important mechanisms involved in transferring thermal energy are lattice vibrations (or phonons) and transfer of free electrons. Valence electrons travel from hot to cold areas with the energy gained, and then transfer their energy to other atoms. Many electrons in ceramics cannot be excited into the conduction band except at quite high temperatures because the energy gap is typically too large (Hammetter, 1991). Consequently, heat transfer in ceramics is primarily a result of lattice vibrations. This results in ceramics typically having a much lower thermal conductivity than metals. However, not all ceramics have low thermal conductivity. Advanced ceramics such as AlN and SiC have good thermal conductivity and low electrical conductivity. Consequently, they are good for electronic applications where dissipating heat is a requirement. Thermal conductivity is an important aspect of materials for microelectronic substrates and electronic packaging materials. The development of higher density circuits makes heat dissipation an increasingly difficult problem to overcome.

42.3.2.1 Guarded Hot Plate The guarded hot plate technique is a comparative method that measures thermal conductivity through the application of thermocouples to a cylindrical ceramic test specimen sandwiched between two cylindrical sections of a reference material with the same diameter and with a known thermal conductivity. The test setup is shown in Figure 42.11. Thermally conductive paste is applied to the mating surfaces to ensure adequate heat transfer between materials. Using a reference material with a thermal conductivity similar to that of the ceramic test specimen will

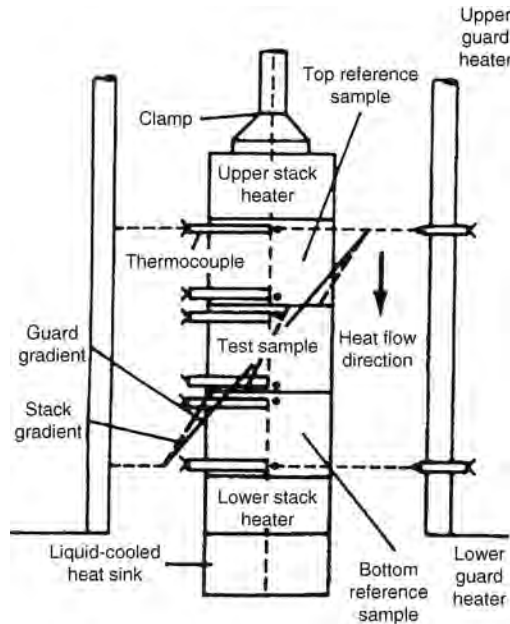


FIGURE 42.11 Guarded hot plate test setup (from Hammett (1991), Figure 2, p. 613).

yield the best results. Stack heaters at the top and bottom of the cylinder stack introduce heat into the materials. The thermocouples are positioned near the mating surfaces on the reference and test cylinders. The thermocouples determine the temperature gradient and then the steady-state heat flux is established between the stack heaters. Using the thermal conductivity of the reference material and the measured temperature gradients in the two test specimen cylinders sandwiching the reference cylinder, the heat flux through the entire stack can be calculated using the thermal expansion equation previously given. These two independent calculations of heat flux through the stack are then averaged, and the average is used as the heat flux value for the ceramic test specimen. In addition, the difference in the two calculated values give the heat loss value for the stack. The coefficient of thermal conductivity, κ , can then be determined because the average heat flux through the stack, the area of the test specimen cylinder, and the measured temperature gradient along the ceramic test specimen's length are all already known. To eliminate convective heat losses, measurements can be made in a vacuum or reduced atmosphere environment.

42.3.2.2 Laser Flash Method Another method for determining thermal conductivity is the laser flash method, which involves quickly heating a thin ceramic specimen on one side via a quick “thermal pulse” from a laser and then using the measured temperature values over time of the back face of the specimen to calculate thermal diffusivity and conductivity. Thermal diffusivity, α , is a different means of expressing a material's heat conduction properties. Thermal diffusivity takes into account that heat can diffuse through a material subject to different boundary conditions, causing both spatial

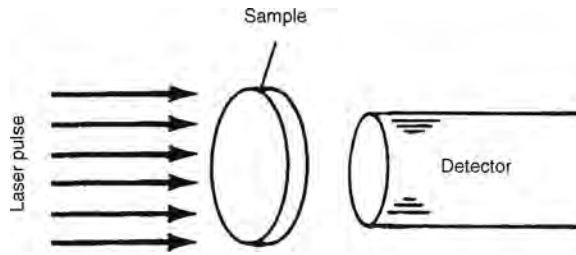


FIGURE 42.12 Schematic of laser flash method test setup (from Hammetter (1991), Figure 3, p. 614).

and temporal variations of temperature (Hammetter, 1991). Thermal diffusivity is a temperature-dependent material tensor property that is defined by the equation

$$\frac{dT}{dt} = \alpha_{ij}(T) \nabla^2 T$$

The relationship between thermal diffusivity and thermal conductivity is given by the equation

$$\alpha_{ij}(T) = \frac{k_{ij}(T)}{\rho(T)c_p(T)}$$

where $\rho(T)$ is the temperature-dependent density and $c_p(T)$ is the specific heat capacity of the test specimen.

The ceramic test specimen is in the shape of a thin disk and is at a constant temperature T , when the laser flash exerts a thermal pulse on the front face. Figure 42.12 shows a schematic of the test setup.

The assumptions here are that the thermal energy is deposited uniformly over the front face of the test specimen, that the heat travels along the thickness of the specimen only to the back face, and that the pulse heats the sample only (Hammetter, 1991). Then, knowing the temperature of the back face at times after the pulse is introduced into the test specimen, the thermal diffusivity of the test specimen can be calculated using various algorithms. Different available algorithms use various temperatures of the back face to calculate thermal diffusivity.

An advantage of the laser flash method is that, in addition to having a simple test setup, it can be used over a large range of thermal conductivities and ceramic specimen temperatures. Also, it is quite popular for measuring thermal conductivities at high temperatures as maintaining a steady-state condition is difficult at higher temperatures.

Other thermal pulse methods using different sources such as electron beams or quartz flash lamps are also possible utilizing very similar test procedures.

42.3.3 Heat Capacity

Heat capacity is a property that refers to the amount of energy that must be added or subtracted from a material to raise or lower its temperature. The amount of energy required to raise the temperature of a material by a degree varies from material to material based on its properties. Specific heat at a constant pressure, c_p , is the most common

expression of a material's heat capacity and is defined as the amount of heat needed to raise the temperature of one gram of a substance by one kelvin at a constant pressure. The specific heat is derived from the enthalpy, H , and this relation is shown in the following equation:

$$c_p \left(\frac{\partial H}{\partial T} \right)_p$$

where T is the temperature and the subscript p indicates constant pressure. From this equation, we can see that the specific heat is the slope of the enthalpy versus temperature change. The specific heat value can be considered a constant for a short range of temperature, but the actual measured specific heat term is expressed as a polynomial to express the nonlinearity of the specific heat with the temperature. This more defined specific heat capacity and its form are given in the equation

$$C_p(T) = a + bT + cT^2 - dT^{-2} + eT^{-1/2}$$

However, the form of equation above does not provide a specific heat for all temperature ranges of many materials. Various equations of this form must be determined for various temperature ranges. There are various methods for determining the heat capacity of ceramics.

42.3.3.1 Calorimetry Calorimetry can be used to determine the heat capacity of a ceramic material. A specific amount of a material test specimen is heated to an initial temperature with an external furnace and then deposited into a calorimeter of a lower temperature. The calorimeter measures the heat energy that the test specimen gives off while cooling to the equilibrium temperature that is between the specimen and calorimeter temperature. Measurements of the enthalpy, H , at various temperatures gives a plot of $H(T)$ versus T . Consequently, the specific heat, c_p , can be calculated from this data.

An advantage of calorimetry is that it can calculate specific heats over a wide temperature range. A drawback, however, is that the method is insensitive to transitions with minor changes in enthalpy (Hammetter, 1991).

42.3.3.2 Differential Scanning Calorimetry Differential scanning calorimetry is the most popular method for measuring the specific heat of a ceramic. This type of calorimeter measures the heat flowrate to a ceramic test specimen while it is heated at a given constant rate. A computer monitors the temperature of the test specimen and makes adjustments to keep the temperature of the ceramic test specimen increasing at a constant rate. Consequently, the specific heat, c_p , of a ceramic test specimen of known mass m can be calculated from the following equation (Hammetter, 1991):

$$\frac{dH(T)}{dt} = mc_p(T) \frac{dT}{dt}$$

This method is commonly used and there are various commercial differential scanning calorimeters available. One of the major drawbacks to this technique is that commercial setups can only be used up to a temperature of about 800°C (Hammetter, 1991). In addition, testing errors resulting from thermal resistance between the test specimen and heat

flow sensing device and other sources requires the use of a comparative process when testing. A different material with a known heat capacity is tested under the same conditions and any difference is noted and incorporated into the error evaluation of the ceramic test specimen being investigated.

42.4 NONDESTRUCTIVE EVALUATION TESTING

Many nondestructive evaluation (NDE) testing techniques are commonly used for evaluating metals and other materials. Applying these techniques, however, to ceramics does not always provide adequate results because of the unique nature of ceramics that impede many of the typical testing techniques. The increasing use of advanced ceramic materials in critical applications along with properties that make them sensitive to quite small defects increases the need to define NDE testing methods directly applicable to ceramics.

Advanced ceramics are typically brittle in nature. Defects as small as 10 μm can be critically detrimental and must be avoided. They can be prevented with careful process control of the fabrication of the ceramic materials from fine powders or with careful inspection of the finished ceramic parts. Both strategies can be applied with NDE testing as a major component.

42.4.1 Ultrasonography

Ultrasonic testing is a common NDE testing technique that can detect and describe flaws and material conditions that other methods are unable to do. Ultrasonic waves are propagated through the material and the waves are disrupted at the discontinuities in the material such as defects, voids, or cracks. The waves are scattered or partially reflected at these discontinuities and from this action, a measure of the discontinuity characteristics such as location, size, and shape are revealed. The ultrasonic method does not always reveal everything about the discontinuity to the degree desired but is still an invaluable tool.

The distinguishing characteristic of ceramics that make typical ultrasonic testing techniques less successful are the high ultrasonic velocities and the smaller critical defect size that must be detected. To detect and characterize small defects, the equipment must be modified so that the ultrasonic wave frequency is increased. An ultrasonic wave with a wavelength of similar size to the defect is required to detect the defect. Consequently, with a ceramic material where a critical defect can be a very small value such as 10 μm , a wavelength of approximately the same size is needed (Lott and Kunerth, 1991). This indicates a very high frequency, f , is required based on the frequency equation

$$f = \frac{v}{\lambda}$$

where λ is the wavelength and v is the longitudinal wave velocity.

An important property to be discerned in a ceramic material is its bulk porosity. Ultrasonic testing can reveal the material's bulk porosity even though the individual pore size is much smaller than the investigating wavelength. The porosity can be characterized by measuring the ultrasonic velocity and attenuation. Measuring the ultrasonic velocity typically entails using a transducer to a face of a ceramic specimen with parallel faces and a known thickness. The round-trip travel time of the acoustic pulse is measured with the

transducer in a pulse-echo mode. The ultrasonic velocity v is given by the equation

$$v = \frac{2d}{t}$$

where d is the thickness of the test specimen and t is the time for round-trip travel. It has been found that there is a linear relationship between ultrasonic velocity and porosity. Once the ultrasonic velocity has been measured, the variations in porosity in a specimen can be shown with the aid of ultrasonic C-scans.

Ultrasonic attenuation is the process by which the ultrasonic beam, when propagating through the test specimen, loses energy. Grain boundaries, pores, voids, and other internal defects cause scattering of the waves, resulting in a lower energy beam. This attenuation in the ultrasonic beam is expressed by the equation

$$I = I_0 \exp(-\alpha x)$$

where I is the beam intensity, x is the distance into the test specimen ($x = 0$ at surface), and α is the ultrasonic attenuation coefficient of the material. Once the attenuation of a material is characterized, various imaging equipment can map the amplitude image of a specimen, which visually describes the porosity of a test specimen.

A standard that applies directly to ultrasonic testing is ASTM E494, "Standard Practice for Measuring Ultrasonic Velocity in Materials."

42.4.2 Radiography

Radiography uses radiation to characterize a material's structure by examining the interaction between the electromagnetic wave and the material itself. The detection of voids, cracks, pores, and other defects is the primary goal of this technique. The detection of these defects is a result of the attenuation and scatter of the radiation as it passes through the material. Various radiation sources can be used in the application of radiography but the most versatile has been found to be X-ray sources. There are two common methods for using X-ray sources to discover discontinuities and defects in a material's structure. One method is to create a two-dimensional image of the test specimen or component where the variation of the image intensity indicates the degree of attenuation. Another approach is to use many attenuation image "slices" and an algorithm to develop a three-dimensional image.

X-ray microradiography techniques use a divergent X-ray beam to produce a two-dimensional image of the test specimen or component being investigated. High contrast and clarity are required when trying to detect the inherently small defects in most ceramics. One common method of applying X-ray microradiography is using a contact method where the test specimen or component being studied contacts an imaging medium, which results in a high-resolution image. Another method uses X-ray sources that are typically less than 100 μm , which allows direct magnification of the component or test specimen image (Lott and Kunerth, 1991). The ability of these methods to detect defects is dependent on the degree of contrast and resolution that is obtained in the image. One disadvantage to this two-dimensional technique is that cracks and other linear defects orientated transverse to the beam direction will be much more difficult to detect.

X-ray computed tomography utilizes many two-dimensional "slice" images of the X-ray attenuation to form a three-dimensional representation of the object being

investigated. This is especially useful when trying to discover defects and discontinuities in a complex three-dimensional component. It is used primarily as a research tool and a few of the main drawbacks to the technique are its high cost and complexity.

42.5 ELECTRICAL TESTING

Ceramic materials have important functions in various electrical and electronic applications. Ceramics, with their unique properties, provide capacitive, insulative, conductive, resistive, and other functions in electronic circuitry. An example of the important role ceramics have in electronics is the use of advanced ceramics such as alumina oxide (Al_2O_3) as the substrate material in electronic packaging. Increases in circuit density have consequently resulted in more stringent requirements being placed on substrate materials. A substrate is desired that has thermal expansion characteristics that closely match with silicon to prevent critical thermal stresses. A low dielectric constant is also needed to improve signal processing. In addition, thermal conductivity is a high priority to dissipate heat from the high-density circuit. Based on stringent requirements such as these, the development of new advanced ceramics and subsequent testing to quantify the development and selection of ceramic materials for important electrical/electronic applications is of increasing importance. Directly adapting current electrical testing techniques to ceramics serves to improve the process of ceramic integration into electronics and other electrical applications.

42.5.1 Electrical Resistance at Elevated Temperatures

An important aspect of ceramic use in electronic applications is the variation of electrical resistance as the temperature rises. With the trend toward higher and higher density circuits, characterizing substrate and other chip component properties at elevated temperatures is quite important.

Measuring electrical resistance at high temperature involves applying a voltage across a test specimen contained in a heating furnace. A ceramic test specimen is mounted between the electrodes in the furnace that has heated the test specimen to the desired temperature. A voltage of 500 V DC is applied across the test specimen for one minute and the volume resistance is measured. The process is repeated for various temperatures until maximum test temperature is reached. The volume resistivity, ρ , is then calculated from the following equation:

$$\rho = \frac{A}{h} R_v$$

where R_v is the measure volume resistance, A is the area of the electrode, and h is the average thickness of the ceramic specimen in the area covered by the electrode. Further specifics are given in ASTM D1829-90, "Electrical Resistance of Ceramic Materials at Elevated Temperatures."

42.5.2 Flexural Strength of Electronic-Grade Ceramics

Another important test for characterizing electronic ceramic components is the flexural strength test. Flexural strength testing is already standardized, but application to

electronic-grade ceramics requires slight modification. The testing sample sizes are much smaller and commonly referred to as “microbars.” Changes in the size of the bending test fixtures, load application, and material preparation are also required. The specifics are outlined in ASTM F417-78, “Standard Test Method for Flexural Strength of Electronic-Grade Ceramics.” The flexural strength is still calculated using the standard equation:

$$S = \frac{3PL}{2bd^2}$$

where S is the flexural strength, P is the force at fracture, L is the distance between supports, and b and d are the width and thickness, respectively, of the specimen.

42.6 SUMMARY

Testing of ceramics has been standardized or formalized for mechanical, thermal, NAE, and electrical aspects of these brittle materials. Although testing for many properties and performance related to these aspects is accepted, methodologies for material selection and design with brittle materials are still being developed. As these methodologies mature and become widely accepted, additional test methods related to the required properties and performance will be proposed, developed, and approved.

REFERENCES

- Baratta FI. Requirements for Flexure Testing of Brittle Materials. In: Freiman SW, Hudson CM, editors. *Methods for Assessing the Structural Reliability of Brittle Materials*. ASTM STP 884. W. Conshohocken (PA): ASTM; 1984. p. 194–222.
- Breder K, Wereszczak AA. Fatigue and Slow Crack Growth. In: Cranmer DC, Richerson DW, editors. *Mechanical Testing Methodology for Ceramic Design and Reliability*. New York: Marcel Dekker; 1998. p. 223–227.
- Butterfield, DM, Clinton, DJ, Morrell, R. The VAMAS Hardness Tests Round-Robin on Ceramic Materials,” Report No. 3, Versailles Advanced Materials and Standards/National Physical Laboratory, April 1989.
- Hammetter WF. Thermophysical Properties In: *Engineered Materials Handbook, Ceramics and Glasses*. Vol. 4. Materials Park (OH): ASM International; 1991. p. 610–615.
- Jenkins MG, Wiederhorn SM, Schiffer RK. Creep Testing of Advanced Ceramics. In: Cranmer DC, Richerson DW, editors. *Mechanical Testing Methodology for Ceramic Design and Reliability*. New York: Marcel Dekker; 1998.
- Khandelwal PK, Vaccari DL. Life prediction methodology of ceramic engine components. Proceedings of the Annual Automotive Technology Development Contractors’ Meeting, p. 256, Dearborn, MI, October 28–31, 1991 Warrendale (PA): Society of Automotive Engineers; 1992. 253–260.
- Kolsky H. An investigation of the mechanical properties of materials at very high rates of loading, *Proceedings of the Royal Society of London B* 1949;62:676–700.
- Lott LA, Kuerth DC. NDE Testing and Inspection. In: *Engineered Materials Handbook, Ceramics and Glasses*. Vol. 4. Materials Park (OH): ASM International; 1991. p. 617–626.
- Market Environment and Objectives of the ISO/TC in 7th Plenary Meeting of ISO/TC 206—Report of the Secretariat, 2000. p. 2–6.

- Miller JH, Liaw, PK. Fracture toughness of ceramics and ceramic matrix composites. In: *ASM Handbook*. Vol. 8. Materials Park (OH): ASM International; 2000. p. 654–64.
- Quinn GD, Baratta F. Flexure data can it be used for ceramics part design?. *Advanced Materials and Proceedings* 1985;129(12):31–35.
- Quinn GD. Flexure strength of advanced structural ceramics: a round robin. *Journal of the American Ceramic Society* 1990;73(8):2374–2384.
- Quinn GD. Indentation hardness testing of ceramics. In: *ASM Handbook*. Vol. 8. Materials Park (OH): ASM International; 2000. p. 243–251.
- Richerson and Associates and Energetics, Inc., *Advanced Ceramics Technology Roadmap: Charting Our Course*. USACA Publications; 2000. p. 1–3.
- Salem JA, Jenkins. MG. Fatigue testing of brittle solids. In: *ASM Handbook*. Vol. 8. Materials Park. (OH): ASM International; 2000. p. 768–778.
- Subhash G, Ravichandran, G. Split-Hopkinson Pressure Bar Testing of Ceramics. In: *ASM Handbook*. Vol. 8. Materials Park (OH): ASM International; 2000. p. 497–504.

PLASTICS TESTING

VISHU SHAH

- 43.1 Introduction
- 43.2 Mechanical properties
 - 43.2.1 Tensile tests (ASTM D 638, ISO 527-1)
 - 43.2.2 Flexural properties (ASTM D 790, ISO 178)
 - 43.2.3 Creep properties
 - 43.2.4 Stress relaxation
 - 43.2.5 Impact properties
 - 43.2.6 Izod–Charpy impact test (ASTM D-256, ISO 179)
 - 43.2.7 High-speed impact tests (ASTM D3763, ISO 6603-2)
 - 43.2.8 Fatigue resistance
 - 43.2.9 Hardness tests
- 43.3 Thermal properties
 - 43.3.1 Tests for elevated temperature performance
- 43.4 Electrical properties
 - 43.4.1 Dielectric strength (ASTM D 149, IEC 243-1)
 - 43.4.2 Dielectric constant and dissipation factor (ASTM D150, IEC 250)
 - 43.4.3 Electrical resistance tests
 - 43.4.4 Arc Resistance (ASTM D 495)
- 43.5 Weathering properties
 - 43.5.1 UV radiation
 - 43.5.2 Accelerated weathering tests
 - 43.5.3 Outdoor weathering of plastics (ASTM D 1435)
- 43.6 Optical properties
 - 43.6.1 Refractive index (ASTM D 542)
 - 43.6.2 Luminous transmittance and haze (ASTM D 1003)
 - 43.6.3 Specular gloss (ASTM D 523)

Further readings

43.1 INTRODUCTION

First commercial plastic material known as Bakelite was developed by Dr. Baekeland in 1909. To commercialize this new product it was necessary to test and develop property data. This was the beginning of the plastics testing technology. Initially the methods developed for testing metals and other traditional materials were employed for testing plastic materials. In very short time the scientist realized that the fundamental differences between the polymeric materials and other traditional materials would necessitate developing slightly different and unique test methodology for plastics. For example, metals usually display unchanged mechanical behavior up to their recrystallization temperature of greater than 300°C. For most applications designers can disregard the effect of temperature, environmental, and long term effects of load. They can rely on instantaneous stress strain properties. Such is not the case with plastics. Plastics are viscoelastic. Viscoelasticity is defined as the tendency of plastics to respond to stress as if they are combination of elastic solids and viscous fluids. This property possessed by all plastics to some degree, dictates that while plastics have solid-like characteristics such as elasticity, strength, and form stability, they also have liquid-like characteristics such as flow depending on time, temperature, rate, and amount of loading. This also means that unlike metals, ceramics and other traditional materials, plastics do not exhibit a linear stress–strain relationship. Plastics are sensitive to change in temperature rate of loading environment etc. Over the years ASTM (American Society for Testing and Materials) has done an excellent job in developing and providing the plastics industry with standard testing procedures. ASTM Committee D20 on Plastics was formed in 1937. The committee oversees the development of test methods, specifications, recommended practices, nomenclature, definitions, and the stimulation of research relating to plastics, their raw materials, components, and compounding ingredients, and to finished products made from plastics. These standards have and continue to play a preeminent role in all aspects important to the effective utilization of plastics, including specimen preparation, material specifications, and methodologies for mechanical, thermal, optical, and analytical testing.

43.2 MECHANICAL PROPERTIES

The mechanical properties, among all the properties of plastic materials, are often the most important properties because virtually all service conditions and the majority of end-use applications involve some degree of mechanical loading. Nevertheless, these properties are the least understood by most design engineers. The material selection for a variety of applications is quite often based on mechanical properties such as tensile strength, modulus, elongation, and impact strength. These values are normally derived from the technical literature provided by material suppliers. More often than not, much emphasis is placed on comparing the published values of different types and grades of plastics and not enough on determining the true meaning of the mechanical properties and their relation to end-use requirements. In practical applications, plastics are seldom, if ever, subjected to a single, steady deformation without the presence of other adverse factors such as environment and temperature. Since the published values of the mechanical properties of plastics are generated from tests conducted in a laboratory under standard test conditions, the danger of selecting and specifying a material from these values is obvious. A thorough understanding of mechanical properties, tests employed to determine such properties, and the effect of adverse conditions on mechanical properties over a long period is extremely important.

43.2.1 Tensile Tests (ASTM D 638, ISO 527-1)

Tensile elongation and tensile modulus measurements are among the most important indications of strength in a material and are the most widely specified properties of plastic materials. Tensile test, in a broad sense, is a measurement of the ability of a material to withstand forces that tend to pull it apart, and to determine to what extent the material stretches before breaking. Tensile modulus, an indication of the relative stiffness of a material, can be determined from a stress–strain diagram. Different types of plastic materials are often compared on the basis of tensile strength, elongation, and tensile modulus data. Many plastics are very sensitive to the rate of straining and environmental conditions. Therefore, the data obtained by this method cannot be considered valid for applications involving load-time scales or environments widely different from this method. The tensile property data are more useful in preferential selection of a particular type of plastic from a large group of plastic materials and such data are of limited use in actual design of the product. This is because the test does not take into account the time-dependent behavior of plastic materials.

The tensile testing machine of a constant-rate-of-crosshead movement is used. It has a fixed or essentially stationary member, carrying one grip and a movable member carrying a second grip. Self-aligning grips employed for holding the test specimen between the fixed member and the movable member prevents alignment problems. A controlled velocity drive mechanism is used. Some of the commercially available machines use a closed loop servo-controlled drive mechanism to provide a high degree of speed accuracy. A load-indicating mechanism capable of indicating total tensile load with an accuracy of $\pm 1\%$ of the indicated value or better is used. An extension indicator, commonly known as the extensometer, is used to determine the distance between the two designated points located within the gauge length of the test specimen as the specimen is stretched. Figure 43.1 shows a commercially available tensile testing machine. The advent of new microprocessor technology has virtually eliminated time-consuming manual calculations. Stress, elongation, modulus, energy, and statistical calculations are performed automatically and presented on a visual display or hard copy printout at the end of the test.

43.2.2 Flexural Properties (ASTM D 790, ISO 178)

The stress–strain behavior of polymers in flexure is of interest to a designer as well as a polymer manufacturer. Flexural strength is the ability of the material to withstand bending forces applied perpendicular to its longitudinal axis. The stresses induced due to the flexural load are a combination of compressive and tensile stresses. This effect is illustrated in Figure 43.2. Flexural properties are reported and calculated in terms of the maximum stress and strain that occur at the outside surface of the test bar. Many polymers do not break under flexure even after a large deflection that makes determination of the ultimate flexural strength impractical for many polymers. In such cases, the common practice is to report flexural yield strength when the maximum strain in the outer fiber of the specimen has reached 5%. For polymeric materials that break easily under flexural load, the specimen is deflected until a rupture occurs in the outer fibers.

There are several advantages of flexural strength tests over tensile tests. If a material is used in the form of a beam and if the service failure occurs in bending, then a flexural test is more relevant for design or specification purposes than a tensile test, which may give a strength value very different from the calculated strength of the outer fiber in the bent



FIGURE 43.1 Tensile testing machine (Courtesy Instron Corporation).

beam. The flexural specimen is comparatively easy to prepare without residual strain. The specimen alignment is also more difficult in tensile tests. Also, the tight clamping of the test specimens creates stress concentration points. One other advantage of the flexural test is that at small strains, the actual deformations are sufficiently large to be measured accurately.

There are two basic methods that cover the determination of flexural properties of plastics. Method 1 is a three-point loading system utilizing center loading on a simple supported beam. A bar of rectangular cross section rests on two supports and is loaded by means of a loading nose midway between the supports. The maximum axial fiber stresses

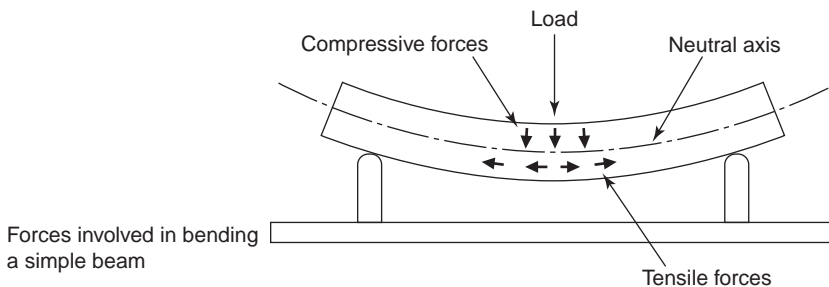


FIGURE 43.2 Forces involved in bending a simple beam (Reprinted by permission of McGraw-Hill Company).

occur on a line under the loading nose. A close-up of a specimen in the testing apparatus is shown in Figure 43.3. This method is especially useful in determining flexural properties for quality control and specification purposes.

Method 2 is a four-point loading system utilizing two load points equally spaced from their adjacent support points, with a distance between load points of one-third of the support span. In this method, the test bar rests on two supports and is loaded at two points (by means of two loading noses), each on equal distance from the adjacent support point. Method 2 is very useful in testing materials that do not fail at the point of maximum stress under a three-point loading system. The maximum axial fiber stress occurs over the area between the loading noses.

Either method can be used with the two procedures. Procedure A is designed principally for materials that break at comparatively small deflections. Procedure B is designed particularly for those materials that undergo large deflections during testing. The basic difference between the two procedures is the strain rate, Procedure A being 0.01 in./in./min, and Procedure B being 0.10 in./in./min.

43.2.2.1 Modulus of Elasticity (Flexural Modulus) The flexural modulus is a measure of the stiffness during the first or initial part of the bending process. This value of the flexural modulus is, in many cases, equal to the tensile modulus.

The flexural modulus is represented by the slope of the initial straight-line portion of the stress-strain curve and is calculated by dividing the change in stress by the corresponding change in strain.



FIGURE 43.3 Close-up of a specimen shown in flexural testing apparatus (Courtesy DuPont Company).

43.2.3 Creep Properties

Today, plastics are used in applications that demand high performance and extreme reliability. Many components, conventionally made in metals, are now made in plastics. The pressure is put on the design engineer to design the plastic products more efficiently. An increasing number of designers have now recognized the importance of thoroughly understanding the behavior of plastics under long-term load and varying temperatures. Such behavior is described in terms of creep properties.

When a plastic material is subjected to a constant load, it deforms quickly to a strain roughly predicted by its stress–strain modulus, and then continues to deform slowly with time indefinitely or until rupture or yielding causes failure. This phenomenon of deformation under load with time is called creep. All plastics creep to a certain extent. The degree of creep depends upon several factors, such as type of plastic, amount of load, temperature, and time.

The short-term stress–strain data is of little practical value in actual designing the part, since such data does not take into account the effect of long-term loading on plastics. Creep behavior varies considerably among types of plastics; however, under proper stress and temperature conditions, all plastics will exhibit a characteristic type of creep behavior. One such generalized creep curve is shown in Figure 43.4. The total creep curve is divided into four continuous stages. The first stage (OP) represents the instantaneous elastic deformation. This initial strain is the sum of the elastic and plastic strain. The first stage is followed by the second stage (PQ) in which strain occurs rapidly but at a decreasing rate. This stage, where creep rate decreases with time, is sometimes referred to as creep or primary creep. The straight portion of the curve (QR) is characterized by a constant rate of creep. This process is called “cold flow.” The final stage (RS) is marked by increase in creep rate until the creep fracture occurs.

If the applied load is released before the creep rupture occurs, an immediate elastic recovery, substantially equal to elastic deformation followed by a period of slow recovery is observed. The material in most cases does not recover to the original shape and permanent set remains. The magnitude of the permanent set depends upon length of time, amount of stress applied, and temperature.

The creep values are obtained by applying constant load to the test specimen in tension, compression, or flexure, and measuring the deformation as a function of time. The values are most commonly referred to as tensile creep, compressive creep, and flexural creep.

43.2.3.1 Tensile Creep Tensile creep measurements are made by applying the constant load to a tensile test specimen and measuring its extension as a function of time. The extension measurement can be carried out several different ways. The simplest way is to make two gauge marks on the tensile specimen and measure the distance between the marks at specified time intervals. The percent creep strain is determined by dividing the extension by initial gauge length and multiplying by 100. The percent creep strain is plotted against time to obtain a tensile creep curve. The tensile stress values are also determined at specified time intervals to facilitate plotting a stress–rupture curve. The more accurate measurements require the use of a strain gauge, which is capable of measuring and amplifying small changes in length with time and directly plotting them on a chart paper. Figure 43.5 illustrates a typical setup for tensile creep testing. The test is also carried out at different stress levels and temperatures to study their effect on tensile creep properties.

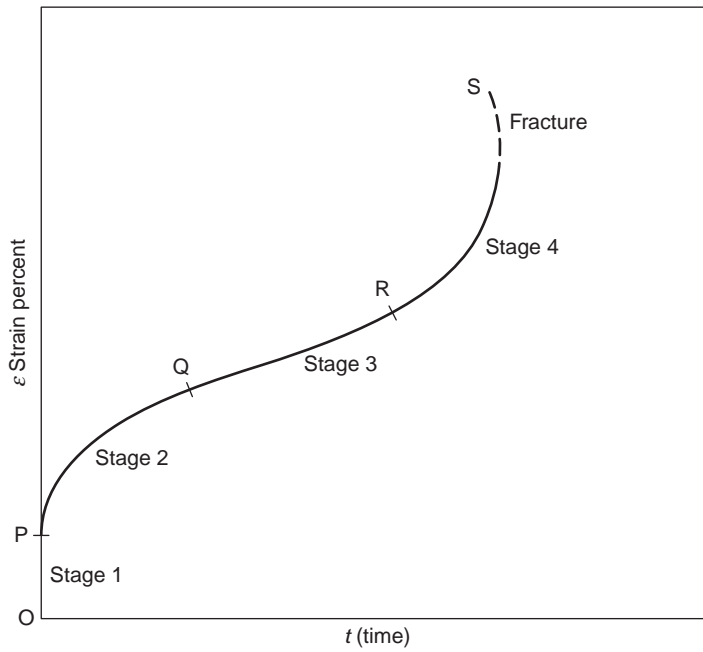


FIGURE 43.4 Generalized creep curve.



FIGURE 43.5 Typical test setup for tensile creep testing (Courtesy Applied Test Systems Inc.).



FIGURE 43.6 Flexural creep testing (Courtesy Ceast U.S.A. Inc.).

43.2.3.2 Flexural Creep Flexural creep measurements are also made by applying a constant load to the standard flexural test specimen and measuring its deflection as a function of time. A typical test setup for measuring creep in flexure is shown in Figure 43.6. As illustrated, the deflection of the specimen at mid-span is measured using a dial indicator gauge. The electrical resistance gauges may also be used in place of a dial indicator. The deflections of the specimen are measured at a predetermined time interval.

43.2.4 Stress Relaxation

Stress relaxation is defined as a gradual decrease in stress with time, under a constant deformation (strain). This characteristic behavior of the polymers is studied by applying a fixed amount of deformation to a specimen and measuring the load required to maintain it as a function of time. This is in contrast to creep measurement, where a fixed amount of load is applied to a specimen and resulting deformation is measured as a function of time.

Stress relaxation behavior of the polymers has been overlooked by many design engineers and researchers, partly because the creep data is much easier to obtain and is readily available. However, many practical applications dictate the use of stress relaxation data. For example, extremely low stress relaxation is desired in the case of a threaded bottle closure, which may be under constant strain for a long period. If the plastic material used in the closures shows an excessive decrease in stress under this constant deformation, the closures will eventually fail. Similar problems can be encountered with metal inserts in molded plastics, Belleville or multiple cantilever springs used in cameras, appliances, and business machines.

Stress relaxation measurements can be carried out using a tensile testing machine such as that described earlier in this chapter. However, the use of such a machine is not always

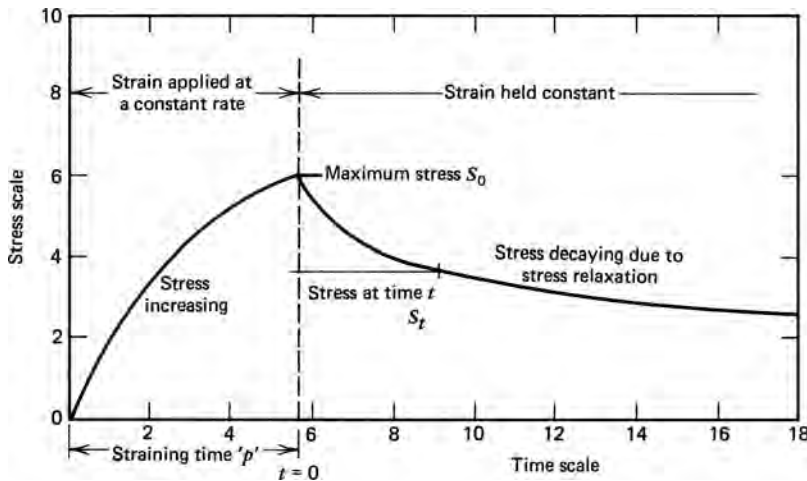


FIGURE 43.7 Stress–time curve (Courtesy Instron Corporation).

practical because the stress relaxation test ties up the machine for a long period of time. The equipment for a stress relaxation test must be capable of measuring very small elongation accurately, even when applied at high speeds. Many sophisticated pieces of equipment now employ a strain gauge or a differential transformer along with a chart recorder capable of plotting stress as a function of time. A typical stress–time curve is schematically plotted in Figure 43.7. At the beginning of the experiment, the strain is applied to the specimen at a constant rate to achieve the desired elongation. Once the specimen reaches the desired elongation, the strain is held constant for a predetermined amount of time. The stress decay, which occurs due to stress relaxation, is observed as a function of time. If a chart recorder is not available, the stress values at different time intervals are recorded and the results are plotted to obtain a stress versus time curve. The stress relaxation experiment is often carried out at various levels of temperature and strain.

The stress data obtained from the stress relaxation experiment can be converted to a more meaningful apparent modulus data by simply dividing stress at a particular time by the applied strain. The curve may be re-plotted to represent apparent modulus as a function of time. The use of logarithmic coordinates further simplifies the stress relaxation data by allowing us to use standard extrapolation methods such as the one used in creep experiments.

43.2.5 Impact Properties

The impact properties of the polymeric materials are directly related to the overall toughness of the material. Toughness is defined as the ability of the polymer to absorb applied energy. The area under the stress–strain curve is directly proportional to the toughness of a material. Impact strength is a measure of toughness. The higher the impact strength of a material, the higher the toughness and vice versa. Impact resistance is the ability of a material to resist breaking under a shock loading or the ability to resist the fracture under stress applied at high speed.

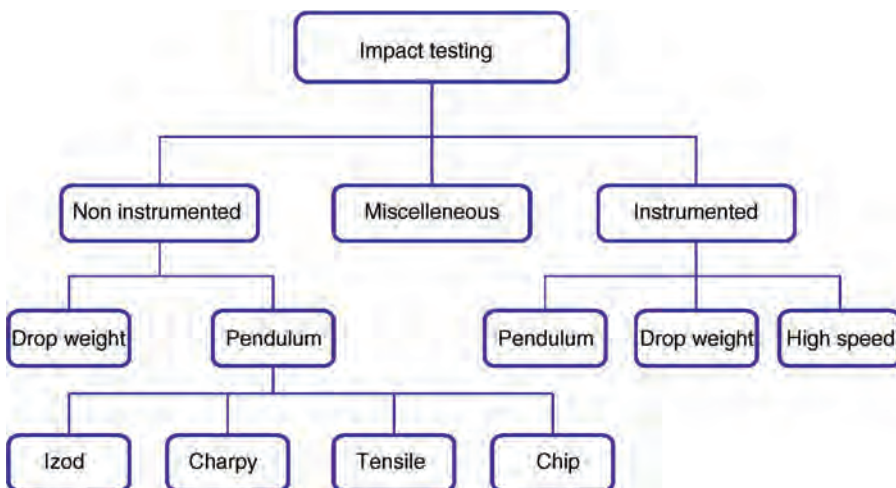
The theory behind toughness and brittleness of the polymers is very complex and therefore difficult to understand. The molecular flexibility plays an important role in determining the relative brittleness or toughness of the material. For example, in stiff polymers like

polystyrene and acrylics, the molecular segments are unable to disentangle and respond to the rapid application of mechanical stress and the impact produces brittle failure. In contrast, flexible polymer such as plasticized PVC has high-impact strength due to the ability of the large segments of molecules to disentangle and respond rapidly to mechanical stress.

Most polymers, when subjected to the impact loading, seem to fracture in a characteristic fashion. The crack is initiated on a polymer surface due to the impact loading. The energy to initiate such a crack is called the crack initiation energy. If the load exceeds the crack initiation energy, the crack continues to propagate. A complete failure occurs when the load has exceeded the crack propagation energy. Thus, both crack initiation and crack propagation contribute to the measured impact strength.

Impact strength is one of the most widely specified mechanical properties of the polymeric materials. However, it is also one of the least understood properties. Predicting the impact resistance of plastics still remains one of the most troublesome areas of product design. One of the problems with some earlier Izod and Charpy impact tests was that the tests were adopted by the plastic industry from metallurgists. The principles of impact mechanisms as applied to metals do not seem to work satisfactorily with plastics because of the plastics' complex structure.

43.2.5.1 Types of Impact Tests In the last two decades, a tremendous amount of time and money have been spent on research and development of various types of impact tests by organizations throughout the world. Attempts have been made to develop different sizes and shapes of specimens as well as impact testers. The specimens have been subjected to a variety of impact loads including tensile, compression, bending, and torsion impacts. Impact load has been applied using everything from a hammer, punches, and pendulums to falling balls and bullets. Unfortunately, very little correlation exists, if any, between the types of tests developed so far. Numerous technical papers and articles have been written on the subject of the advantages of one method over the other. To this date, no industry-wide consensus exists regarding an ideal impact test method. In this chapter, an attempt is made to discuss as many types of impact tests as possible along with the respective advantages and limitations of each test. Impact testing is divided into three major classes and subdivided into several classes as follows:



43.2.6 Izod–Charpy Impact Test (ASTM D-256, ISO 179)

The objective of the Izod–Charpy impact test is to measure the relative susceptibility of a standard test specimen to the pendulum-type impact load. The results are expressed in terms of kinetic energy consumed by the pendulum in order to break the specimen. The energy required to break a standard specimen is actually the sum of energies needed to deform it, to initiate its fracture, and to propagate the fracture across it, and the energy expended in tossing the broken ends of the specimen. This is called the “toss factor.” The energy lost through the friction and vibration of the apparatus is minimal for all practical purposes and usually neglected.

The specimen used in both tests is usually notched. The reason for notching the specimen is to provide a stress concentration area that promotes a brittle rather than a ductile failure. A plastic deformation is prevented by such type of notch in the specimen. The impact values are seriously affected because of the notch sensitivity of certain types of plastic materials. Figure 43.8 illustrates a typical pendulum-type impact testing machine.

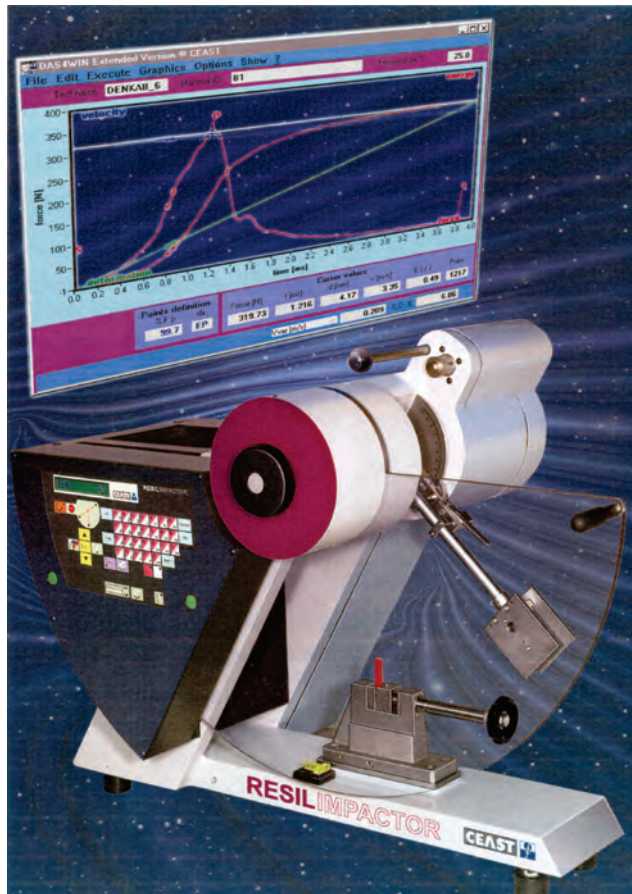


FIGURE 43.8 Pendulum impact tester (Courtesy Ceast U.S.A. Inc.).

The Izod test requires a specimen to be clamped vertically as a cantilever beam. The specimen is struck by a swing of a pendulum released from a fixed distance from the specimen clamp. A similar setup is used for the Charpy test except for the positioning of the specimen. In the Charpy method, the specimen is supported horizontally as a simple beam and fractured by a blow delivered in the middle by the pendulum. The obvious advantage of the Charpy test over the Izod test is that the specimen does not have to be clamped and, therefore, it is free of variations in clamping pressures.

43.2.6.1 Izod Impact Test The test specimen is clamped into position so that the notched end of the specimen is facing the striking edge of the pendulum. The pendulum hammer is released, allowed to strike the specimen, and swing through. If the specimen does not break, more weights are attached to the hammer and the test is repeated until failure is observed. The impact values are read directly in inch pound-force or feet pound-force from the scale. The impact strength is calculated by dividing the impact values obtained from the scale by the thickness of the specimen. For example, if a reading of 2 ft lbf is obtained using a 1/8 in. thick specimen, the impact value would be 16 ft lbf/in. of notch. The impact values are always calculated on the basis of 1 in. thick specimens even though much thinner specimens are usually used. The unnotched impact strength is obtained by reversing the position of a notched specimen in the vise. In this case, the notch is subjected to compressive rather than tensile stresses during impact. As discussed earlier in this chapter, the energy required to break a specimen is the sum of the energies needed to deform it, initiate and propagate the fracture, and toss the broken end (toss factor).

43.2.6.2 Falling-Weight Impact Test The falling-weight impact test, also known as the drop impact test or the variable-height impact test, employs a falling weight. This falling weight may be a tup with a conical nose, a ball, or a ball-end dart. The energy required to fail the specimen is measured by dropping a known weight from a known height onto a test specimen. The impact energy is normally expressed in feet pound-force and is calculated by multiplying the weight of the projectile by the drop height.

The biggest advantage of the falling-weight impact test over the pendulum impact test or high-rate tension test is its ability to duplicate the multidirectional impact stresses that a part would be subjected to in actual service. The other obvious advantage is the flexibility to use specimens of different sizes and shapes, including an actual part. Unlike the notched Izod impact test which measures the notch sensitivity of the material and not the material toughness, falling-weight impact tests introduce polyaxial stresses into the specimen and measure the toughness. The variations in the test results due to the fillers and reinforcements, clamping pressure, and material orientation are virtually eliminated in the falling-weight impact test. This type of test is also very suitable for determining the impact resistance of plastic films, sheets, and laminated materials. This falling-weight impact test is primarily designed to determine the relative ranking of materials according to the energy required to break flat rigid plastic specimens under various conditions of impact of a striker impacted by a falling weight. A free-falling weight or a tup is used to determine the impact strength of the material.

Many different versions of test equipment exist today. They all basically operate on the same principle. Figure 43.9 illustrates one such typical commercially available testing machine. It consists of a cast aluminum base, a slotted vertical guide tube, a round-nose striker and striker holder, an 8 lb weight, a die and die support, and a sample platform. The sample platform is used to position a sheet of desired thickness for impact testing.



FIGURE 43.9 Drop impact tester (Courtesy Byk-Gardner USA).

The die is removable from the base in order that the actual parts of complex shapes can be placed onto the base and impact tested.

The test is carried out by raising the weight to a desired height manually or automatically with the use of a motor-driven mechanism and allowing it to fall freely onto the other side of the striker. The striker transfers the impact energy to the flat test specimen positioned on a cylindrical die or a part lying on the base of the machine. The kinetic energy possessed by the falling weight at the instant of impact is equal to the energy used to raise the weight to the height of drop and is the potential energy possessed by the weight as it is released. Since the potential energy is expressed as the product of weight

and height, the guide tube can be marked with a linear scale representing the impact range of the instrument in inch pound. Thus, the toughness or the impact strength of a specimen or a part can be read directly off the calibrated scale in inch pound-force. The energy loss due to the friction in the tube or due to the momentary acceleration of the punch is negligible.

An alternate method for achieving the same result utilizes an instrument that employs a free-falling dart dropped from a specified height onto a test specimen. The dart with a hemispherical head is constructed of smooth, polished aluminum or stainless steel. An electromagnetic, air-operated or other mechanical release mechanism with a centering device is used for releasing the dart. The dart is also fitted with a shaft long enough to accommodate removable incremental weights. A two-piece annular specimen clamp is used to hold the specimen.

43.2.6.3 Instrumented Impact Testing One of the biggest drawbacks of the conventional impact test methods is that it provides only one value—the total impact energy—and nothing else. The conventional tests cannot provide additional information on the ductility, dynamic toughness, fracture, and yield loads or the behavior of the specimen during the entire impact event.

This effectively limits the application of noninstrumented impact test methods to quality control and material ranking. Instrumented impact testers are generally suited for research and development as well as advance quality control.

Instrumented impact testers measure force continuously while the specimen is penetrated. The resulting data can be used to determine type of failure and maximum load, in addition to the amount of energy required to fracture the specimen. One of the most common types of failures occurring from ductile to brittle transition at low temperatures can only be observed by studying load–energy–time curve. The fracture mode of a plastic is sensitive to the changes in temperature, and can change abruptly at or near the materials transition temperature. Manufacturers of plastic automotive components routinely test materials at low temperatures (-20 to -40°F) to assure that they will not become brittle in cold weather. By studying the shape of the load–time or load–deflection curve, the type of failure can be analyzed, and important information about its performance in service can be gathered. The new piezoelectric equipped strikers offer increased sensitivity, opening the doors for testing a whole new range of materials. Applications involving light-weight products such as foam containers for eggs and ultra-thin films used in packaging industry can now be meaningfully tested.

All standard impact testers can be instrumented to provide a complete load and energy history of the specimen. Such a system monitors and precisely records the entire impact event, starting from the acceleration (from rest) to the initial impact and plastic bending to fracture initiation and propagation to the complete failure. The instrumentation is done by mounting the strain gauges or load cell onto the striking bit in the case of pendulum impact tester or onto the tup in the case of a drop impact tester. During the test, a fiber optic device triggers an oscilloscope just before striking the specimen. The output of the strain gauge is recorded by the oscilloscope depicting the variation of the load applied to the specimen throughout the entire fracturing process. A complete load–time history of the entire specimen is obtained. The apparent total energy absorbed by the specimen can be calculated and plotted against time. The specimen displacement can be calculated by the double integration of the load–time curve and the load–displacement curve can be plotted. With the advent of microprocessor technology, some manufacturers are now



FIGURE 43.10 Instrumented impact tester (Courtesy Instron Corporation).

capable of offering a unit that automatically calculates the sample displacement and provide a load–displacement curve eliminating the need for calculations. Many other useful data such as the impact rate, force, and displacement at yield, and break, yield, and failure energies as well as modulus are calculated and printed out. A commercially available instrumented impact tester is shown in Figure 43.10.

43.2.7 High-Speed Impact Tests (ASTM D3763, ISO 6603-2)

An ever-increasing demand for engineering plastics and the need for sophisticated and meaningful impact test methods for characterizing these materials have forced the industry into developing new high-speed impact tests. These tests not only provide the basic information regarding the toughness of the polymeric materials but also provide other important data of interest, such as the load–deflection curve and the total energy absorption. The high-speed impact test overcomes the basic limitations of conventional impact

testing methods as discussed previously. The rate of impact can be varied from 30 to 570,000 in./min.

High speed impact testing has gained considerable popularity in recent years because of its ability to simulate actual impact failures at high speeds. For example, conventional impact testing methods are useless in testing to meet advanced automotive crash standards that require the impact simulation at 28 mph (30,000 in./min). High-speed impact testers are able to meet the challenge. As discussed earlier in this chapter, almost all polymers are strain-rate sensitive. Two polymers, when impact tested at one strain rate, may show similar impact strength values. The same two polymers tested at a high strain rate show a completely different set of values.

Most versatile high-speed impact testing machine is capable of testing everything from the thin film which may require as low an impact rate as 30 in./min to the plastic automotive bumper which may require a high impact rate up to 30,000 in./min. The specimen or product can be tested under a controlled environment of temperature and humidity. The equipment basically consists of a tup attached to a motor wound spring or pneumatically powered actuator along with plunger displacement measuring system. The force is detected with a fast responding quartz load cell mounted directly on the actuator. The velocity can be set digitally from 30 to 30,000 in./min. Some type of clamp assembly to hold the specimen in place is used. The equipment can also be fitted with environmental chamber for specialized testing. The tester is equipped with a monitor and x - y plotter that automatically displays load versus displacement data. A built-in microprocessor provides more useful information, such as modulus, yield, and failure energies.

High-speed impact testers have been proven very useful in material evaluation. At low strain rates, some polymers fail in ductile manner. The same polymers appear to show brittle failure at high strain rates. The point at which this ductile to brittle transition takes place is of particular importance. A high-rate impact test can provide such information in a graphical form. Tests can also be carried out at different temperatures to find ductile-brittle transition points at various temperatures. Other useful applications of the high-speed impact tester include the process quality control, design evaluation, and assembly evaluation.

43.2.7.1 Abrasion Resistance Tests The material's ability to resist abrasion is most often measured by its loss in weight when abraded with an abraser. The most widely accepted abraser in the industry is called the Taber abraser. A variety of wheels with varying degree of abrasiveness is available. The grade of "calibrase" wheel designated CS-17 with 1000-g load seems to produce satisfactory results with almost all plastics. For softer materials less abrasive wheels with smaller load on the wheels may be used. The test specimen is usually a 4 in. diameter disc or a 4 sq. in. plate having both surfaces substantially plane and parallel. A 1/2 in. diameter hole is drilled in the center. Specimens are conditioned employing standard conditioning practices prior to testing. To commence testing, the test specimen is placed on a revolving turntable. Suitable abrading wheels are placed on the specimen under certain set dead weight loads. The turntable is started and an automatic counter records the number of revolutions. Most tests are carried out to at least 5000 revolutions. The specimens are weighed to the nearest mg. The test results are reported as weight loss in milligrams/1000 cycles. The grade of abrasive wheel along with amount of load at which the test was carried out is always reported along with results.

Test methods such as ASTM D 1044 (resistance of transparent plastic materials to abrasion) are also developed for estimating the resistance of transparent plastic materials



FIGURE 43.11 Abrasion tester (Courtesy Taber Industries).

to one kind of abrasion by measurement of its optical affects. The test is carried out in similar manner to that described above, except that 100 cycles with a 500 g. load is normally used. A photoelectric photometer is used to measure the light scattered by abraded track. The percentage of the transmitted light that is diffused by the abraded specimens is reported as a test result. Figure 43.11 illustrates a commercially available abrasion tester.

43.2.8 Fatigue Resistance

The behavior of materials subjected to repeated cyclic loading in terms of flexing, stretching, compressing, or twisting is generally described as fatigue. Such repeated cyclic loading eventually constitutes a mechanical deterioration and progressive fracture that leads to complete failure. Fatigue life is defined as the number of cycles of deformation required to bring about the failure of the test specimen under a given set of oscillating conditions.

The failures that occur from repeated application of stress or strain are well below the apparent ultimate strength of the material. Fatigue data are generally reported as the number of cycles to fail at a given maximum stress level. The fatigue endurance curve, which represents stress versus number of cycles to failure, also known as $S-N$ curve, is generated by testing a multitude of specimens under cyclic stress, each one at different stress levels. At high stress levels, materials tend to fail at relatively low numbers of cycles. At low stresses, the materials can be stressed cyclically for an indefinite number of times and the failure point is virtually impossible to establish. This limiting stress below which material will never fail is called the fatigue endurance limit. The fatigue endurance limit can also be defined as the stress at which the $S-N$ curve becomes asymptotic to the horizontal (constant stress) line. For most polymers, the fatigue endurance limit is between 25% and 30% of the static tensile strength. The fatigue resistance data are of practical importance in the design of gears, tubing, hinges, parts on vibrating machinery, and pressure vessels under cyclic pressures.

Two basic types of tests have been developed to study the fatigue behavior of plastic materials:

1. flexural fatigue test,
2. tensile fatigue test.

43.2.8.1 Flexural Fatigue Test (ASTM D 671) The ability of a material to resist deterioration from cyclic stress is measured in this test by using a fixed cantilever-type testing machine capable of producing a constant-amplitude-of-force on the test specimen each cycle. The main feature of a fatigue testing machine is an unbalanced, variable eccentric, mounted on a shaft that is rotated at a constant speed by a motor. This unbalanced movement of an eccentric produces alternating force. The specimen is held as a cantilever beam in a vice at one end and bent by a concentrated load applied through a yoke fastened to the opposite end. A counter is used to record the number of cycles along with a cutoff switch to stop the machine when the specimen fails. A typical commercially available fatigue testing machine is illustrated in Figure 43.12. The test specimen of two different geometries are used. If machined specimens are used, care must be taken to eliminate all scratches and tool marks from the specimens. Molded specimen must be stress-relieved before using.

The test is carried out by first determining the complementary mass and effective mass of the test specimen. The load required to produce the desired stress is calculated from these values. The number of cycles required to produce failure is determined. The test is repeated at varying stress levels. A curve of stress versus cycles-to-failure ($S-N$ diagram) is plotted from the test results.

43.2.8.2 Tensile Fatigue Test Unlike the flexural fatigue test which uses the constant deflection (strain) principle, the tensile fatigue test is conducted under constant load (stress) conditions. The specimen is dumbbell-shaped, about 2 in. long with a cylindrical cross section.

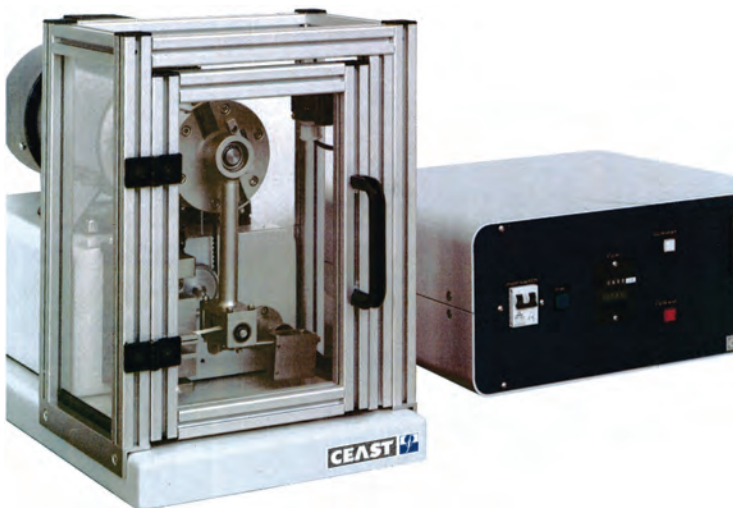


FIGURE 43.12 Flexural fatigue tester (Courtesy Ceast U.S.A. Inc.).

The test is conducted by mounting both ends of the dumbbell specimen in the testing machine. The specimen is rotated between two spindles, and stress in the form of tension and compression is applied. The specimen is subjected to the number of cycles of stress specified or until fracture occurs.

43.2.9 Hardness Tests

43.2.9.1 Rockwell Hardness (ASTM D 785) The Rockwell hardness test measures the net increase in depth impression as the load on an indenter is increased from a fixed minor load to a major load and then returned to a minor load. The hardness numbers derived are just numbers without units. Rockwell hardness numbers are always quoted with a scale symbol representing the indenter size, load, and dial scale used. The hardness scales in order of increasing hardness are *R*, *L*, *M*, *E*, and *K* scales. The higher the number in each scale, the harder the material. There is a slight overlap of hardness scales and, therefore, it is quite possible to obtain two different dial readings on different scales for the same material. For a specific type of material, correlation in the overlapping regions is possible. However, due to differences in elasticity, creep, and shear characteristics between different plastics, a general correlation is not possible.

43.2.9.2 Durometer Hardness (ASTM D 2240, ISO 868) The Durometer hardness test is mostly used for measuring the relative hardness of soft materials. The test method is based on the penetration of a specified indenter forced into the material under specified conditions.

The Durometer hardness tester consists of a pressure foot, an indenter, and an indicating device. The indenter is spring loaded and the point of the indenter protrudes through the hole in the base. The test specimens are at least 1/4 in. thick and can be either molded or cut from a sheet. Several thin specimens may be piled to form a 1/4 in. thick specimen but one piece specimens are preferred. The poor contact between the thin specimens may cause results to vary considerably.

The test is carried out by first placing a specimen on a hard, flat surface. The pressure foot of the instrument is pressed onto the specimen, making sure that it is parallel to the surface of the specimen. The durometer hardness is read within 1 sec after the pressure foot is in firm contact with the specimen.

Two types of durometers are most commonly used—Types A and D. The basic difference between the two types is the shape and dimension of the indenter. The hardness numbers derived from either scale are just numbers without any units. Type A durometer is used with relatively soft material while Type D is used with slightly harder material.

43.3 THERMAL PROPERTIES

Thermal properties of plastic materials are equally as important as mechanical properties. Unlike metals, plastics are extremely sensitive to changes in temperature. The mechanical, electrical, or chemical properties of plastics cannot be looked at without looking at the temperature at which the values were derived. Crystallinity has a number of important effects upon the thermal properties of a polymer. Its most general effects are the introduction of a sharp melting point and the stiffening of thermal mechanical properties. Amorphous plastics, in contrast, have a gradual softening range. Molecular orientation also has

a significant effect on thermal properties. Orientation tends to decrease dimensional stability at higher temperatures. The molecular weight of the polymer affects the low-temperature flexibility and low-temperature brittleness. Many other factors such as intermolecular bonding, cross linking, and copolymerization all have a considerable effect on thermal properties. From the above discussion, it is very clear that the thermal behavior of polymeric materials is rather complex. Therefore, in designing a plastic part or selecting a plastic material from the available thermal property data, one must thoroughly understand the short term as well as the long-term effect of temperature on properties of that plastic material.

43.3.1 Tests for Elevated Temperature Performance

Designers and material selectors of plastic products constantly face the challenge of selecting a suitable plastic for elevated temperature performance. The difficulty arises due to the varying natures and capabilities of various types and grades of plastics at elevated temperatures. Many factors are considered when selecting a plastic for a high-temperature application. The material must be able to support a design load under operating conditions without objectionable creep or distortion. The material must not degrade or lose necessary additives that will cause drastic reduction in the physical properties during the expected service life.

All the properties of plastic materials are not affected in a similar manner by elevating temperature. For example, electrical properties of a particular plastic may show only a moderate change at elevated temperatures, while the mechanical properties may be reduced significantly. Also, since the properties of plastic materials vary with temperature in an irregular fashion, they must be looked at as a function of temperature in order to obtain more meaningful information. From the foregoing, it is quite clear that a single maximum use temperature that will apply to all the important properties in high-temperature applications is simply not possible.

One of the most important considerations while studying the performance of plastics at elevated temperatures is the dependence of key properties such as modulus, strength, chemical resistance, and environmental resistance on time. Therefore, the short-term heat resistance data alone is not adequate for designing and selecting materials that require long-term heat resistance. For the sake of convenience and simplicity, we divide the elevated temperature effects into two categories:

1. short-term effects:
 - a. heat deflection temperature,
 - b. vicat softening temperature,
 - c. torsion pendulum,
2. long-term effects:
 - a. long-term heat resistance test,
 - b. UL temperature index,
 - c. creep modulus/creep rupture tests.

43.3.1.1 Heat Deflection Temperature (HDT) (ASTM D 648, ISO 75-1 & 75-2) Heat deflection temperature is defined as the temperature at which a standard test bar ($5 \times 1/2 \times 1/4$ in.) deflects 0.010 in. under a stated load of either 66 or 264 psi. The heat

deflection temperature test, also referred to as the heat distortion temperature test, is commonly used for quality control and for screening and ranking materials for short-term heat resistance. The data obtained by this method cannot be used to predict the behavior of plastic materials at elevated temperature nor can it be used in designing a part or selecting and specifying material. Heat deflection temperature is a single point measurement and does not indicate long-term heat resistance of plastic materials. Heat distortion temperature, however, does distinguish between those materials that lose their rigidity over a narrow temperature range and those that are able to sustain light loads at high temperatures.

The apparatus for measuring heat deflection temperature consists of an enclosed oil bath fitted with a heating chamber and automatic heating controls that raise the temperature of the heat transfer fluid at a uniform rate. A cooling system is also incorporated to fast cool the heat transfer medium for conducting repeated tests. The specimens are supported on steel supports, 4 in. apart with the load applied on top of the specimen vertically and midway between the supports. The contact edges of the support and of the piece by which pressure is applied is rounded to a radius of 1/4 in. A suitable deflection measurement device, such as a dial indicator, is normally used. A mercury thermometer is used for measuring temperature. The unit is capable of applying 66 or 264 psi fiber stress on specimens by means of a dead weight. A commercially available heat deflection measuring device with a close-up of a specimen holder is illustrated in Figure 43.13.

The specimen is positioned in the apparatus along with the temperature and deflection measuring devices and the entire assembly is submerged into the oil bath kept at room temperature. The load is applied to a desired value (66 or 264 psi fiber stress). Five minutes after applying the load, the pointer is adjusted to zero and the oil is heated at the rate of $2 \pm 0.2^\circ\text{C}/\text{min}$. The temperature of the oil at which the bar has deflected 0.010 in. is recorded as the heat deflection temperature at the specified fiber stress.

43.3.1.2 Long-Term Effects The long-term effects of elevated temperature on properties of plastics are extremely important, especially when one considers the fact that the majority of applications involving high heat are long-term applications. During long-term



FIGURE 43.13 DTUL/Vicat tester (Courtesy Ceast U.S.A. Inc).

exposure to heat, plastic materials may encounter many physical and chemical changes. A plastic material that shows little or no effect at elevated temperature for a short time may show a drastic reduction in physical properties, a complete loss of rigidity, and severe thermal degradation when exposed to elevated temperature for a long time. Along with time and temperature, many other factors such as ozone, oxygen, sunlight, and pollution combine to accelerate the attack on plastics. At elevated temperatures, many plastics tend to lose important additives such as plasticizers and stabilizers, causing plastics to become brittle or soft and sticky.

Three basic tests have been developed and accepted by the plastics industry. If the application does not require the product to be exposed to elevated temperature for a long period under continuous load, a simple heat-resistance test is adequate. The applications requiring the product to be under continuous significant load must be looked at from creep modulus and creep rupture strength test data. Yet, another one of the most widely accepted methods of measuring maximum continuous use temperature has been developed by Underwriters Laboratories. The UL temperature index, established for a variety of plastic materials to be used in electrical applications, is the maximum temperature that the material may be subjected to without fear of premature thermal degradation.

43.3.1.3 Long-Term Heat-Resistance Test (ASTM D 794) The long-term heat-resistance test was developed to determine the permanent effect of heat on any property by selection of an appropriate test method and specimen. In ASTM recommended practice, only the procedure for heat exposure is specified and not the test method or specimen.

Any specimen, including sheet, laminate, test bar, or molded part may be used. If a specific property, such as tensile strength loss is to be determined, a standard tensile test bar specimen and procedures must be used for comparison of test results before and after the test. The test requires the use of a mechanical convection oven with a specimen rack of suitable design to allow air circulation around the specimens. The test is carried out by simply placing the specimen in the oven at a desired exposure temperature for a predetermined length of time. The subsequent exposure to temperatures may be increased or decreased in steps of 25°C until a failure is observed. Failure due to heat is defined as a change in appearance, weight, dimension, or other properties that alter plastic material to a degree that it is no longer acceptable for the service in question. Failure may result from blistering, cracking, loss of plasticizer, or other volatile material that may cause embrittlement, shrinkage, or change in desirable electrical or mechanical properties.

There are many factors that affect the reproducibility of the data. The degree of temperature control in the oven, the type of molding, cure, air velocity over the specimen, period of exposure, and humidity of the oven room are some of these factors. The amount and type of volatiles in the molded part or specimen may also affect the reproducibility.

43.4 ELECTRICAL PROPERTIES

The unbeatable combination of characteristics such as ease of fabrication, low cost, light weight, and excellent insulation properties have made plastics one of the most desirable materials for electrical applications. Although, the majority of applications involving plastics are insulation-related, plastics can be made to conduct electricity by simply modifying the base material with proper additives such as carbon black.

Until recently, plastics were considered a relatively weaker material in terms of load-bearing properties at elevated temperatures. Therefore, the use of plastics in electrical applications was limited to nonload-bearing, general-purpose applications. The advent of new high performance engineering materials has altered the entire picture. Plastics are now specified in a majority of applications requiring resistance to extreme temperatures, chemicals, moisture, and stresses. The primary function of plastics in electrical applications has been that of an insulator. This insulator or dielectric separates two field-carrying conductors. Such a function can be served equally well by air or vacuum. However, neither air nor vacuum can provide any mechanical support to the conductors. Plastics not only act as effective insulators but also provide mechanical support for field-carrying conductors. For this very reason, the mechanical properties of plastic materials used as insulators become very important. Typical electrical applications of plastic material include plastic-coated wires, terminals, connectors, industrial and household plugs, switches, and printed circuit boards.

The key electrical properties of interest are dielectric strength, dielectric constant, dissipation factor, volume and surface resistivity, and arc resistance.

43.4.1 Dielectric Strength (ASTM D 149, IEC 243-1)

The dielectric strength of an insulating material is defined as the maximum voltage required to produce a dielectric breakdown. Dielectric strength is expressed in volts per unit of thickness such as volt per mil. All insulators allow a small amount of current to leak through or around themselves. Only a perfect insulator, if there is such an insulator in existence, can be completely free from small current leakage. The small leakage generates heat, providing an easier access to more current. The process slowly accelerates with time and the amount of voltage applied until a failure in terms of dielectric breakdown or what is known as puncture occurs. Obviously, dielectric strength, which indicates electrical strength of a material as an insulator, is a very important characteristic of an insulating material. The higher the dielectric strength, the better the quality of an insulator. Three basic procedures have been developed to determine dielectric strength of an insulator. Figure 43.14 illustrates the basic setup for a dielectric strength test. A variable transformer and a pair of electrodes are normally employed. Specimens of any desirable thickness prepared from the material to be tested are used. Specimen thickness of 1/16 in. is fairly common. The first procedure is known as the short-times method. In this method, the voltage is increased from zero to breakdown at uniform rate.

The second method is known as the slow-rate-of-rise method. The test is carried out by applying the initial voltage approximately equal to 50% of the breakdown voltage as determined by the short-time test or as specified. Next, the voltage is increased at a uniform rate until the breakdown occurs.

43.4.2 Dielectric Constant and Dissipation Factor (ASTM D150, IEC 250)

43.4.2.1 Dielectric Constant (Permittivity) Dielectric constant of an insulating material is defined as the ratio of the charge stored in an insulating material placed between two metallic plates to the charge that can be stored when the insulating material is replaced by air (or vacuum). Defined another way, the dielectric constant is the ratio of the capacitance by two metallic plates with an insulator placed between them and the capacitance of the same plates with a vacuum between them. Simply stated, the dielectric constant indicates the ability of an insulator to store electrical energy.

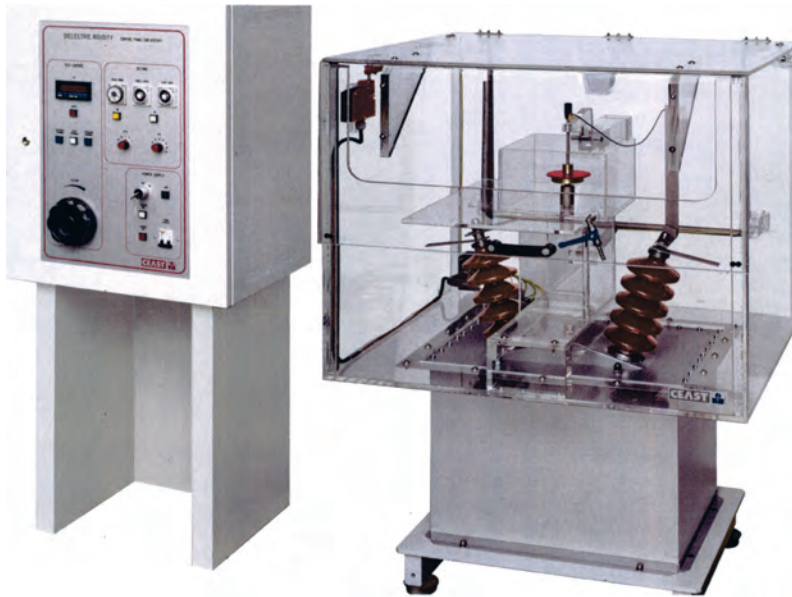


FIGURE 43.14 Dielectric strength tester (Courtesy Ceast U.S.A. Inc.).

The dielectric constant test is fairly simple. The test specimen is placed between the two electrodes, as shown in Figure 43.15, and the capacitance is measured. Next, the test specimen is replaced by air and once again, the capacitance value is measured. The dielectric constant value is determined from the ratio of the two measurements. Dielectric constant values are affected by factors such as frequency, voltage, temperature, humidity, and so on.

43.4.2.2 Dissipation Factor In all electrical applications, it is desirable to keep the electrical losses to a minimum. Electrical losses indicate the inefficiency of an insulator. The

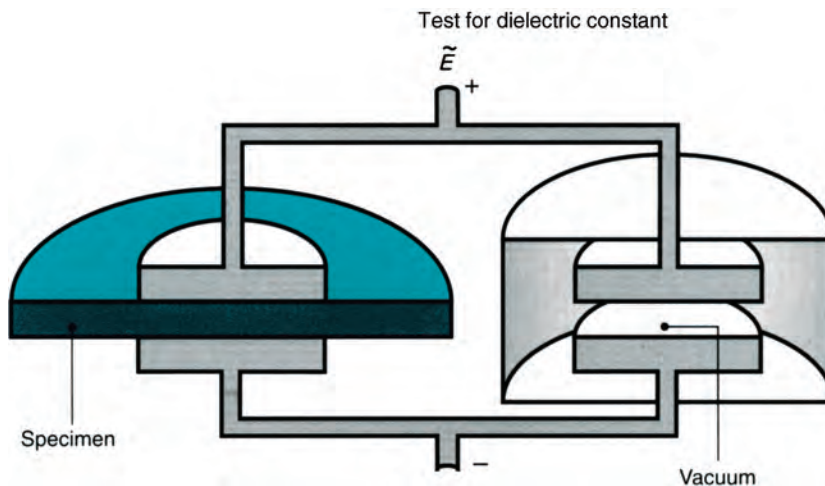


FIGURE 43.15 Schematic of dielectric constant test (Courtesy Bayer Corporation).

dissipation factor is a measure of such electrical inefficiency of the insulating material. The dissipation factor indicates the amount of energy dissipated by the insulating material when the voltage is applied to the circuit. The dissipation factor is defined as the ratio of the conductance of a capacitor in which the material is the dielectric to its susceptance or the ratio of its parallel reactance to its parallel resistance. Most plastics have a relatively lower dissipation factor at room temperature. However, at high temperatures, the dissipation factor is quite high, resulting in greater overall inefficiency in electrical system. Loss factor, which is the product of dielectric constant and the dissipation factor, is a frequently used term, since it relates to the total loss of power occurring in insulating materials.

43.4.3 Electrical Resistance Tests

As was stated earlier, the primary function of an insulator is to insulate current-carrying conductors from each other as well as from ground and to provide mechanical support for components. Naturally, the most desirable characteristic of an insulator is its ability to resist the leakage of the electrical current. The higher the insulation resistance, the better the insulator. Failure to recognize the importance of insulation resistance values while designing products such as appliances and power tools could lead to fire, electrical shock, and personal injury.

Insulation resistance can be subdivided into

1. volume resistance,
2. surface resistance.

Volume resistance is defined as the ratio of the direct voltage applied to two electrodes that are in contact with a specimen to that portion of the current between them that is distributed through the volume of the specimen. Or, simply stated, the volume resistance is the resistance to leakage through the body of the material. Volume resistance generally depends upon the material. The term most commonly used by designers is volume resistivity. It is defined as the ratio of the potential gradient parallel to the current in the material to the current density or simply stated, the volume resistivity of a material is the electrical resistance between the opposite faces of a unit cube for a given material and at a given temperature.

High volume resistivity materials are desirable in applications requiring superior insulating characteristics.

The surface resistance of a material is defined as the ratio of the direct voltage applied to the electrodes to that portion of the current between them that is primarily in a thin layer of moisture or other semiconducting material that may be deposited on the surface. Or simply stated, surface resistance is the resistance to leakage along the surface of an insulator. The surface resistance of a material depends upon the quality and cleanliness of the surface of the product. A product with oil or dirt particles on it gives lower surface resistance values. Temperature and humidity both seem to affect the insulation resistance appreciably.

43.4.4 ARC Resistance (ASTM D 495)

Arc resistance is the ability of a plastic material to resist the action of a high-voltage electrical arc, usually stated in terms of time required to form material electrically conductive. Failure is characterized by carbonization of the surface, tracking, localized heating to incandescence, or burning. In all applications in which conducting elements are

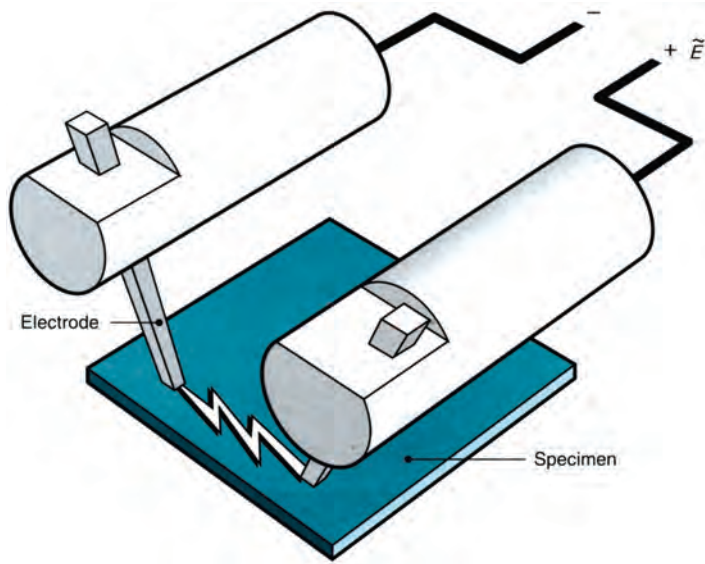


FIGURE 43.16 Arc resistance test (Courtesy Bayer Corporation).

brought into contact, arcing is inevitable. Switches, circuit breakers, and automotive distributor caps, are a few good examples of applications where arcing is known to cause failure. Another term that is generally associated with arcing is tracking. Tracking is defined as a phenomenon where a high voltage source current creates a leakage or fault path across the surface of an insulating material by slowly but steadily forming a carbonized path appearing as a thin, wiry line between the electrodes. Tracking is accelerated by the presence of surface contaminants such as dirt and oil and by the presence of moisture. Figure 43.16 illustrates a typical setup for an arc resistance test. The voltage is applied intermittently and severity is increased in steps until the failure occurs. Arc resistance is measured in seconds to failure.

43.5 WEATHERING PROPERTIES

The increased outdoor use of plastics has created a need for a better understanding of the effect of the environment on plastic materials. The environmental factors have significant detrimental effects on appearance and properties. The severity of the damage depends largely on factors such as the nature of the environment, geographic location, type of polymeric material, and duration of exposure. The effect can be anywhere from a mere loss of color or a slight crazing, and cracking to a complete breakdown of the polymer structure. Any attempt to design plastic parts without a clear understanding of the degradation mechanisms induced by the environment would result in a premature failure of the product.

43.5.1 UV Radiation

All types of solar radiation have some sort of detrimental effect on plastics. Ultraviolet radiation is the most destructive of all radiations. The energy in ultraviolet radiation is

sufficiently strong enough to break molecular bonds. This activity in the polymer brings about thermal oxidative degradation which results in embrittlement, discoloration, and an overall reduction in physical and electrical properties. Xenon ARC lamps, fluorescent lighting, sun lamps, and other artificial sources also emit a similar type of harmful radiation. Other factors in the environment such as heat, humidity, and oxygen accelerate the UV degradation process.

One of the best methods of protecting plastics against UV radiation is to incorporate UV absorbers or UV stabilizers into the plastic materials. The UV absorbers provide preferential absorption to most of the incident UV light and are able to dissipate the absorbed energy harmlessly. Thus, the polymer is protected from harmful radiation at the cost of UV absorbers which are destroyed in the process with time. Several types of organic and inorganic UV absorbers are developed for this purpose. Almost all inorganic pigments absorb UV radiation to a certain extent and provide some degree of protection. Perhaps the most effective pigments are certain types of carbon black that absorb over the entire range of UV and visible radiation and transform the energy into less harmful radiation.

UV stabilizers, unlike UV absorbers, inhibit the bond rupture by chemical means or dissipate the energy to lower levels that do not attack the bonds. The effectiveness of such additives when incorporated with the polymer can be determined by various test methods.

43.5.2 Accelerated Weathering Tests

Most data on the aging of plastics are acquired through accelerated tests and actual outdoor exposure. The latter being a time-consuming method, accelerated tests are often used to expedite screening the samples with various combinations of additive levels and ratios. A variety of light sources are used to simulate the natural sunlight. The artificial light sources include carbon arc lamps, xenon arc lamps, fluorescent sun lamps, and mercury lamps. These light sources, except fluorescent, are capable of generating a much higher intensity light than natural sunlight. Xenon arc lamps can be operated over a wide range from below peak sunlight to twice the sunlight levels. Quite often, a condensation apparatus is used to simulate the deterioration caused by sunlight and water as rain or dew. Modern instruments have direct specimen spray on front and/or back side of the specimen.

There are three major accelerated weathering tests:

1. exposure to carbon arc lamps,
2. Exposure to xenon arc lamps,
3. exposure to fluorescent UV lamps.

The xenon arc, when properly filtered, most closely approximates the wavelength distribution of natural sunlight.

43.5.2.1 Exposure of Plastics to Fluorescent UV Lamps and Condensation (ASTM G 53, ISO 4892) This method is meant to simulate the deterioration caused by sunlight and dew by means of artificial ultraviolet light and condensation apparatus (Figures 43.17 and 43.18). Solar radiation ranges from ultraviolet to infrared. Ultraviolet light of wavelengths between 290 and 350 nm is the most efficient portion of terrestrial sunlight that is damaging to plastics. In the natural sunlight spectrum, energy below 400 nm accounts for less than 6% of the total radiant energy (3). Since the special fluorescent UV lamps radiate between 280 and 365 nm, they



FIGURE 43.17 UV light and condensation apparatus (Courtesy Q-Panel Lab Products).

accelerate the degradation process considerably. In recent years, the UVA-340 lamps have increased in popularity because of the poor results of the conventional FS-40 lamps.

The test apparatus basically consists of a series of UV lamps, a heated water pan, and test specimen racks. The temperature and operating times are independently controlled

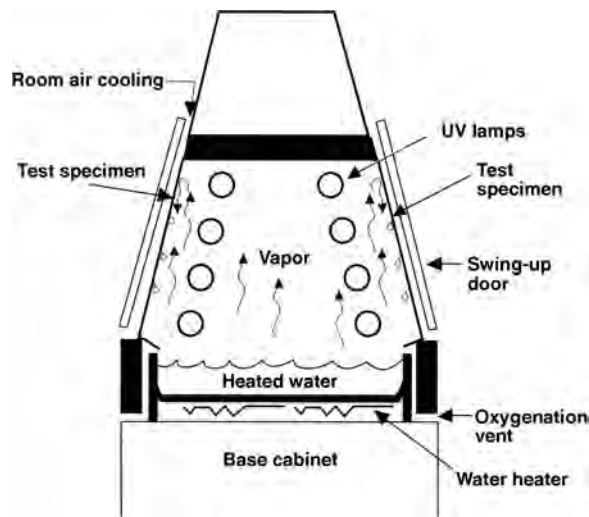


FIGURE 43.18 Cross section of a UV light and condensation apparatus (Courtesy Q-Panel Lab Products).

both for UV and the condensation effect. The test specimens are mounted in specimen racks with the test surfaces facing the lamp. The test conditions are selected based on requirements and programmed into the unit. The specimens are removed for inspection at a predetermined time to examine color loss, crazing, chalking, and cracking.

43.5.2.2 Exposure of Plastics to Carbon Arc-Type Light and Water (ASTM D 1499, ISO 4892) This method is very useful in determining the resistance of plastic materials when exposed to radiation produced by carbon-arc lamps. There are basically two different types of carbon-arc lamps used as the source of radiation. The first type is enclosed carbon arc lamp. The second type is known as an open-flame sunshine carbon-arc. Both apparatuses revolve around centrally mounted arc lamps. The provision is also made to expose the specimen to water, which is sprayed through nozzles. The light-on, light-off, and water-spray cycles are independent of each other and the apparatus can be programmed to operate with virtually any combination. A black panel temperature inside the test chamber can be monitored and controlled by a sensor mounted directly on the revolving specimen rack.

43.5.2.3 Exposure of Plastics to Xenon Arc-Type Light and Water (ASTM D 2565, G 26, ISO 4892) A filtered xenon arc-type light source is one of the most popular indoor exposure tests since it exhibits a spectral energy distribution of sunlight at the surface of the earth. The xenon-arc lamp consists of a burner tube and a light filter system consisting of interchangeable glass filters used in combination to provide a spectral distribution that approximates natural sunlight exposure conditions. The apparatus typically has a built-in recirculating system that recirculates distilled or deionized water through the lamp. The water cools the xenon burner and filters out long wavelength infrared energy. For air-cooled lamps, this is accomplished by the use of optical filters.

Two basic procedures are recommended. Procedure A is a normal operating procedure for comparative evaluation within a series exposed simultaneously in one instrument. Procedure B is used for comparing results among instruments. Both procedures are described in detail in the ASTM Standards Manual.

43.5.3 Outdoor Weathering of Plastics (ASTM D 1435)

The test is devised to evaluate the stability of plastic materials exposed outdoors to varied influences that comprise weather exposure conditions that are complex and changeable. Important factors are climate, time of year, and the presence of industrial atmosphere. It is recommended that repeated exposure testing at different seasons and over a period of more than one year be conducted to confirm exposure at any one location. Since weathering is a comparative test, control samples are always utilized and retained at standard conditions of temperature and humidity. The control samples must also be covered with inert wrapping to exclude light exposure during the aging period. However, dark storage does not insure stability.

Test sites are selected to represent various conditions under which the plastic product will be used. Arizona is often selected for intense sunlight, wide temperature cycle, and low humidity. Florida, on the other hand, provides high humidity, intense sunlight, and relatively high temperatures.

Exposure test specimens of suitable shape or size are mounted in a holder directly applied to the racks. Racks are positioned at a 45° angle and facing the equator. Many



FIGURE 43.19 Typical aluminum exposure racks (Courtesy Atlas Electric Devices Company).

other variations in the position of the racks are also employed, depending upon the requirements.

The specimens are removed from the racks after a specified amount of time and subjected to various tests such as appearance evaluation, electrical tests, and mechanical tests. The results are compared with the test results from testing control specimens. Typical aluminum exposure racks are shown in Figure 43.19.

43.6 OPTICAL PROPERTIES

43.6.1 Refractive Index (ASTM D 542)

Refractive index is a fundamental property of transparent materials. Refractive index values are very important to a design engineer involved in designing lenses for cameras, microscopes, and other optical equipment. The refractive index, also known as the index of refraction, is defined as the ratio of the velocity of light in a vacuum (or air) to its velocity in a transparent medium.

$$\text{Index of refraction} = \frac{\sin \text{ angle of incidence}}{\sin \text{ angle of refraction}}.$$

The Abbe refractometer is the refractometer most widely used to determine the index of refraction (Figure 43.20). The test also requires a source of white light and a contacting liquid that will not attack the surface of the plastic. The contacting liquid must also have a higher refractive index than the plastic being measured. A test specimen of any size may be used as long as it conveniently fits on the face of the fixed half of the refractometer prism. The surface of the specimen in contact with the prism must be flat and polished.

The test is carried out by placing a specimen in contact with the prism using a drop of contacting liquid. The polished edge of the specimen is kept towards the light source. The refractive index is determined by moving the index arm of the refractometer so that the field seen through the eyepiece is half dark. The compensator is adjusted to remove all



FIGURE 43.20 Abbe refractometer.

color from the field. Next, the index arm is adjusted using the vernier to coincide the dark and light portion of the field at the intersection of the cross hairs. The value of the index of refraction is read for sodium D lines.

43.6.2 Luminous Transmittance and Haze (ASTM D 1003)

Luminous transmittance is defined as the ratio of transmitted light to the incident light. The value is generally reported in percentage of light transmitted. Polymethyl methacrylate, for example, transmits 92% of the normal incident light. Haze is the cloudy appearance of an otherwise transparent specimen caused by light scattered from within the specimen or from its surface. Haze is defined as the percentage of transmitted light that in passing through a specimen deviates from the incident beam by forward scattering. Haze is normally caused by surface imperfections, density changes, or inclusions that produce light scattering. Haze is also reported in percentage.

Two procedures have been developed to measure light transmittance and light scattering properties: Procedure A requires the use of a hazemeter while Procedure B requires the use of a recording spectrophotometer.

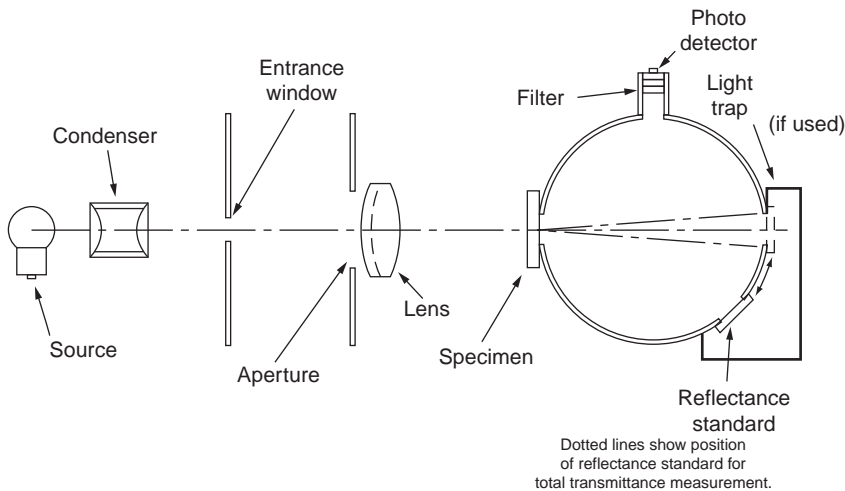


FIGURE 43.21 Schematic of hazemeter (reprinted with permission of ASTM).

43.6.2.1 Procedure A: Hazemeter This procedure employs an integrating sphere hazemeter as illustrated schematically in Figure 43.21. The test specimen must be large enough to cover the aperture, but small enough to be tangent to the sphere wall. A disc of 1.375 in. in diameter is most commonly used. A commercially available hazemeter is shown in Figure 43.22. The test is conducted by taking four different consecutive readings and measuring the photocell output as follows:

T_1 = specimen and light trap out of position, reflectance standard in position.

T_2 = specimen and reflectance standard in position, light trap out of position.



FIGURE 43.22 Hazemeter (Courtesy BYK-Gardner USA).

T_3 = light trap in position, specimen and reflectance standard out of position.

T_4 = specimen and light trap in position, reflectance standard out of position.

The quantities represented in each reading are incident light, total light transmitted by specimen, light scattered by instrument, and light scattered by instrument and specimen, respectively.

43.6.2.2 Procedure B: Recording Spectrophotometer This procedure is somewhat similar to Procedure A. The recording spectrophotometer is used to generate four different curves. From the recorded curves T_1 , T_2 , T_3 , and T_4 values are computed using automatic integrator. Calculations are carried out in a similar manner to determine total and diffuse luminous transmittance and the percentage haze.

43.6.3 Specular Gloss (ASTM D 523)

Specular gloss is defined as the relative luminous reflectance factor of a specimen at the specular direction. This method has been developed to correlate the visual observations of surface shininess made at roughly corresponding angles. The light beam is directed towards the specimen at a specified angle and the light reflected by the specimen is collected and measured. All specular gloss values are based on a primary reference standard—a highly polished black glass with an assigned specular gloss value of 100.

The operation of a glossimeter is very simple. The instrument is turned on and placed on black glass primary standard. The control knob is adjusted so that the meter indicates the value assigned to the primary standard. Next, the sensor is placed on the specimen surface and the gloss value is read directly from the analog or digital display. The linearity of the instrument is routinely checked by placing the sensor on a white secondary standard, which should read within 1.0 gloss unit of the assigned value of that standard. Figure 43.23 illustrates a commercially available glossmeter.



FIGURE 43.23 Glossmeter (Courtesy BYK-Gardner USA).

FURTHER READINGS

- Abolins V. Gardner impact versus Izod, which is better for plastics?. *Materials Engineering* (Nov. 1973).
- ASTM D 1709 (Part 36); ASTM D 3029 (Part 35), Annual Book of ASTM Standards, Philadelphia, PA., 1978.
- Baer E. *Engineering Design for Plastics*, New York: Reinhold; 1964. Chap. 4.
- Bergen RL. Tests for selecting plastics. *Metal Progress* (Nov. 1966).
- Billmeyer FW. *Textbook of Polymer Science*. New York: Interscience; 1962.
- Bragaw CG. Tensile Impact. *Modern Plastics* 1956;33 (10).
- Deanin RD. *Polymer Structure, Properties and Applications*. Boston, MA: Cahners Publishing Co.; 1972.
- Delatycki O. Mechanical performance and design in polymers. *Applied Polymer Symposia* 1971; 17 (134); Interscience, New York,
- “Design Guide,” *Modern Plastics Encyclopedia*. New York: McGraw-Hill; 1979–1980.
- “Design Guide,” *Modern Plastics Encyclopedia*. New York: McGraw-Hill; 1980–1981.
- Dregger DR. How dependable are accelerated weathering tests for plastics and finishes? *Machine Design* (Nov. 1973).
- Dubois JH, Levy S. *Plastics Product Design Engineering Handbook*. New York: Reinhold; 1977.
- Goldman TD, Lutz JT. Developing low temperature impact resistance pvc: a new testing approach. *ANTEC* 1979; 25.
- Harper CA. Short course in electrical properties. *Plastics World* (April 1979), p. 73.
- Heap RD, Norman RH. *Flexural Testing of Plastics*, London, England: The Plastics Institute; 1969.
- Ives GC, Mead JA, Riley MM. *Handbook of Plastics Test Methods*. London, England: Iliffe Books; 1971.
- Kamal MR. Weatherability of Plastic Materials, *Applied Polymer Symposium*, No. 4, New York: Interscience Publishers; 1967.
- Kinmonth RA, Saxon R, King RM. Sources of variability in laboratory weathering. *Polymer Engineering Science* 1970;10(5):309–313.
- Kinney GF. *Engineering Properties and Applications*. New York: John Wiley & Sons; 1957.
- Levy S, Dubois H. *Plastics Product Design Engineering Handbook*. New York: Reinhold; 1977.
- LNP Corporation, *Technical Bulletin, Predict Shrinkage and Warpage of Reinforced and Filled Thermoplastics*. Malvern, PA. 1978.
- Lubin G. *Handbook of Fiberglass and Advanced Plastics Composites*. New York: Reinhold; 1969.
- Mascia L. *The Role of Additives in Plastics*. London, England: Edward Arnold Publishing Company; 1974.
- Maxwell B, Harrington JP. Effect of velocity on tensile impact properties of PMMA. *Transactions on ASME* 1952; 74.
- McMichael S, Fischer S. *Understanding Materials with Instrumented Impact*. Materials Engineering, April 1989.
- Milby R. *Plastics Technology*, New York: McGraw-Hill; 1973.
- Miller RW. Considerations in the evaluation of plastics in electrical equipment. *Plastics Design and Processing* (July 1980).
- Nielsen LE. *Mechanical Properties of Polymers*. New York: Reinhold; 1962.
- O'Toole, JL. *Creep Properties of Plastics*. *Modern Plastics Encyclopedia*. New York: McGraw-Hill 1968.

- Reymers H. A new temperature index, who needs it, what does it tell. *Modern Plastics* (March 1970).
- Rodriguez F. *Principles of Polymer Systems*. New York: McGraw-Hill; 1970, Chap. 8.
- Shah Vishu *Handbook of Plastics Testing & Failure Analysis*. 3rd ed. New York: John Wiley & Sons; 2007.
- Spath W. *Impact Testing of Materials*. New York: Grodon and Breach Science Publishers; 1961.
- Starita JM. Impact testing. *Plastics World* (April 1977).
- Symposium, on Plastics Testing—Present and Future, ASTM Publication No. 132, Philadelphia, PA: American Society for Testing and Materials; 1953.
- Tanzillo JD. Development of an impact test for evaluation of toughness of rigid plastic building components. *ANTEC* 1969; 15.
- Tardif HP, Marquis H. Impact testing with an instrumented machine. *Metal Progress* (Feb. 1964).
- “The Perils of Izod—Part 1,” *Plastics Design Forum* (May/June 1980).
- Tryson GR, Takemori MT, Yee AF. Puncture testing of plastics: effects of test geometry. *ANTEC* 1979; 25.
- Underwriters Laboratories Publication, Polymeric Materials—Long Term Property Evaluations. UL 746B.
- Underwriters Laboratories Standards for Safety, UL 746 A-B-C-D, U. L. Inc. Melville. L.I. New York.
- Vincent PI. *Impact Tests and Service Performance of Thermoplastics*. London, England: The Plastics Institute; 1971.
- Westover RF. The thirty years of plastics impact testing. *Plastics Technology* 1958; 4.
- Westover RF, Warner WC. Tensile impact test for plastic materials. *Research and Standards* 1961;1 (11).
- Winolow RM, Matreyek W, Trozzolo AM. Polymers under weather. *Journal of Petroleum Science and Engineering* 1972;28(7).

TESTING AND INSTRUMENTAL ANALYSIS FOR PLASTICS PROCESSING: KEY CHARACTERIZATION TECHNIQUES

MARIA DEL PILAR NORIEGA

- 44.1 FTIR spectroscopy
- 44.2 Chromatography (GC, GC-MSD, GC-FID, and HPLC)
 - 44.2.1 GC
 - 44.2.2 Liquid chromatography, HPLC
 - 44.2.3 SEC or gel permeation chromatography (GPC)
- 44.3 DSC and thermogravimetry (TGA)
 - 44.3.1 DSC
 - 44.3.2 Modulated DSC, MDSC
 - 44.3.3 Oxygen induction time (OIT)
 - 44.3.4 TGA
 - 44.3.5 Modulated TGA (MTGA)
- 44.4 Rheometry
 - 44.4.1 Torque rheometry
 - 44.4.2 Capillary rheometry
 - 44.4.3 Rotational rheometry

References

44.1 FTIR SPECTROSCOPY

The infrared electromagnetic radiation produces vibrational and rotational changes in the molecule, distinctive of the chemical structure of the analyzed substance that can be used for the characterization of any molecule including polymers and additives. With the

introduction of the Fourier transform infrared spectrophotometer (FTIR) at the end of 1960s, some of the disadvantages of traditional dispersive instruments were solved and a faster and more sensitive technique was developed (Fuller et al.).

The FTIR spectroscopy could be a very useful technique in characterization of polymeric materials, since the following type of analysis could be done:

- identification and quantification of additives and polymers,
- determination of the polymer structure of multilayer films. This analysis has to be done in conjunction with other techniques such as optical microscopy and differential scanning calorimetry (DSC)
- analysis of contaminants in polymeric samples and additives,
- study of thermo-oxidative and environmental degradation of polymers,
- analysis of isomerism and functionalization of polymeric materials,
- analysis of chain branching and end functional groups in polymers,
- analysis of crystalline structure of polymeric materials.

Nowadays, a wide range of FTIR techniques are available, and the most useful ones for polymer characterization are as follows:

- *Transmission*: The sample is placed in the infrared path emitted by the source, and the transmitted infrared radiation is analyzed by the detector. The sample could be a film, a sheet, and a powder or a liquid supported in a transparent window.
- *Attenuated Total Reflectance ATR*: In this technique, the infrared radiation emitted by the source is internally reflected in a high-refractive-index crystal and only the evanescent waves interact with the sample in direct contact with the crystal. The ATR technique allows to obtain a low-penetration infrared spectrum of the sample, so it is used in the analysis of coatings and multilayer films, when only the surface has to be measured.

The chemical structure of polymers and additives could be correlated with the infrared absorption, with the use of available literature and Figure 44.1 (Conley, 1972; ASTM D 5576; Lomonte, 1962; Gartner et al., 1998; Sierra et al., 2000; Naranjo et al., 2008).

44.2 CHROMATOGRAPHY (GC, GC-MSD, GC-FID, AND HPLC)

The chromatography is a separation technique in which the components of the mixture are distributed in two phases (stationary phase and mobile phase) based on the partition coefficients. Depending on the type of stationary phase and mobile phase, the chromatography techniques can be grouped into the following categories:

- *Gas Chromatography (GC)*: The mobile phase is a gas, and the stationary phase is a solid or a liquid. *Example*: capillary GC.
- *Liquid Chromatography*: The mobile phase is a liquid, and the stationary phase is a solid or a gel. For example: high-performance liquid chromatography (HPLC) and thin-layer chromatography (TLC).

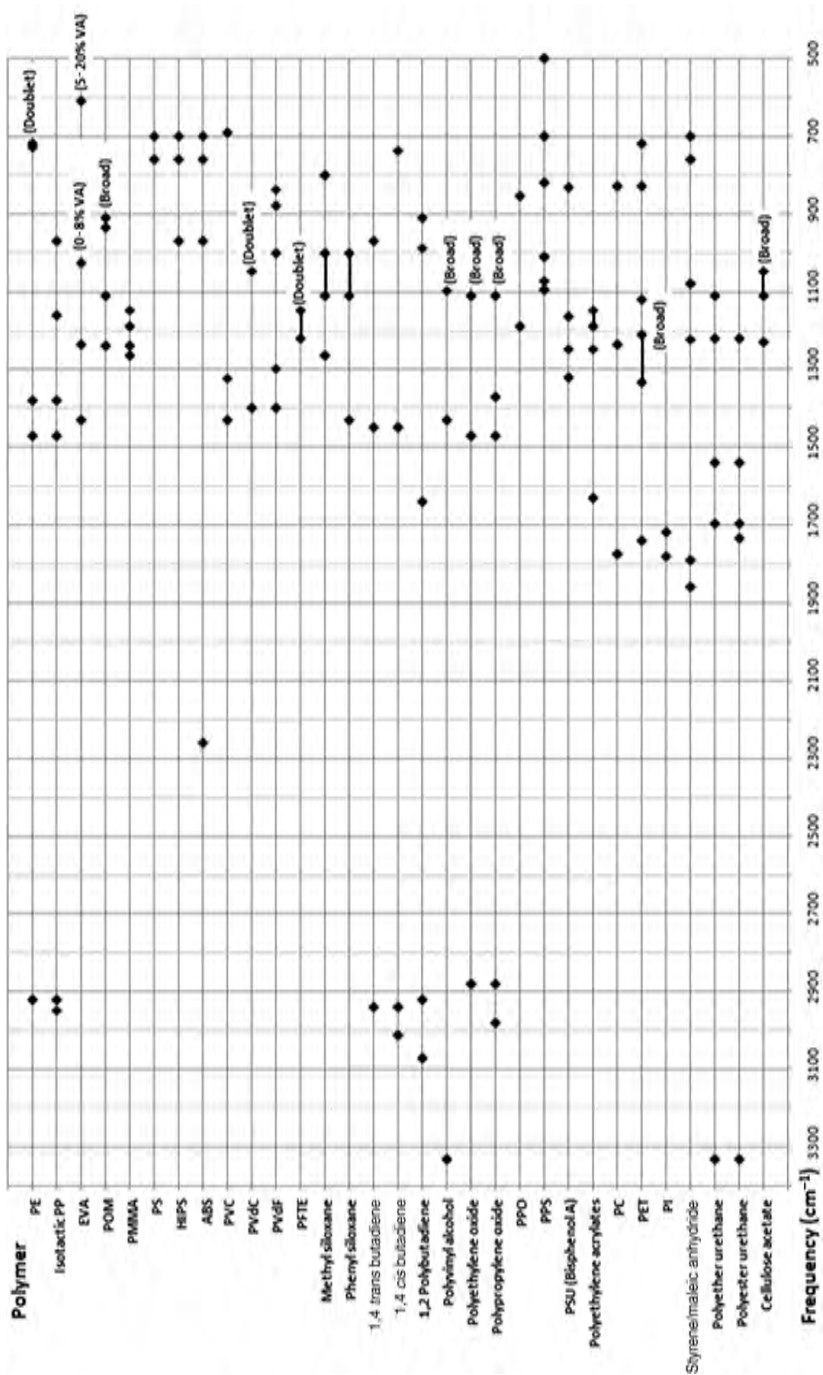


FIGURE 44.1 Characteristic infrared absorption bands of polymers (Conley, 1972; Gartner et al., 1998; Sierra et al., 2000).

Once the components are separated in the chromatographic columns, it is possible to quantify (and sometimes identify) the substances of the mixture with the help of a proper detector.

44.2.1 GC

The GC could be a very useful technique in characterization of polymeric materials, since the following types of analyses can be done:

- qualitative and quantitative analyses of comonomers in polymeric materials (after pyrolysis, Stilianos Rowis and Fedora, 1996; Frontier Laboratories Ltd; Wampler, 2005),
- residual monomer analysis in polymeric resins,
- qualitative and quantitative analyses of impurities and additives in polymeric materials,
- odor and volatiles analysis in polymeric materials,
- residual pesticides in polymeric materials,
- purity analysis of additives,
- specific migration studies in polymeric materials.

The GC technique could be used for the analysis of any substance that meets the following requirements: volatilization temperature below 320 °C and a molecular weight below 600 g/mole. This chromatographic technique includes the following components and instrumentation:

- *Injectors*: The injector is the component of gas chromatograph used to receive, volatilize, mix with the carrier gas, and transport the sample to the capillary column. The most universally used in the polymer and additives analysis are the split/splitless and the on-column.
 - *Split/Splitless Injector*: In the split operation, a fraction of the vaporized sample is carried into a column (according to a predefined split ratio), whereas in the splitless operation, the majority of the sample is carried into the column. For this reason, split injection is preferred for concentrated samples and splitless for trace analysis.
 - *On-Column*: This type of injector allows the injection of sample inside the column, for the analysis of substances with a molecular weight up to 1000 g/mole. The main advantages of this injector are less discrimination of low-volatility compounds, less adsorption of active compounds, less degradation of thermal sensitive compounds, and better sensitivity for trace compounds.
- *Columns*: The column is the component of gas chromatograph that supports the stationary phase and performs the separation of the sample components. For the majority of polymers and additives analysis, capillary columns are used with a typical internal diameter within the range of 0.32–0.20 mm. The polarity of columns depends on the type of substances to be separated. As a general rule, columns with low polarity are selected to separate compounds according to the boiling point, columns with high polarity are used to separate compounds by dipole–dipole interactions, and columns with intermediate polarity are selected to separate by means of both mechanisms.

- **Detectors:** The detector is the component of gas chromatograph that measures the signal of each component that is separated and eluted from the column. The most used detectors in the polymer and additives analysis are the flame ionization detector (FID) and the mass selective detector (MSD).
 - **Flame Ionization Detector (FID):** In this type of detector, the components that are separated from the sample are ionized in a hydrogen–air flame. Because of the high response to the majority of organic compounds, FID is considered as a universal detector.
 - **Mass Selective Detector (MSD):** In this type of detector, the compounds that eluted from the column are identified and quantified according to the fragmentation spectrum, once they are impinged with high-energy electrons and separated based on the mass-to-charge ratio.
- **Control and Data Analysis Unit:** It is the unit of the chromatograph that receives the signals from the detectors, provides control signal, and controls the other components including pressure and flow of carrier gas, oven temperature, injectors, and detectors.

In case of odor and volatile analysis, a very powerful technique for sampling is the headspace. In the headspace, the volatile compounds in equilibrium with the sample (solid or liquid) are removed from an enclosed space and injected into the gas chromatograph. To concentrate the volatiles in the enclosed space, the sample vial is shaken and heated up to 120 °C. This sampling technique is particularly useful for the analysis of residual monomers, residual solvents in printed plastics, and undesired odors from plastics and contamination. For quantitative analysis or when the partition coefficient is unknown, multiple extraction of the headspace is injected into the chromatograph. As expected, the area of the chromatographic peak decreases with the number of extractions in an exponential trend (see Figure 44.2a). The total area can be calculated from an exponential approximation of the infinite series. The k parameter can be calculated from the slope of the linear correlation of $\ln(A_i)$ against the number of extractions minus one (see Figure 44.2b).

As we mentioned before in the MSD detector, the compounds separated and eluted from the column are fragmented in characteristic mass spectrum. Smaller fragments are produced by high-energy electrons. This fragmentation process (under controlled conditions) produces a mass spectrum characteristic of the chemical structure of the compound that can be used not only for quantification purpose, but also for identification purpose. Table 44.1 presents the representative mass fragments of a selected group of additives used in polymers.

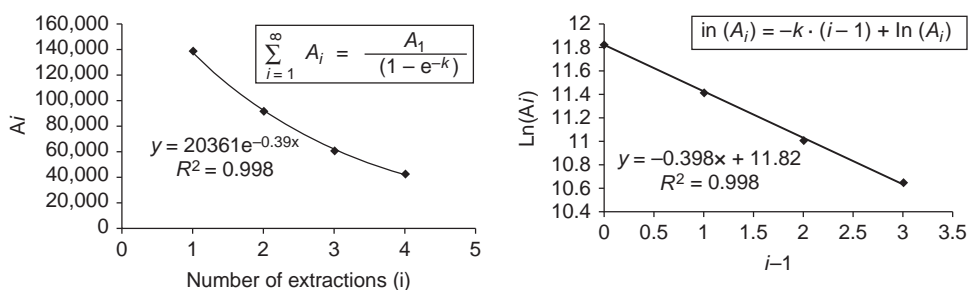


FIGURE 44.2 Calculation of the total area in a multiple headspace extraction.

TABLE 44.1 Representative Mass Fragments of Additives Used in Polymers (McLafferty and Turecek, 1993; National Institute for Standard Technology, 2005)

Type	Additive	MW (g/mol)	Mass Fragments (<i>m/z</i>)
Flame retardant plasticizers	Triphenyl phosphate (TPP)	326	326, 325, 77, 65, 39, 51, 169, 170, 233, 96
	o-Tricresyl phosphate (TOCP)	368	165, 91, 179, 181, 368, 277
Fast gelation plasticizers	Tricresyl phosphate (TCP)	368	368, 367, 91, 165, 179, 369, 198
	Dimethyl phthalate (DMP)	222	163, 77, 164, 135, 92, 50, 149
General-purpose plasticizers	Dibutyl phthalate (DBP)	278	149, 106, 104, 80, 61, 223, 205
	Bis(2-ethylhexy) phthalate (BEHP)	390	149, 150, 41, 76, 104, 223, 167, 279
Low-volatility plasticizers	Di-isooctyl phthalate (DIOP)	390	149, 167, 57, 70 41, 71
	Diisodecyl phthalate (DIDP)	446	149, 57, 167, 307
Adipates and sebacates plasticizers	Diundecyl phthalate (DUP)	474	149, 43, 57, 41, 55, 71
	Bis(2-ethylhexy) sebacate (DEHS)	426	185, 57, 112, 71, 70, 43
	Bis(2-ethylhexy) adipate, (DEHA)	370	129, 57, 71, 70, 112, 147, 43
	Diisooctyl adipate (DIOA)	370	129, 57, 55, 41, 43, 70
Slip additives	Butyl bencyl adipate	292	91, 129, 111, 101
	Oleamide	281	59, 72, 55, 41, 43, 281
	Erucylamide	337	59, 72, 55, 41, 43, 69, 137, 337
	Stearamide	283	59, 72, 28, 43, 57, 55, 41, 283
Antioxidants and UV stabilizers	Benzophenone	182	105, 77, 182, 51, 50
	Benzotriazole	119	119, 64, 91, 63, 38, 52
	BHT	220	205, 57, 220, 206, 145, 177

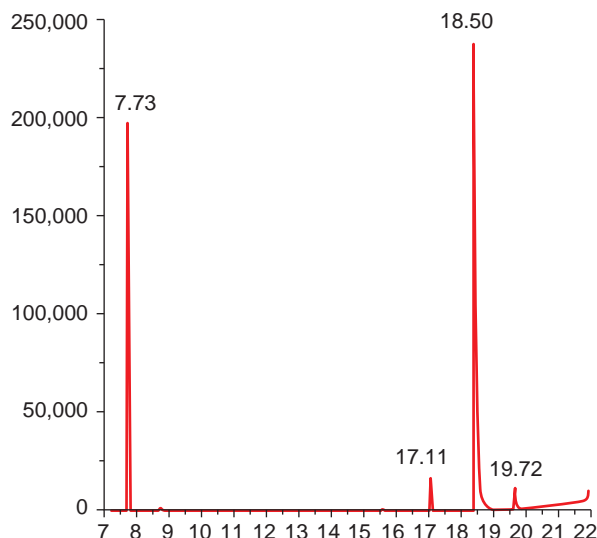


FIGURE 44.3 GC/MSD for plasticizers extracted from vinyl sample (measured at ICIPC¹).

EXAMPLE 44.1: Identification of plasticizers using GC/MSD

The plasticizers from a vinyl sample were extracted by Soxhlet technique with ethylic ether in 16 h. Then, the solvent was evaporated and injected directly into the GC/MSD equipment. The GC/MSD reports (Figure 44.3) several peaks related to plasticizers of the vinyl sample. In Table 44.2, the retention time and the most intense mass fragments are reported.

The eluted compound at 7.73 min presents mass fragment characteristics of di-methyl phthalate (DMP) (according to the Table 44.1, the characteristic mass fragments are 163, 77, 164, 92, 50, 135, and 149). The eluted compound at 18.53 min presents mass fragment characteristics of bis(2-ethyl hexyl) phthalate or DOP (according to Table 44.1, the characteristic mass fragments are 149, 167, 57, 279, 71, 113, 83, 150, 41, 76, 104, and 223) or di-isooctyl phthalate DIOP (according to Table 44.1, the characteristic mass fragments are 149, 167, 57, 70, 41, 83, 104, 113, 279, and 71). The eluted compound at 17.11 min presents mass fragment characteristics of triphenyl phosphate (TPP; according to Table 44.1, the characteristic mass fragments are 326, 325, 77, 65, 39, 51, 169, and 170). The eluted compound at 19.72 min presents mass fragment characteristics of tricresyl phosphate (TCP; according to Table 44.1, the characteristic mass fragments are 367, 77, 115, 65, 91, 382, 41, 51, 178, 152, and 215).

44.2.2 Liquid Chromatography, HPLC

The liquid chromatography can be a very useful technique in the characterization of polymeric materials, since the following types of analyses can be done:

¹ ICIPC: Research Institute for Plastics and Rubber, Colombia

TABLE 44.2 Mass Spectra of Plasticizers Extracted from Vinyl Sample

Retention Time (min)	Mass Fragment (m/z)
7.73	163, 77, 164, 75, 92, 133, 50, 135, 104, 105, 120, 64, 149
17.11	326, 325, 77, 94, 65, 215, 169, 170, 233, 232, 168, 51, 228, 141
18.50	149, 167, 57, 71, 70, 150, 279, 55, 113, 83, 104, 112, 94, 84, 132
19.72	367, 368, 77, 207, 94, 382, 65, 115, 281, 178

- determination of the molecular weight distribution of polymeric resins,
- residual monomer analyses in polymeric resins,
- qualitative and quantitative analyses of impurities and additives in polymeric materials,
- purity analysis of additives,
- specific migration studies in polymeric materials.

This chromatographic technique includes the following components and instrumentation (Patnaik, 2004; Kazakevich and Lobrutto, 2007; Scott, 2003):

- *Mobile Phase Supply System*: This component of HPLC consists of glass reservoirs where the mobile phase is stored and degassed by a helium gas flow. The solvent is also filtered to eliminate solid contaminants in a 0.45- μm membrane filter.
- *Solvent Delivery System*: This consists of a pump, pulse damper, and a programmer. The pumps force the mobile phase through the system.
- *Sample Valve*: It is a loop that is able to inject a sample volume into the mobile phase just at the head of the column.
- *Column and Thermostat*: In the case of size-exclusion chromatography (SEC), the columns are packed with gels of different porosities where the separations are done according to the hydrodynamic size of macromolecules. Although most of the analysis are carried out at room temperature, in the gel permeation chromatography of very insoluble polymers (such as polyolefins), the column is heated in an oven or in a thermostat.
- *Detector*: It is the device to measure the solutes in the mobile phase and record the resulting chromatogram. The typical detectors in HPLC are the following: ultra-violet–visible detectors, the fluorescence detector, the electrochemical detectors, and the refractive index detector. For SEC, laser light-scattering detectors and viscometer detectors are used as their response is proportional to molecular weight and concentration (Table 44.3).
- *Data Acquisition and Processing Unit*

44.2.3 SEC or Gel Permeation Chromatography (GPC)

The SEC is a chromatographic technique that separates polymer molecules based on differences in their molecular size. The smaller-sized macromolecules have the ability to penetrate inside the pore space and the movement through the column is retarded. The macromolecules in a solvent solution adopt a coiled form characterized by their hydrodynamic radius, which is proportional to the molecular weight (see Equation (44.1))

TABLE 44.3 Examples of Polymer Additives Analysis by Using HPLC

Method	Sample Preparation	Mobile Phase	Column	Detector	Source
Phthalates plasticizers in PVC	Extraction with 1% isopropal nol in hexane	Iso-octane	Zipax 0.5% oxydipropionitrile, 2.1 mm × 1 m, flow rate 0.5 ml/min, injection vol. 10.6 µL	Diode array detector	Crompton (2007)
Antioxidants and slip additives	Polymer were dissolved in tetrahydrofuran and filtered after extraction with ultrasonic bath for 30 min	A: Water + 0.001 m tetrabutylammoniumhydrogensulfate, pH = 3.0 with H ₂ SO ₄ , B = acetonitrile, gradient: start with 30% B, to 98% B in 10 min	Column 125 × 3 mm BDS, 3 µm, flow rate 0.5 mL/min, injection vol. 5 µL oven temp. 40°C	UV-detector DAD, 280/20 nm, reference 900/50 nm	Grazzfeld-Huesgen (1997)
Antioxidants		A: water, B: methanol	Column: microsorb-MV, C8 3.5 µm, 4.6 × 50 mm, vol. 1 µL, injection vol. 5 µL, flow rate 0.4 ml/min	Ion trap mass spectrometer	Rudrabhatla
		LC program:			
		Time	A (%)	B (%)	
		00 : 00	40	60	
		10 : 00	0	100	
		15 : 00	0	100	
		15 : 01	40	60	
		20 : 00	40	60	

(continued)

TABLE 44.3 (Continued)

Method	Sample Preparation	Mobile Phase				Column	Detector	Source
Antistatic additives	Extraction using microwave energy	A: Isopropanol, B: Isopropanol/water (60/40)		LC program:		SunFire TMC18 (3.0 × 150 mm; 3.5 µm), injection vol. 20 µL, oven temp. 45°C	Photodiode array (220 nm)	Carballeira-Amarelo et al.
		Time	A (%)	B (%)	Curve			
		0	0	100	Linear			
		2	100	0	Linear			
		10	100	0	Linear			
		13	0	100	Linear			

Kazakevich and Lobrutto, 2007).

$$R = \left(\frac{3}{4} \pi M [\eta] \right)^{1/8}, \quad (44.1)$$

where R is the hydrodynamic radius, M is the molecular weight, and $[\eta]$ is the intrinsic viscosity (the limit of the ratio of the specific viscosity of the polymer solution to its concentration, when concentration tends to zero).

To obtain the molecular weight distribution (MWD) of a particular polymeric sample, a calibration of the elution time with its molecular weight is required. For that purpose, polystyrene PS standards with a narrow MWD and known average molecular weight are commonly used. Because macromolecules of PS standards are of different nature of the polymer sample, Benoit et al. (1966) Grubisic et al. (1967) proposed a universal calibration for elution time with the PS standard's molecular weight (see Equation (44.2)).

$$M_{\text{sample}} = \left(\frac{K_{\text{PS}}}{K_{\text{sample}}} \right)^{\frac{1}{1+\alpha_{\text{sample}}}} M_{\text{PS}}^{\frac{1+\alpha_{\text{PS}}}{1+\alpha_{\text{sample}}}}, \quad (44.2)$$

where M_{sample} is the molecular weight of sample, M_{PS} is the molecular weight obtained by calibration with PS standards, K_{sample} and α_{sample} are the Mark–Houwink constants for sample with same solvent and temperature, and K_{PS} and α_{PS} are the Mark–Houwink constants for polystyrene with same solvent and temperature.

EXAMPLE 44.2: MWD for a polylactic acid (PLA)

During the analysis of MWD of a PLA polymer by GPC, the following information was obtained. Table 44.4 and Figure 44.4 registered the calibration data with PS standards.

To perform the universal calibration of MW for PS standard and the MW of PLA, the Mark–Houwink constants obtained from literature were used (see Table 44.5). The

TABLE 44.4 Calibration for Polystyrene Standards

Elution Time (min)	MW for PS	MW Corrected to PLA (Equation (2.3))
16.20	19,825,000	17,883,161
19.65	5,000,000	4,224,008
21.41	1,112,000	874,547
22.69	300,000	221,675
24.77	50,000	33,925
28.22	5,000	3,040
30.67	1700	982
33.20	666	368

TABLE 44.5 Mark–Houwink Constants for PS and PLA in THF at 35 °C (Kubies et al., 2006)

Mark-Houwink Constants	PS	PLA
K	1.25e-4 dL/g	5.49e-4 dL/g
α	0.717	0.639

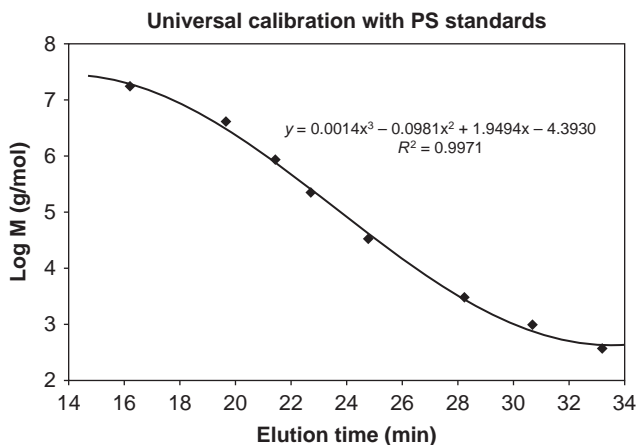


FIGURE 44.4 Calibration graph for polystyrene standards.

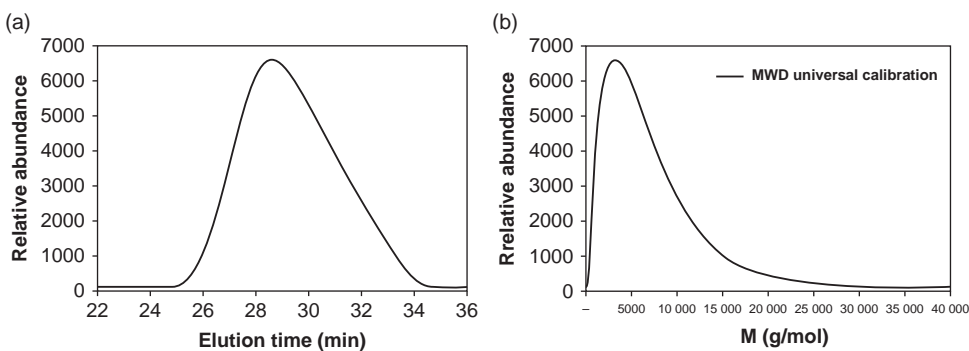


FIGURE 44.5 MWD for a PLA using GPC data and the universal calibration.

universal calibration calculated with the Mark–Houwink constants is presented in Equation (44.3)

$$M_{\text{PLA}} = 0.4054 \cdot M_{\text{PS}}^{1.0476}. \quad (44.3)$$

The data obtained from GC analysis using THF solvent at 35 °C could be converted into the MWD by using the polynomial of polystyrene standards (Figure 44.4) after applying the universal calibration (Equation (44.3)). The raw data of GPC is presented in Figure 44.5a, and the MWD obtained is presented in Figure 44.5b.

44.3 DSC AND THERMOGRAVIMETRY (TGA)

44.3.1 DSC

The DSC is a thermal technique that measures the enthalpy changes, coupled with diverse physical and chemical events, experienced by a sample under a certain temperature–time program. By using the DSC technique, it is possible to evaluate heat absorbed or emitted by a sample during several thermal events, including glass transition, melting, crystallization, crosslinking, chemical reaction, evaporation, and chemical decomposition. The thermal

TABLE 44.6 Thermal Properties and Events of a Polymer Measured by DSC and TGA (Naranjo et al., 2008; Ehrenstein et al., 2004)

Property/Event	DSC	TGA
Heat capacity, C_p	X	
Melting feat, ΔH	X	
Melting temperature, T_m	X	
Glass transition temperature, T_g	X	
Crystallization temperature, T_k	X	
Crystallinity	X	
Degree of crystallinity, χ	X	
Thermal and mechanical history	X	
Compatibility and miscibility	X	
Chemical stability	X	X
Thermal stability		X
Chemical composition		X
Moisture content		X
Filler content		X
Drying time		X
Crosslinking rate	X	X
Crosslinking degree	X	

properties and possible thermal events of a polymer measured by DSC and TGA are presented in Table 44.6. In addition, the peak temperature, start and end temperatures, and onset temperature of the event can be determined. The heat capacity of a sample as a function of temperature can also be determined, if a proper standard is used (i.e., a sapphire standard). The rate of change of heat capacity or enthalpy as a function of time (or temperature) can be determined, and therefore the DSC is an important tool for kinetic studies of chemical reactions, crosslinking, and thermo-oxidative decomposition.

The DSC permits to determine thermal transitions of polymers in a range of temperatures between -180 and $+600$ °C. The DSC test requires samples that are in the milligram range (<20 mg).

44.3.1.1 Glass Transition Temperature, T_g The glass transition temperature, usually denoted by T_g , is a common property of amorphous and semicrystalline polymers. It is the temperature at which the relaxation of the macromolecules stops when the polymer is cooled. The mobility of the chain segments is much below T_g (it is known as “frozen material”). As a consequence, as the polymer reaches the glass transition temperature, a very rigid and fragile nature is observed. T_g is registered in the DSC technique as a step in the heat capacity or in the heat flow. By convention, T_g is calculated as the temperature at which half of the change in specific heat capacity has occurred.

44.3.1.2 Melting Temperature, T_m As a semicrystalline polymer is heated above T_g , at a certain temperature, the crystalline domains are disaggregated and a viscoelastic fluid is obtained. This thermodynamic transition is usually denominated melting temperature, T_m . T_m is observed in the DSC technique as an endothermic peak in the heat capacity or in the heat flow, so a peak temperature (as well as start and end temperatures) and a peak area (enthalpy change during melting) can be determined. The differential scanning calorimeter is used to measure the melting, T_m , and the glass transition temperatures of polymers using the ISO 11357 and ASTM 3418 standards.

44.3.1.3 Crystallization Temperature, T_k As a semicrystalline polymer is cooled, at a certain temperature, the crystalline domains are reordered and crystalline polymer is obtained. This thermodynamic transition is usually denoted as crystallization temperature, T_k . T_k is observed in the DSC technique as an exothermic peak in the heat capacity or in the heat flow, so a peak temperature (as well as start and end temperatures) and a peak area (enthalpy change during crystallization) can be determined. As a rule, the crystallization temperature is always lower than the melting temperature and depends on the cooling rate.

Only under exceptional conditions, a 100% crystalline polymer is obtained, so a measurement of the degree of crystallinity is very useful. The degree of crystallinity, χ , is determined from the ratio of the heat of fusion of a polymer sample, ΔH_{SC} , and the enthalpy of fusion of a 100% crystalline sample ΔH_C .

$$\chi = \frac{\Delta H_{SC}}{\Delta H_C}. \quad (44.4)$$

The importance of the degree of crystallinity can be explained by the fact that it is possible to relate it with processing conditions (as cooling rate) and the physical properties of the final product.

Figure 44.6 shows a typical DSC curve measured under a heating program using a partly crystalline polymer, that is, PET sample. In Figure 44.6, the transition or detectable deviation from the baseline is T_g ($T_g \approx 83.2^\circ\text{C}$); the first area that is enclosed between the trend line and the baseline is a direct measurement for the amount of heat, ΔH_k , needed for transition. In this case, the transition is a cold crystallization ($T_k \approx 120.4^\circ\text{C}$) and the area corresponds to the heat of crystallization. Finally, the second area that is enclosed between the trend line and the baseline is a direct measurement for the amount of heat,

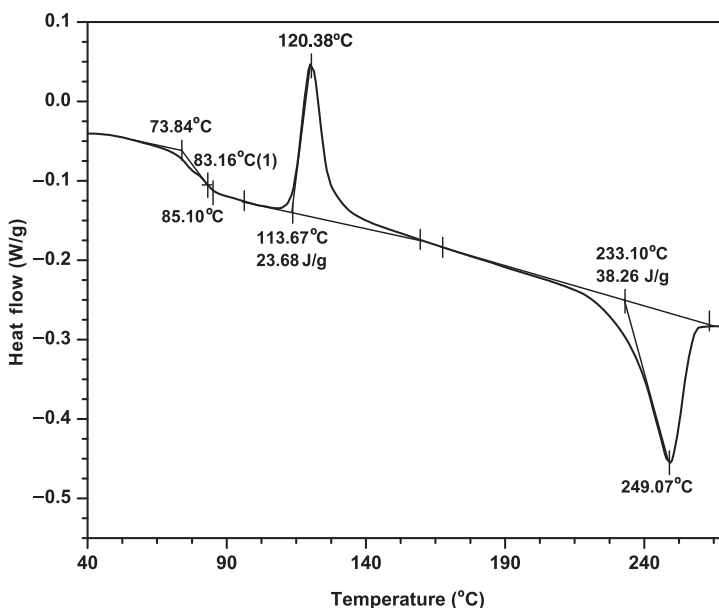


FIGURE 44.6 DSC heat flow of a PET sample (measured at ICIPC).

TABLE 44.7 DSC Characteristic Data for Selected Polymers (Knappe et al., 1993; Hemminger and Cammenga, 1988)

Polymer	T_g (°C)	T_m (°C) Peak	100% crystalline ΔH_c (J/g)	Comments
LDPE	−120 to −70	105–115	140	
LLDPE	−120 to −70	120–130		
HDPE	−130 to −80	130–135	290	
PP	−10 to 0	165–176	207	Isotactic homopolymer Random copolymer Block copolymer VA content from 35 to 3%
EVA	−40 to 20	65–110		
PVC	81–99			
PS	80–113			Atactic homopolymer
HIPS	80–113 −80 to −20			Styrene blocks Butadiene bocks
SAN	105–130			
ABS	−80 to −50 100–110 120–130			Butadiene bocks Styrene blocks Acrylonitrile blocks
PA6	40–85 5	180–230 180–230	190 190	Dry Equilibrium moisture
PA66	50–70 5	225–265 225–265	200 200	Dry Equilibrium moisture
PA11	40–50	180–200	224	
PA12	40–50	175–190	95	
PA610	50–65	210–233	209	
PET	60–70 80	250–285	115	Amorphous Semicrystalline
PLA (Sierra et al.,)	55–65 40–50	150	93	Amorphous Semicrystalline
PBT	25–75	190–250	142	
PC	140–150			
POM	−85 to −75 −30	175–190 140–170	325 250	Homopolymer Copolymer
PMMA	106–115			Atactic homopolymer
PTFE	−73 to −20	190–335	82–115	

ΔH_m , needed for another transition. In this case, the transition is melting ($T_m \approx 249^\circ\text{C}$) and the area corresponds to the heat of fusion.

Table 44.7 is an useful review of glass transition temperatures, melting temperatures, and 100% crystalline melting temperatures for selected thermoplastic polymers following the standards ASTM D3417 and ASTM D3418.

Specific heat, C_p , is one of the many material properties that can be measured with DSC. During a DSC temperature sweep, the sample pan and the reference pan are maintained at the same temperature. This allows the measurement of the differential energy required to maintain identical temperatures. The sample with the higher heat capacity will absorb a larger amount of heat, which is proportional to the difference between the heat capacity of the measuring sample and the reference sample. It is also possible to determine the purity of a polymer sample when additional peaks or curve shifts are detected in a DSC measurement.

44.3.2 Modulated DSC, MDSC

MDSC overcomes the limitations of the standard DSC offering higher sensitivity and resolution. MDSC can separate overlapping thermal events of a polymer, which is difficult or almost impossible to achieve by conventional DSC. MDSC improves upon standard DSC as it measures the total heat flow plus its heat capacity component and obtains the kinetic component from their difference (Thomas, 2006). The possibility to resolve complex transitions into specific components facilitates a much better data interpretation. It is also possible to distinguish thermally reversible repeatable transitions, such as glass transition and melting, from irreversible events, such as crosslinking, decomposition, and crystallization, among others (Ehrenstein et al., 2004).

Figure 44.7 shows MDSC results obtained in one single run under a slow heating program using a quenched PA6 specimen. In the lower left curve (reversible heat flow), the first transition from the baseline is $T_g \approx 43.7^\circ\text{C}$. The next area that is enclosed between the trend line and the baseline is a direct measurement of ΔH_m , needed for melting; in this case, the melting peak is $T_m \approx 220.7^\circ\text{C}$. The second area, upper left curve (nonreversible heat flow), that is enclosed between the trend line and the baseline is a direct

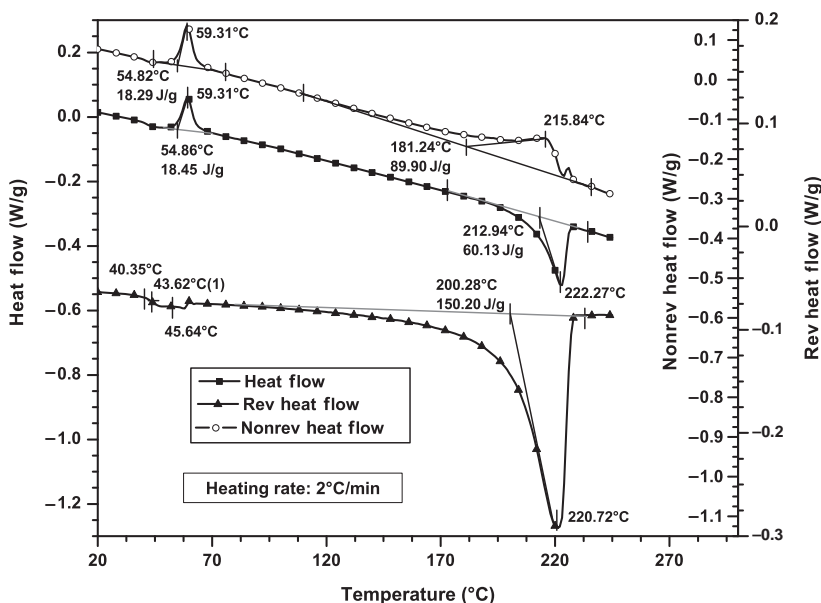


FIGURE 44.7 MDSC heat flow of a quenched PA6 (measured at ICIPC).

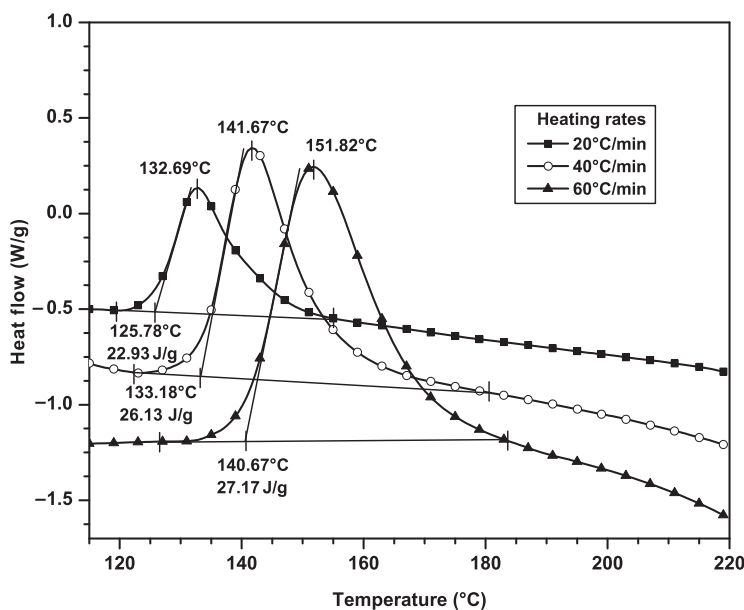


FIGURE 44.8 MDSC crystallization results for PET (measured at ICIPC).

measurement of the amount of heat, ΔH_k , needed for cold crystallization; in this case, $T_k \approx 59.3^\circ\text{C}$. Finally, the third area, upper right curve, is an irreversible and nonrepeatable effect that may be related to crystal rearrangement during melting.

Figure 44.8 presents a MDSC crystallization study for an amorphous PET sample at different heating rates. From the figure, it is possible to assess that at higher heating rates the crystallization temperature shifts to higher values in almost 12%. At higher heating rates, the heat of crystallization shows an increase in almost 15%, implying that a crystallization may take place during heating (Thomas, 2006).

44.3.3 Oxygen Induction Time (OIT)

Oxygen induction time is a technique or method that uses DSC to make comparisons of the stability of polymeric materials to thermal oxidation. Most plastics need to be stabilized against thermal oxidation such as stabilizers, antioxidants, and UV absorbers. There are two methods: the static method (less sensitive but faster) and the dynamic method (more sensitive but time-consuming) (Ehrenstein et al., 2004).

The ASTM D3895-04 standard is a procedure for the determination of OIT of polymers using DSC. This test is applicable to polyolefin materials that are in a stabilized form or in a compounded form. An OIT program requires a constant heating rate under an inert gas atmosphere (nitrogen) until 200°C . Afterward, the atmosphere is switched to oxygen; and this is the start of the recording for the experiment until oxidation is observed. Figure 44.9 shows the OIT measurement of a polypropylene (PP) masterbatch.

In this test, the switch to oxidative atmosphere takes place at 19.91 min, and the oxidation occurs at 20.89 min (read from Figure 44.9). Therefore, the OIT of this PP material is 0.98 min.

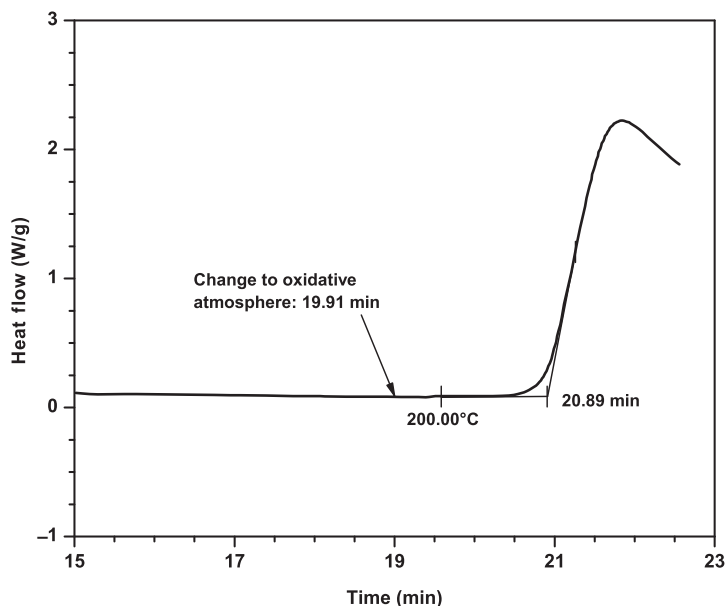


FIGURE 44.9 OIT measurement of a PP material (measured at ICIPC).

44.3.4 TGA

The thermogravimetric analysis (TGA) is a thermal analysis technique that measures the weight changes suffered by a sample under a certain temperature–time program, working on the principle of a beam balance. By using the TGA, it is possible to evaluate weight changes due to the following thermal events: volatilization of moisture, volatilization of additives, decomposition of polymers and additives, decomposition of organic pigments, and decomposition of some mineral fillers, such as calcium carbonate (CaCO_3). This measurement technique is typically used for thermal stability. The testing chamber can be heated (up to $\sim 1200^\circ\text{C}$) and rinsed with gases (inert or reactive).

44.3.4.1 Decomposition Temperatures of Thermoplastics Table 44.8 gives the typical peak decomposition temperatures obtained by TGA for several polymers. A thermogravimetric analyzer can detect weight changes of less than $10\ \mu\text{g}$ as a function of temperature and time (Table 44.8).

Figures 44.10 and 44.11 show TGA measurements of a PLA or polylactide at different heating rates: 1, 5, 10, 20, and $40^\circ\text{C}/\text{min}$. Based on this information, besides the peak decomposition temperature (T_d) at each heating rate, it is also possible to estimate the decomposition kinetics of this PLA sample based on weight changes. At higher heating rates, the T_d shifts to higher values.

44.3.5 Modulated TGA (MTGA)

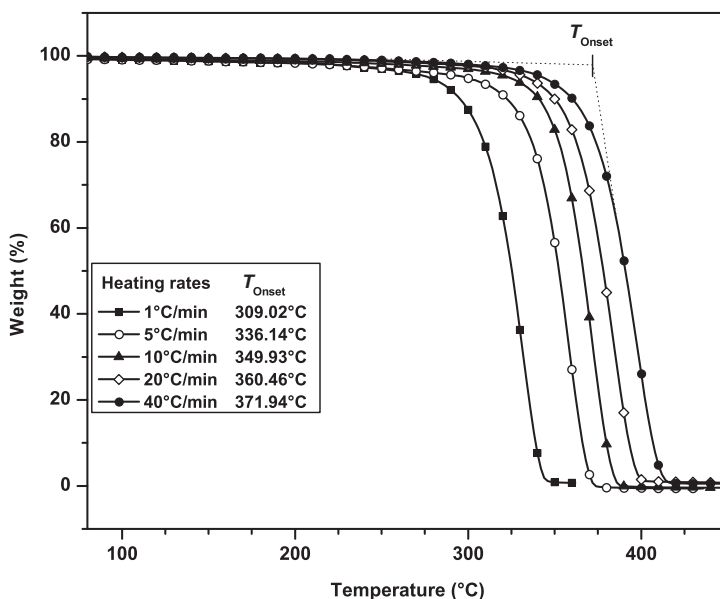
Modulated TGA (MTGA), compared to the standard TGA technique, offers advantages for polymer decomposition studies. This thermal analysis method generates model-free

TABLE 44.8 Peak Decomposition Temperatures Obtained by TGA for Several Polymers (Knappe et al., 1993)

Polymer	10 °C/min and N2 50 mL/min		20 °C/min and N2 50 mL/min		30 °C/min and N2 50 mL/min	
	Td1 °C	Td2 °C	Td1 °C	Td2 °C	Td1 °C	Td2 °C
PVC			333	466	280	460
LDPE				487		480
PP						480
PS				443		
ABS		420–425				
PA 66				430–473		
POM			315	370		

kinetic data from which activation energy can be calculated and studied as a function of several parameters, such as time, temperature, and conversion, among others. Another improvement is that the kinetic data for optimizing polymer processing can be produced in a single run (Blaine).

In general, problems related to thermal machine setup in polymer processing can be investigated and solved (troubleshooting) with the use of the different thermal analysis techniques (Noriega et al., 2001). TGA registers the different losses of weight of a sample as a function of temperature. This rules out degradation of polymers and polymer blends under temperature.

**FIGURE 44.10** TGA of PLA at different heating rates (measured at ICIPC).

44.4 RHEOMETRY

Rheometry deals with experimental methods and devices called rheometers, to measure the rheological behavior of materials (Macosko, 1994). Rheometry delivers essential information for the descriptive rheology that quantifies the relationship between two quantities: the stress and the rate of deformation of materials under the action of external forces. In our case, these materials are polymers. The mathematical expressions that establish the relationship between both quantities are called “constitutive equations.” This definition is not restricted to fluids; solids also undergo deformation under the action of external forces. The distinction between solids and fluids is a rheological issue and depends on the time scale of observation for the “rate of deformation.” If the time scale is big enough, it can be observed that a solid can flow under an external force; if the time scale of observation is small enough, a fluid can show a finite deformation like a solid under external forces. The Deborah number (Osswald and Menges, 2003) delimits this, and it is the ratio of a characteristic relaxation time of a material and the characteristic time scale of observation. Polymers are viscoelastic materials. It is not necessary for the polymers to stop deforming under stress in solid state; they show an elastic behavior under flow conditions.

Constitutive equations for flows are tensorial expressions. Viscosity is the material or fluid property that establishes the relationship between two tensors: stress and rate of deformation. Both the tensors have shear and elongation components that define two different kinds of viscosities. They are called shear viscosity and elongational or extensional viscosity.

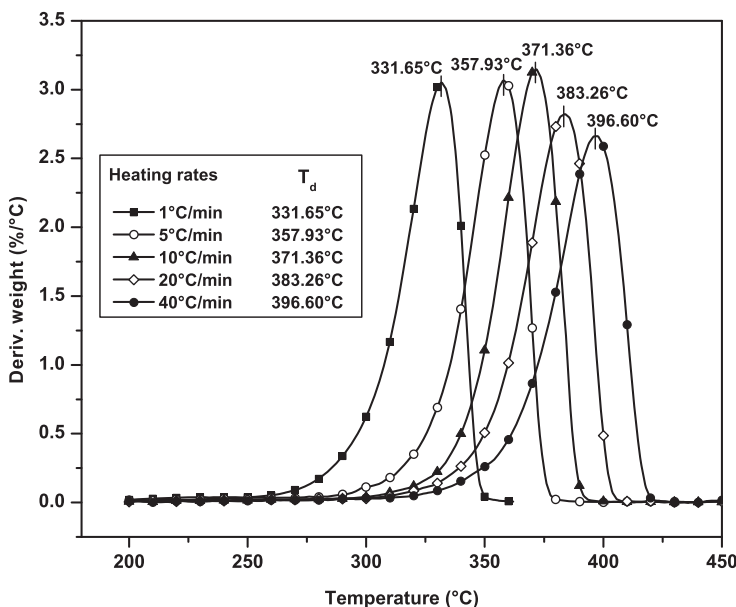


FIGURE 44.11 Decomposition temperatures of PLA at different heating rates (measured at ICIPC).

Rheometers are built to produce simple isothermal flows or simple deformation rate patterns. The tensors for description of stress and rate of deformation inside them are very simple. Therefore, viscosity can be expressed with very simple mathematical expressions. For the construction and use of rheometers there should be only one component for stress and deformation rate. The “shear viscosity” or the “elongational viscosity” can be calculated from them easily.

There are many possibilities to built rheometers, but nowadays, the most common and commercially available types are the capillary and the rotational rheometers. The flow in a capillary rheometer is a steady-state flow. These devices allow the study of shear viscosity as a function of shear rate in a wide range, usually from 10 s^{-1} to 10^4 s^{-1} , and depending on temperature under high pressure.

Rotational rheometers can produce a steady-state deformation rate, a steady-state shear stress, or both in an oscillatory and inclusive in a transient mode. These rheometers allow the study of viscosity as a function of shear rate, under the effect of temperature, frequency, and time of deformation. They normally work under atmospheric pressure.

44.4.1 Torque Rheometry

These devices are wrongly denoted as rheometers. The flow pattern inside them is so complex that it is not possible to measure stress and deformation rate components individually. Therefore, it is difficult to solve the tensorial equation for viscosity as well. However, they have found an important role in mixing studies, first-order decomposition kinetics, and thermal stability measurements, among others. For this reason, most people denote them as rheometers.

44.4.2 Capillary Rheometry

This rheometric technique deals with the experimental description of the material flow through a channel with a geometrical well-defined cross section (circular, rectangular, or annular), whose characteristic dimensions (diameter, width, and height) are small enough compared with the channel or capillary length. This condition is necessary to obtain a significant portion of flow pattern inside the channel, free of distortions caused by inlet and outlet phenomena in the capillary. In case of rectangular capillaries, the width of their cross section must be greater than 10 times the height to minimize “border effects” in the cross section of the flow pattern. For annular capillaries, the ratio of the gap between cylinders and the diameter of the internal cylinder should be greater than 20 to minimize the effect of curvature on the flow. Annular capillaries are bigger than the others. As annular capillaries need significantly more material during measurements, it is hard to keep material temperature constant, not common in commercial and industrial applications (Figure 44.12).

The constitutive tensor equation for a fully developed, isothermal, incompressible, and steady-state flow in a channel is reduced to the following simple expression in terms of the wall shear stress τ_{wap} and the wall shear rate $\dot{\gamma}_{\text{wap}}$, where η_{ap} is the shear viscosity

$$\tau_{\text{wap}} = \eta_{\text{ap}} \cdot \dot{\gamma}_{\text{wap}}. \quad (44.5)$$

The wall shear stress τ_{wap} can be calculated from (44.6) having the capillary radius R and the pressure drop Δp along a portion of the capillary, where the flow can be

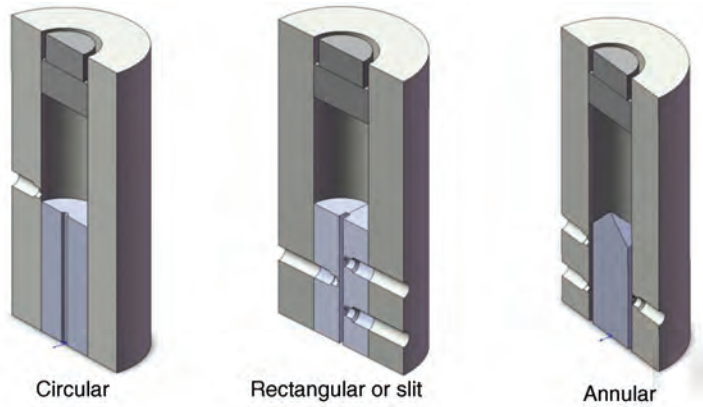


FIGURE 44.12 Types of capillary rheometers.

considered as fully developed.

$$\tau_{\text{wap}} = \frac{R}{2} \cdot \frac{\Delta p}{l}. \quad (44.6)$$

If the shear viscosity η_{ap} is independent of shear rate as for a Newtonian fluid, and of pressure, the wall shear rate $\dot{\gamma}_{\text{wap}}$ is obtained from the Hagen–Poiseuille equation for the flow in a cylindrical channel and the volumetric flow Q in Equation (44.7)

$$\dot{\gamma}_{\text{wap}} = \frac{4Q}{\pi R^3}. \quad (44.7)$$

The subindex “ap” stands for “apparent” because the above equations are valid for Newtonian fluids. Because we are dealing with non-Newtonian fluids, those results have to be corrected.

The flow through the capillary can be generated by controlled pressure or by controlled volume displacement inside a cylinder or reservoir connected to the capillary. Some manufactures install a small extruder to the reservoir to guarantee a permanent and homogeneous supply of material. Because it is easier to measure pressure than volumetric flow, the majority of capillary rheometers use the controlled movement of a piston inside the cylinder of the reservoir to generate a constant flow rate. Q in Equation (44.7) is then a known and controlled magnitude for the experiment. It is only necessary to measure the pressure losses Δp along a length inside the channel where the flow can be supposed as fully developed. This is not an easy task.

Cylindrical capillary is normally small in diameter, and it is not possible to measure the gradient pressure for the flow portion free of entrance effects inside the capillary. The pressure transducer is located closed to the capillary entrance and its readings include entrance losses. Ryder–Bagley’s correction has to be applied to eliminate this error (Macosko, 1994; Osswald and Menges, 2003; Menges, 2002; Carreau et al., 1997). The method to apply this correction implies to run several measurements using at least three capillaries with the same diameter but with different length. The pressure value obtained from a linear extrapolation of the curves for an ideal null length capillary represents the pressure drop at the capillary inlet. Some authors use the intersection of the curve for null

pressure as the equivalent length of capillary for the pressure readings. Both methods are equivalent.

Because the viscosity of polymer melts increases with pressure (Sedlacek et al., 2004; Liang, 2001), the Ryder–Bagley’s method delivers, in some cases, negative values for the pressure corrections or equivalent length shorter than the actual length of capillaries. Some authors recommend to use a quadratic approximation instead of the linear one to calculate Bagley’s correction.

Rectangular and annular capillaries are usually big enough to locate at least two transducers but preferably three pressure transducers in a region of the capillary with a fully developed steady flow to measure the pressure gradient. Ryder–Bagley’s correction is not necessary in this case. With three pressure transducers, it is possible to detect a nonlinear profile for the pressure gradient, which has to be considered during the interpretations of readings.

As mentioned above, a viscosity value obtained from Equation (44.5) is only valid for Newtonian fluids and it is called *apparent shear viscosity* when applied to non-Newtonian fluids like polymers. To obtain the actual shear viscosity, the apparent one should be corrected. The Weissenberg–Rabinowitsch (Macosko, 1994; Osswald and Menges, 2003; Menges, 2002; Carreau et al., 1997) correction is the most used method for it. To apply this correction method, it is necessary to run several measurements under different shear rates, using the same capillary and keeping a constant polymer temperature for all runs. Some rheometer manufacturers offer software solutions as an integral part of equipment to make both Bagley and Weissenberg–Rabinowitsch corrections.

The Bagley’s correction is very time-consuming. Some manufacturers offer rheometers equipped with several capillaries arranged in a parallel array, usually three, with different lengths and one of them with null length. With them, it is possible to obtain from one experimental run three readings for the Bagley’s correction, which is carried out automatically by the equipment.

Other rheological properties of polymer melts, regarding elastic behavior, can be measured using capillary rheometers. When the material comes out from the capillary, it increases the cross section due to the tendency of oriented macromolecules to recover their nonoriented state. The ratio between the diameter of the capillary and the strand, called die swell ratio, is used to characterize this behavior. Based on die swell measurement, it is possible to calculate the first normal stress difference and indirectly the elongational viscosity using the Cogswell’s principle (Cogswell, 1981) but this method is not applicable for all polymers.

Capillary rheometers deliver information of shear viscosity for shear rate from about 10 s^{-1} to 10^4 s^{-1} combining rectangular capillaries for low shear rates (10 s^{-1} to 10^2 s^{-1}) with circular (100 s^{-1} to 10^4 s^{-1}) ones. This range is usually enough to get information about the rheological behavior of melts for most common polymer processing techniques such as extrusion, blow molding, and injection molding. To obtain a more complete description of this behavior, it is necessary to get information about shear viscosity at very low shear rates or null viscosity (η_0). This is a very important parameter for the most used rheological models in commercial software (Johannaber and Michaeli, 2001) to express the shear viscosity as a function of shear rate: the Carreau’s model in Equation (44.8).

$$\eta = \frac{\eta_0}{(1 + B \cdot \dot{\gamma})^C} \quad (44.8)$$

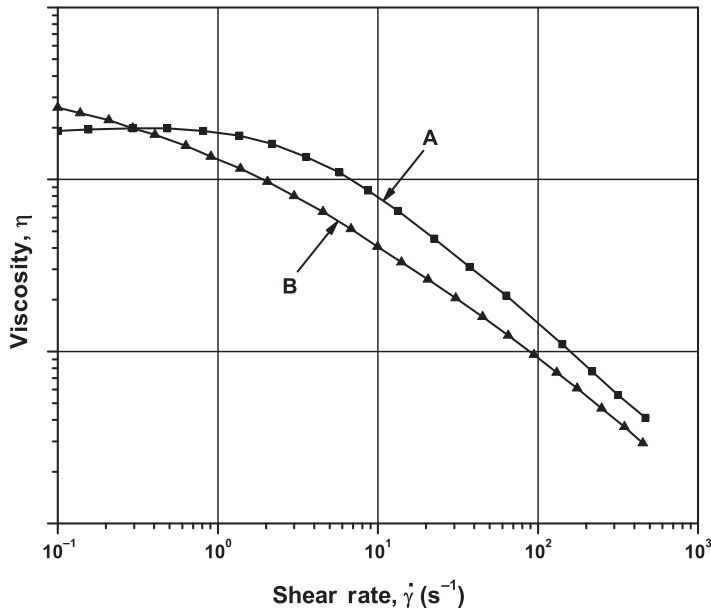


FIGURE 44.13 Shear viscosity for two different polymers at the same temperature.

The use of a rectangular channel instead of a circular one to obtain low shear rates is not the only reason for its use. The bigger dimension of those capillaries make possible to measure the gradient pressure in the channel region where the entrance effects are not present. This makes unnecessary the time-consuming Bagley's correction. Because of the relatively big cross section of capillary, the pressure to measure highly viscous materials such as rubber is not so high as with circular capillaries.

Shear viscosity curves are not only indispensable for rheological calculations of polymer flows in injection molding, extrusion dies, plasticating screws, and so on. They deliver important information about the molecular structure of polymers. They are very useful in comparing different materials, if the viscosity curves are valid for the same temperature. In Figure 44.13, for example, polymer A has a narrower molecular weight distribution (MWD) than polymer B and shows a more extended Newtonian behavior with a more accentuated dependency of viscosity with shear rate in the non-Newtonian region. Polymer B has a higher molecular weight than polymer A and exhibits higher viscosity values in the Newtonian region. This is valid for the majority of polymer melts.

44.4.3 Rotational Rheometry

In this rheometric technique, the substance being studied is deformed between two bodies where one of them remains stationary and the other one rotates. The bodies can be two concentric cylinders (Couette's rheometer) where usually only the internal cylinder rotates, or two horizontal discs, where the inferior disc is stationary. The Couette's rheometer is suited to measure low-viscosity substances (under 100 Pa.s) and at high shear rates. The substance remains in an almost closed space and cannot be expelled by centrifugal

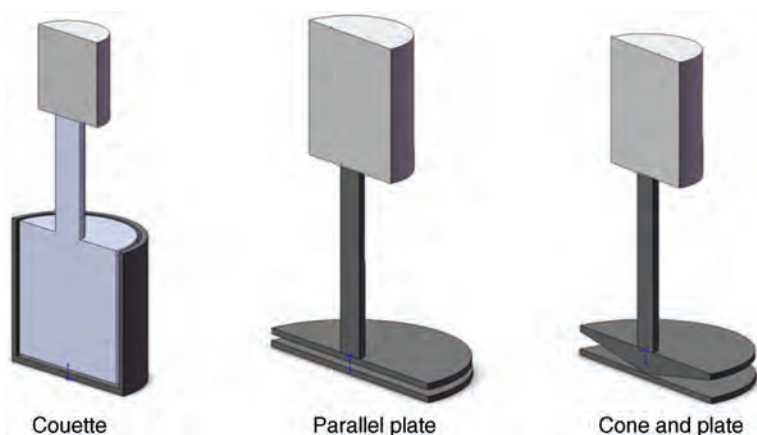


FIGURE 44.14 Types of rotational rheometers.

force. The most commonly used rotational rheometers for characterization of polymer melts use discs. One of the discs, usually the rotating one, may be a circular plate or a cone. The stationary disc is flat. The combination of cone and plate has an advantage to generate a constant shear rate in the whole region between both discs. Both discs are placed in an open space inside an oven. This limits the rotation speed, especially for substances with low viscosity at test temperature.

Theoretically, rotational rheometers do not have limitations for low shear rates (only limited by sensitivity of transducers and control unit). They deliver more information about material rheological behavior than capillary rheometers do, for example, information about the elongational and dynamic viscosities. This rheological information is essential for structural rheology, which deals with the relationship between molecular structure of polymers (molecular weight, MWD, molecular branches) and their rheological behavior (Figure 44.14).

Because of the viscoelastic properties of polymer melts during the shear deformation of the material between the discs it appears a normal force that tries to separate both discs (Weissenberg effect). This normal force to the discs generates primary and secondary normal stresses that can be expressed like shear viscosity as a function of shear rate and temperature. The primary normal stress can be best measured by cone-plate rheometers.

There are two basic types of rotational rheometers depending on which quantity is being controlled and which is being measured during an experiment: the shear stress or the deformation rate. Rheometers with controlled stress are called controlled stress rheometer (CSR, with controlled torque). Rheometers with controlled rate of deformation are known as controlled rate rheometer (CRR, with controlled rotational speed). CSR allows the generation of smaller shear rates than CRR. CRR can generate higher shear rates in continuous operation mode or higher frequencies in oscillatory mode than CSR. Modern rheometers can operate in both modes.

Rotational rheometers can carry out stationary, oscillatory, and transient experiments, depending on what is controlled and measured: torque (stress) or angular velocity (deformation).

44.4.3.1 Stationary Mode:

- *Deformation Rate is Kept Constant:* angular velocity is adjusted to a desired value and the torque is measured.
- *Stress is Kept Constant:* torque is adjusted to a desired value and the angular velocity is measured.

Stationary measurements are suitable to obtain the shear viscosity for very low shear rates, typically from 10^{-6} to 10^2 s^{-1} , and are an excellent complement to measurements with capillaries.

Shear viscosity can be calculated, like in (5), as the quotient of shear stress and shear rate. For the case of a cone and plate rheometer, where one disc is rotating with a constant angular velocity Ω and the cone angle θ_0 is very small ($<5^\circ$), the shear rate $\dot{\gamma}$ is constant in the region between discs and is given by

$$\dot{\gamma} = \frac{\Omega}{\theta_0}. \quad (44.9)$$

The shear stress can also be considered as constant in the region between discs and can be calculated from the measured torque T using

$$\tau = \frac{3T}{2\pi R^3}. \quad (44.10)$$

The primary normal stress coefficient function, ψ_1 can be obtained from the force F , trying to separate both discs, using

$$\psi_1 = \frac{2F}{\pi R^2 \cdot \dot{\gamma}^2}. \quad (44.11)$$

The secondary normal stress coefficient function, ψ_2 is very difficult to measure and it is usually approximated to $0.1\psi_1$.

For the case of a parallel plate rheometer, where one disc, with radius R , is rotating with a constant angular velocity Ω and the distance between plates is h , the shear rate $\dot{\gamma}$ can be calculated using the following equation and supposing that the fluid is Newtonian.

$$\dot{\gamma} = \frac{\Omega \cdot R}{h}. \quad (44.12)$$

The shear stress is given by

$$\tau = \frac{2T}{\pi \cdot R^3}. \quad (44.13)$$

Here again, the viscosity must be corrected because the equations above are only valid for Newtonian fluids. The Weissenberg–Rabinowitsch correction can be used for it.

Compared with cone and plate rheometers, parallel plate rheometers have the disadvantage to need the Weissenberg–Rabinowitsch correction, but they have several advantages. The shear rate can be easily changed varying the distance h between discs, as can be observed in Equation (44.12), still having a low rotational speed. With this strategy, it is possible to keep border and inertial effects at low levels. The specimen

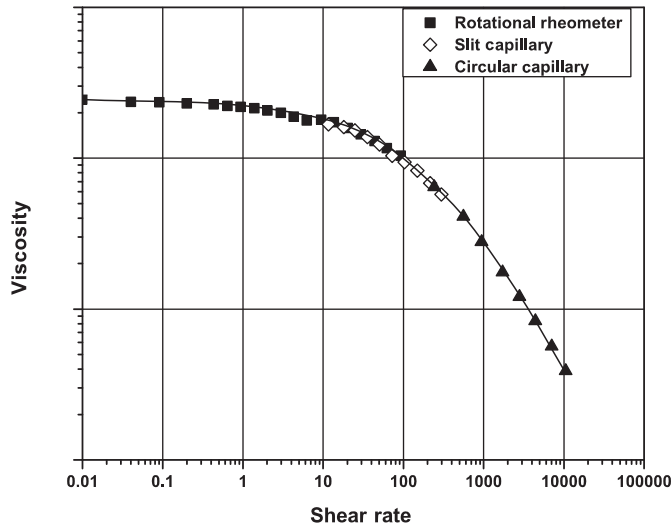


FIGURE 44.15 Viscosity curve using rotational and capillary rheometers.

preparation for parallel plate rheometers is simpler. They are more suitable to measure viscous materials at low shear rates.

Border and inertial effects are difficult to detect because the discs are confined in an isolated and temperature-controlled closed chamber. Some manufacturers offer an optical probe to observe the material during the experiment.

Figure 44.15 illustrates the use of several techniques (rotational, rectangular, and circular capillary) to obtain a complete viscosity curve.

44.4.3.2 Oscillatory Mode The amplitude and the frequency of angular oscillation are adjusted to desired values (in the linear range of deformation) and the torque variation is measured.

The amplitude and the frequency of torque oscillation are adjusted to desired values (assuring that a linear range of deformation is not exceeded), and the torque variation is measured.

In oscillatory tests, viscoelastic materials show a retarded response of deformation under oscillatory stress (Macosko, 1994; Carreau et al., 1997). For the case of shear experiments, the quotient of the amplitude of shear stress and the amplitude of shear deformation is the dynamic shear module G^* . It has two components: G' describes the elastic component of the response and is called the storage module. G'' is called the loss module and describes the viscous component of G^* .

The equations are simple and the modules can be calculated from the oscillatory (sinusoidal) variation of torque T , the frequency Ω , and the phase difference δ of the disc using Equations (44.9) and (44.10) for a cone and plate rheometer.

$$\gamma(t) = \gamma_0 \cdot \sin(\Omega \cdot t) \quad (44.14)$$

$$\tau(t) = \tau_0 \cdot \sin(\Omega \cdot t + \delta). \quad (44.15)$$

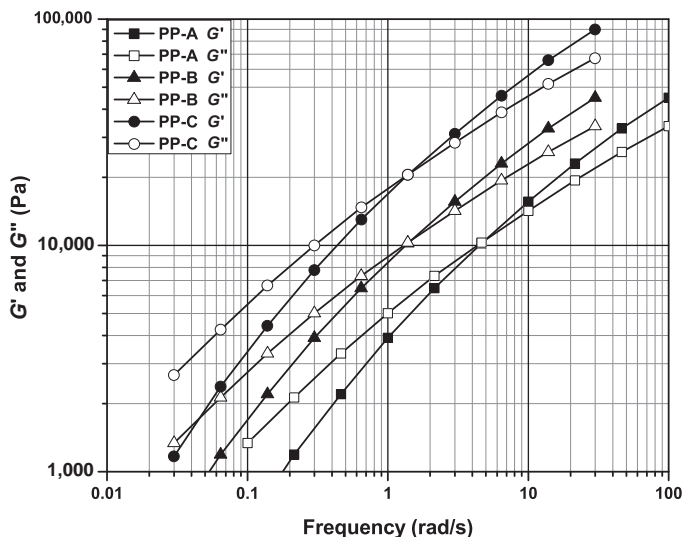


FIGURE 44.16 Modules for three different types of polypropylene.

44.4.3.3 Transient Mode

Creep or Retardation Test a shear stress is suddenly applied (stress step), the shear strain as angular displacement is followed over a certain period of time, and the stress is released suddenly.

Relaxation Test a strain as angular displacement is applied and the variation of the resulting shear stress is registered over time.

The following tests are other transient methods:

Relaxation Under Shear Rate a constant angular deformation rate is suddenly applied, and the resulting declining shear stress is registered over time until a steady-state condition is established. After this, the angular deformation ceases, and the declining shear stress is registered over time.

In all these experiments it should be assured that the material never exceeds the linear viscoelastic range.

44.4.3.4 Sweeps There are special tests for rotational rheometer called “sweeps.” It means that one operation parameter can be varied continuously inside a range. For example, to identify the linear viscoelasticity region, a torque sweep is recommended. In oscillatory modus a frequency sweep allows to measure the shear viscosity as a function of frequency. This curve has a univocal translation to the viscosity as a function of shear rate (Cox–Merz principle) if deformation amplitude remains in the linear viscoelastic range. A temperature sweep is very useful to find out how properties such as viscosity, storage or loss modules vary with the temperature at a constant frequency. With this information, and using the WLF principle of time–temperature superposition, it is possible to convert temperature into time. Using time

sweeps, the sample undergoes an oscillatory deformation at a frequency and fixed amplitude is a way to obtain information about other rheological behaviors such as rheopexy and tixotropy.

44.4.3.5 WLF Principle of Time–Temperature Superposition Figure 44.16 shows the storage and the loss modules for three different types of polypropylene. The intersection of both curves for the same material gives information about its molecular structure. The higher the modules at crossing point, the narrower the molecular weight distribution is. The lower the frequency at the crossing point, the higher the molecular weight is. Material A has the lowest molecular weight, and Material C has the narrower molecular weight distribution of these three materials.

REFERENCES

- ASTM D 5576: *Practice for determination of structural entities in Polyolefins by FTIR*. ASTM vol. 8.03, p. 584–586.
- Benoit H, Grubisic Z, Rempp P, Decker D, Zillox J-G. Liquid-phase chromatographic study of branched and linear polystyrenes of known structure. *Journal de Chimie Physique* 1966;63: 1507–1514.
- Blaine R. Method and apparatus of modulated-temperature thermogravimetry. TA Instruments, Inc., U.S. Patents No. 6,113,261 (2000) and 6,336,741 (2002).
- Carballeira-Amarello T, Castro López MM, Noguerol Cal R, Dopico García MS, López-Vilariño JM, González-Rodríguez MV. *Antistatics in Food Contact Materials*. España: Dpto. de Química Analítica, E. U. Politécnica.
- Carreau, PJ, De Kee D, Chhabra RP. *Rheology of Polymeric Systems – Principles and Applications*. München: Carl Hanser Verlag; 1997.
- Cogswell, FN. *Polymer Melt Rheology*. London: Wiley; 1981.
- Conley RT. *Infrared Spectroscopy*. 2nd ed. Allyn and Bacon Inc.; 1972.
- Crompton R. *Determination of Additives in Polymers and Rubbers*. Rapra Technology; 2007.
- Ehrenstein G, Riedel G, Trawiel P. *Thermal Analysis of Plastics*. Munich: Hanser Publishers; 2004.
- Frontier Laboratories Ltd. *Composition Analysis of Fully Aromatic Polyester by Py-GC Utilizing Reactive Pyrolysis*. Double-Shot Pyrolyzer-Application Note PYA2-004E. Frontier Laboratories Ltd.
- Fuller MP, Garry MC, Stanek Z. *An FTIR Liquid Analyser*. American Laboratory.
- Gartner C, Sierra JD, Avakian R. *New Polyolefins Characterization by Instrumental Analysis*. Atlanta: ANTEC; April 1998.
- Gratzfeld-Huesgen A. *Analysis of Antioxidants and UV Stabilizers in Polymers using HPLC*. Agilent Technologies; 1997.
- Grubisic Z, Rempp P, Benoit HJ. Universal calibration for gel permeation chromatography. *Polymer Science Part B: Polymer Letters* 1967;5:753–759.
- Hemminger WF, Cammenga HK. *Methoden der Thermischen Analyse*. Springer-Verlag, Berlin; 1988.
- Johannaber F, Michaeli W. *Handbuch Spritzgießen*. München: Carl Hanser Verlag; 2001.
- Kazakevich Y, Lohrutto R. *HPLC for Pharmaceuticals Scientist*. Wiley; 2007.
- Knappe S, Kaisersberger E, Moehler H. *Netsch Annual for Science and Industry*. NETZSCH-Geratebau GmbH. Selb Germany; 1993;2.

- Kubies D, Scudla J, Puffr R, Sikora A, Baldrian J, Kovarova J, Šlouf M, Rypacek F. Structure and mechanical properties of poly(L-lactide)/layered silicate nanocomposites. *European Polymer Journal* 2006;42:888–899.
- Liang, J-Z. Pressure effect of viscosity for polymer fluids in die flow. *Polymer* 2001;42:3709–3712.
- Lomonte JN. IR determination of vinylidene unsaturation in polyethylene. *Analytical Chemistry* 1962;34(1):129–131.
- Macosko, CW. *Rheology Principles, Measurements and Applications*. 1st ed. New York: Wiley-VCH; 1994. p. 550.
- McLafferty F, Turecek F. *Interpretation of Mass Spectra*. Wiley; 1993.
- Menges, C *et al.* *Werkstoffkunde Kunststoffe*. 5th ed. München: Carl Hanser Verlag; 2002. p. 411.
- Naranjo A, Noriega M del P, Osswald T, Roldán A, Sierra J. *Plastics Testing and Characterization, Industrial Applications*. Munich: Hanser Publishers; 2008.
- National Institute for Standard Technology. *NIST standard reference database number 69*, 2005.
- Noriega, M, del P, Rauwendaal, C. *Troubleshooting the Extrusion Process*. Munich: Hanser Publishers; 2001.
- Osswald T, Menges G. *Materials Science of Polymers for Engineerings*. 2nd ed. München: Hanser Publisher; 2003.
- Patnaik P. *Dean's Analytical Chemistry Handbook*. 2nd ed. McGraw-Hill; 2004.
- Rudrabhatla M. *Analysis of Polymer Antioxidant Additives on the Varian 500-MS LC Ion Trap*. Application Note #020, Varian Inc.
- Scott RPW. *Liquid Chromatography*. Chrom-Ed Book Series; 2003.
- Sedlacek, T, C *et al.* On the effect of pressure on the shear and elongational viscosities of polymer melts. *Polymer* 2004;44(7):1328–1337.
- Sierra JD, Ospina S, Montoya N, Noriega MP, Osswald TA. *Characterization of Polyethylene Blends by Using Novel Techniques Such as the Successive Self-Nucleation and Annealing (SSA) and the Fourier Self-Deconvolution IR Spectroscopy (FSD-IR)*. Orlando: ANTEC; 2000.
- Sierra J, Noriega M, del P, Ospina S, Cardona E. The Effect of a Citrate Plasticizer on the Thermal and Rheological Properties of Polylactic Acid. ANTEC 2009 Conference Proceedings, SPE, Chicago, IL, USA, June, 2009.
- Stilianos Rowis G, Fedora JW. Use of a thermal extraction unit for furnace-type pyrolysis: Suitability for the analysis of polymers by pyrolysis/GC/MS. *Rapid Communications in Mass Spectrometry* 1996;10:82–90.
- Thomas L. *Modulated DSC, TA Instruments*. Springer-Verlag, Berlin; 2006.
- Wampler T. *Analysis of an Acrylic Copolymer Using Pyrolysis-GC/MS*. Perkin Elmer; 2005.

ANALYTICAL TOOLS FOR ESTIMATION OF PARTICULATE COMPOSITE MATERIAL PROPERTIES

T. I. ZOHDI AND MAGD E. ZOHDI

- 45.1 Introduction
- 45.2 Concepts in statistical quality control
- 45.3 Effective property estimates
 - 45.3.1 Elementary estimates
 - 45.3.2 Improved estimates
- 45.4 Summary
- References

45.1 INTRODUCTION

Most modern devices owe a significant amount of their success to the tailored electromagnetic and mechanical material behavior of the components that comprise them. A relatively inexpensive way to obtain macroscopically desired responses is to enhance an easy-to-form matrix material's properties by introducing microscale particles possessing different electromagnetic and mechanical properties (Figure 45.1). The particles are chosen to produce an overall desired "effective property." The aggregate response of the material is an outcome of the interaction between the smaller-scale (microstructure) constituents that comprise the "effective" material. In the construction of such materials, the basic philosophy is to select matrix/particle material combinations in order to produce desired aggregate responses. For example, in many engineering applications, the classical choice is to add a particulate phase with suitable properties in order to modify the overall properties of a modable base matrix material.

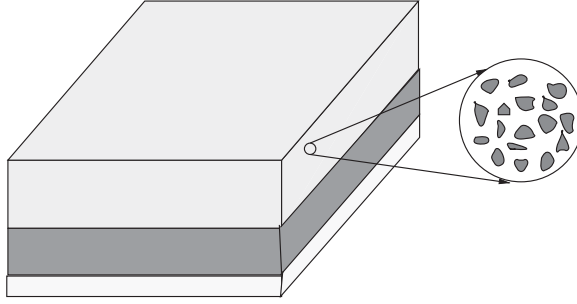


FIGURE 45.1 A representative sample of a material with heterogeneous microstructure.

The design of desired effective electrical properties of materials, by manipulating the heterogeneous microstructure, is a critical to the success of modern devices (Figure 45.1). In the context of electrical materials, the microscale properties are characterized by a spatially variable permittivity ϵ . Typically, in order to characterize the structural-scale effective response of such materials, a relation between averages

$$\langle \mathbf{D} \rangle_{\Omega} = \epsilon^* \cdot \langle \mathbf{E} \rangle_{\Omega} \quad (45.1)$$

is sought, where $\langle \cdot \rangle_{\Omega} \stackrel{\text{def}}{=} \frac{1}{|\Omega|} \int_{\Omega} \cdot \, d\Omega$ is the averaging operator and \mathbf{D} and \mathbf{E} are the electrical flux (density) and electric field within a statistically representative volume element (RVE) of volume $|\Omega|$. The quantity, ϵ^* , is known as the effective permittivity and is the property used in usual (homogenized) macroscale analyses. Similarly, for magnetic properties one has

$$\langle \mathbf{B} \rangle_{\Omega} = \mu^* \cdot \langle \mathbf{H} \rangle_{\Omega} \quad (45.2)$$

where \mathbf{B} and \mathbf{H} are the magnetic flux and magnetic fields, and μ^* is the effective magnetic permeability.

Remark: The same framework holds for linear elasticity

$$\langle \boldsymbol{\sigma} \rangle_{\Omega} = \mathbf{E}^* : \langle \boldsymbol{\epsilon} \rangle_{\Omega} \quad (45.3)$$

where $\boldsymbol{\sigma}$ and $\boldsymbol{\epsilon}$ are the stress and strain fields, and \mathbf{E}^* is the effective linear-elasticity tensor.

45.2 CONCEPTS IN STATISTICAL QUALITY CONTROL

In classical quality control techniques involving inspection, the following steps must be taken:

1. sample a stream of products by taking N samples, each of size Ω ,
2. measure the desired quantity of interest inherent to the sample,
3. calculate the deviations of the quantity of interest and
4. plot/tabulate the data on a control chart/table.

The arithmetic mean of the samples is the primary measure of central tendency. The symbol \bar{A} is used to designate the arithmetic mean of the sample and may be expressed in

algebraic terms as $\bar{A} = \frac{1}{N} \sum_{i=1}^N A_i$, where the A_i is the quantity of interest extracted from various samples, $i = 1, 2, \dots, N$. The standard deviation is another useful number to characterize the dispersion of between the sets of numbers and is defined by $\delta_A = \sqrt{\frac{1}{N} \sum_{i=1}^N (A_i - \bar{A})^2}$. One important measure of the deviation, in fact the one we will focus on, is the range, R_A , defined as

$$R_A = A_{\max} - A_{\min}. \quad (45.4)$$

Although the distribution of the A values can be of any shape, the distribution of \bar{A} values tends to be close to a normal distribution. The larger the sample size and the more normal the universe, the closer the distribution of the average \bar{A} 's approach the normal curve. According to statistical theory, (the Central Limit Theorem), in the "long run," the average of \bar{A} values will be the same as A^* , the average of the universe (a statistically representative sample). The focus of this presentation is to introduce the reader to estimation techniques for the effective property of a sample of composite material, in particular, the response of an infinitely large sample, relative to the microscopic constituents that comprise it. Furthermore, we show the reader estimates on the range (upper and lower bounds) of possible properties, for a given set of constituents (components). Such concepts are standard in manufacturing; however, they are also useful for characterizing the properties of heterogeneous materials, which inherently require sample size selection to ascertain statistically representative effective material properties for use in structural scale calculations. *If test samples of material are too small (not truly statistically representative), then the overall effective property will have variations in it from sample to sample. Here, we focus on the associated problem of a priori prediction of the effective property that would result if an infinitely large statistically representative sample or, alternatively, an infinite number of finite-sized samples were tested.* This allows an engineer to determine confidence intervals for the material properties that they can employ in structural-scale calculations.¹ For an introduction to cross-disciplinary classical statistical quality control techniques, such as (a) measurements and quality control, (b) dimension and tolerance, (c) quality control, (d) interrelationship of tolerances of assembled products, (e) control charts for attributes, and (f) acceptance sampling, see Zohdi (2006).

45.3 EFFECTIVE PROPERTY ESTIMATES

For an RVE sample to be statistically representative, it must relatively large compared with the length-scale of the constituents (e.g., particles in a matrix), containing a significant amount of microstructure. Therefore, the computations to determine effective properties, which would involve some type of numerical discretization posed over the RVE, are nontrivial. Essentially, a finite element or finite difference mesh/grid (Figure 45.2) must be fine enough to resolve the microscopic features of the composite material.

Because of the difficulties in computing effective properties directly, a variety of approximate techniques have been developed to estimate the overall macroscopic properties of materials consisting of a matrix containing distributions of particles, dating back at least to Maxwell (1867, 1873) and Lord Rayleigh (1892).

¹ An exhaustive treatment of sample size selection and estimates of statistical variation based on ensemble averaging is given in Zohdi and Wriggers (2008).

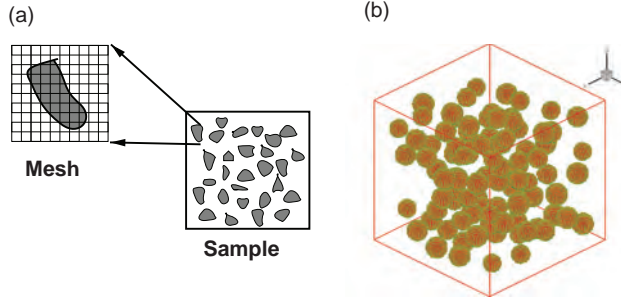


FIGURE 45.2 (a) A typical numerical grid needed to resolve a composite material microstructure. (b) Numerical resolution of the morphology of a test sample's microstructure (Zohdi, 2010).

45.3.1 Elementary Estimates

In the realm of solid mechanics, Voigt (1889) is usually cited with the first analysis of the linear effective *mechanical* properties of the microheterogeneous solids, $\langle \sigma \rangle_{\Omega} = \mathbf{I} \mathbf{E} : \langle \epsilon \rangle_{\Omega}$, where σ is the stress and ϵ is the strain. Voigt approximated the strain field within an aggregate sample of heterogeneous material as being uniform, leading to $\langle \mathbf{I} \mathbf{E} \rangle_{\Omega}$ as an expression of the effective property. Reuss (1929) approximated the stress fields within the aggregate of polycrystalline material as uniform, leading to $\langle \mathbf{I} \mathbf{E}^{-1} \rangle_{\Omega}^{-1}$ as the effective property. In 1952, Hill (1952) proved that these assumptions provide bounds on the effective property, namely, $\langle \mathbf{I} \mathbf{E}^{-1} \rangle_{\Omega}^{-1} \leq \mathbf{I} \mathbf{E}^* \leq \langle \mathbf{I} \mathbf{E} \rangle_{\Omega}$. These inequalities mean that the eigenvalues of the tensors $\mathbf{I} \mathbf{E}^* - \langle \mathbf{I} \mathbf{E}^{-1} \rangle_{\Omega}^{-1}$ and $\langle \mathbf{I} \mathbf{E} \rangle_{\Omega} - \mathbf{I} \mathbf{E}^*$ are non-negative. Therefore, one can interpret the Voigt and Reuss fields as providing two microfield extremes, since the Voigt stress field is one where the tractions at the phase boundaries cannot be in equilibrium (statically inadmissible), while the implied Reuss strains are such that the heterogeneities and the matrix could not be perfectly bonded, that is (kinematically inadmissible). These results can be easily reinterpreted for electrical and magnetic and properties and fields to yield

$$\langle \epsilon^{-1} \rangle_{\Omega}^{-1} \leq \epsilon^* \leq \langle \epsilon \rangle_{\Omega} \quad (45.5)$$

where the upper bound is generated by assuming that the electric field is uniform throughout the medium and the lower bound is generated by assuming that the electric field flux is uniform throughout the medium. For the magnetic properties one has

$$\langle \mu^{-1} \rangle_{\Omega}^{-1} \leq \mu^* \leq \langle \mu \rangle_{\Omega} \quad (45.6)$$

where the upper bound is generated by assuming that the magnetic field is uniform throughout the medium and the lower bound is generated by assuming that the magnetic field flux is uniform throughout the medium. In the electromagnetics literature, the bounds in Equations (45.5) and 45.6 are frequently referred to as the Wiener bounds (Wiener, 1910). These inequalities mean that the eigenvalues of the tensors $\epsilon^* - \langle \epsilon^{-1} \rangle_{\Omega}^{-1}$ and $\langle \epsilon \rangle_{\Omega} - \epsilon^*$ are non-negative. Typically, the bounds are quite wide and provide only rough qualitative information.

45.3.2 Improved Estimates

Within the last 50 years improved estimates have been pursued, with a notable contribution being the Hashin–Shtrikman bounds (Hashin and Shtrikman, 1962a, 1962b, 1963; Hashin, 1983). The Hashin–Shtrikman bounds are the tightest possible bounds on scalar isotropic effective responses (ε^*, μ^*) , generated from isotropic microstructures $(\varepsilon_1, \mu_1, \varepsilon_2, \mu_2)$, where the volumetric data and phase contrasts of the constituents are the only data known. For the overall permittivity

$$\langle \varepsilon^{-1} \rangle_{\Omega}^{-1} \leq \varepsilon_1 + \underbrace{\frac{v_2^\varepsilon}{\frac{1}{\varepsilon_2 - \varepsilon_1} + \frac{1 - v_2^\varepsilon}{3\varepsilon_1}}}_{\varepsilon^{*, -}} \leq \varepsilon^* \leq \varepsilon_2 + \underbrace{\frac{1 - v_2^\varepsilon}{\frac{1}{\varepsilon_1 - \varepsilon_2} + \frac{v_2^\varepsilon}{3\varepsilon_2}}}_{\varepsilon^{*, +}} \leq \langle \varepsilon \rangle_{\Omega} \quad (45.7)$$

and for the overall permeability

$$\langle \mu^{-1} \rangle_{\Omega}^{-1} \leq \mu_1 + \underbrace{\frac{v_2^\mu}{\frac{1}{\mu_2 - \mu_1} + \frac{1 - v_2^\mu}{3\mu_1}}}_{\mu^{*, -}} \leq \mu^* \leq \mu_2 + \underbrace{\frac{1 - v_2^\mu}{\frac{1}{\mu_1 - \mu_2} + \frac{v_2^\mu}{3\mu_2}}}_{\mu^{*, +}} \leq \langle \mu \rangle_{\Omega} \quad (45.8)$$

where $\varepsilon_2 \geq \varepsilon_1$, $\mu_2 \geq \mu_1$, v_2^ε is the volume fraction of phase with the higher ε value (“phase 2” in the former expression) for the permittivity-mismatch and v_2^μ is the volume fraction of the phase with the higher μ value (“phase 2” in the latter expression) for the permeability-mismatch.² The typical behavior of the bounds and their range, plotted against volume fraction, is shown in Figures 45.3 and 45.4. Typically, dielectric materials are measured relative to the vacuum-level permittivity, $\varepsilon_0 = 8.854 \times 10^{-12}$ farads/m. The usual representation is $\varepsilon_{\text{ir}} \stackrel{\text{def}}{=} \frac{\varepsilon_{\text{r}}}{\varepsilon_0}$, where ε_{ir} is known as the relative permittivity.³ In this example, we chose a mixture of polyethylene ($\varepsilon_{1r} = \frac{\varepsilon_1}{\varepsilon_0} = 2.25$) and silicon ($\varepsilon_{2r} = \frac{\varepsilon_2}{\varepsilon_0} = 11.68$), both measured at STP, for 0.9 MHz (Hector and Schultz, 1936).

For linearized elasticity applications, another form of the Hashin and Shtrikman bounds (Hashin and Shtrikman, 1962a, 1962b, 1963; Hashin, 1983), for isotropic materials with isotropic effective (mechanical) responses are commonly used. For the bulk modulus

$$\kappa^{*, -} \stackrel{\text{def}}{=} \kappa_1 + \frac{v_2}{\frac{1}{\kappa_2 - \kappa_1} + \frac{3(1 - v_2)}{3\kappa_1 + 4G_1}} \leq \kappa^* \leq \kappa_2 + \frac{1 - v_2}{\frac{1}{\kappa_1 - \kappa_2} + \frac{3v_2}{3\kappa_2 + 4G_2}} \stackrel{\text{def}}{=} \kappa^{*, +}, \quad (45.9)$$

and for the shear modulus

$$G^{*, -} \stackrel{\text{def}}{=} G_1 + \frac{v_2}{\frac{1}{G_2 - G_1} + \frac{6(1 - v_2)(\kappa_1 + 2G_1)}{5G_1(3\kappa_1 + 4G_1)}} \leq G^* \leq G_2 + \frac{(1 - v_2)}{\frac{1}{G_1 - G_2} + \frac{6v_2(\kappa_2 + 2G_2)}{5G_2(3\kappa_2 + 4G_2)}} \stackrel{\text{def}}{=} G^{*, +}, \quad (45.10)$$

² For either case, the volume fraction of the other phase is v_1 , where $v_1 + v_2 = 1$.

³ Similar representations can be made for the magnetic properties using the vacuum-level permeability, $\mu_0 = 8.854 \times 10^{-12}$ farads/m and $\mu_{\text{ir}} \stackrel{\text{def}}{=} \frac{\mu_{\text{r}}}{\mu_0}$, where μ_{ir} is known as the relative permeability.

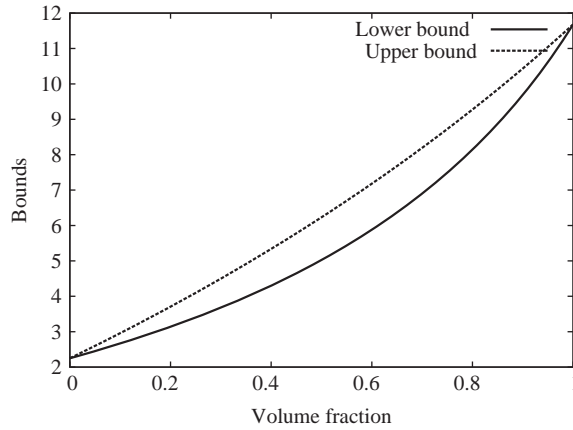


FIGURE 45.3 An example (Zohdi, 2012) of the Hashin–Shtrikman bounds for a mixture of polyethylene ($\varepsilon_{1r} = \frac{\varepsilon_1}{\varepsilon_0} = 2.25$) and silicon ($\varepsilon_{2r} = \frac{\varepsilon_2}{\varepsilon_0} = 11.68$), both measured at STP for 0.9 MHz (Hector and Schultz, 1936).

where κ_2 and κ_1 are the bulk moduli, and G_2 and G_1 are the shear moduli of the respective phases ($\kappa_2 \geq \kappa_1$ and $G_2 \geq G_1$), and where v_2 is the second phase volume fraction. Such bounds are the tightest possible on isotropic effective responses, with isotropic two phase microstructures, where only the volume fractions and phase contrasts of the constituents are known. Note that no geometric or distributional information is required for the bounds. Thus, in summary, for a composite materials designer, the range of possible effective responses of composite material for a given set of matrix and particle materials is

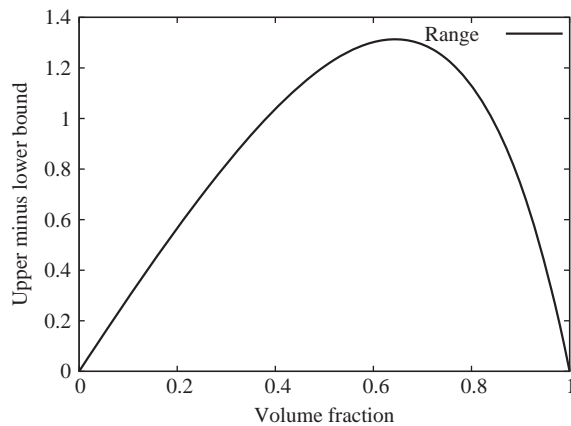


FIGURE 45.4 An example (Zohdi, 2012) of the *difference/range* in the upper and lower Hashin–Shtrikman bounds, $R_{\varepsilon^*} = \frac{\varepsilon^{*,+} - \varepsilon^{*,\circ}}{\varepsilon_0}$, for a mixture of polyethylene ($\varepsilon_{1r} = \frac{\varepsilon_1}{\varepsilon_0} = 2.25$) and silicon ($\varepsilon_{2r} = \frac{\varepsilon_2}{\varepsilon_0} = 11.68$), both measured at STP for 0.9 MHz (Hector and Schultz, 1936).

- For the electrical permittivity

$$R_{\epsilon^*} = \epsilon^{*,+} - \epsilon^{*,-} \quad (45.11)$$

- For the magnetic permeability

$$R_{\mu^*} = \mu^{*,+} - \mu^{*,-} \quad (45.12)$$

- For the elastic bulk modulus

$$R_{\kappa^*} = \kappa^{*,+} - \kappa^{*,-} \quad (45.13)$$

- For the elastic shear modulus

$$R_G = G^{*,+} - G^{*,-} \quad (45.14)$$

It is critical to note that the proofs of these bounds require sample sizes that are truly statistically representative (infinitely large compared with the particles in the sample). If the samples are too small (not statistically representative), then the bounds may be violated, and we again must resort to the estimates of statistical variation based on ensemble averaging, using Central Limit Theorem-type techniques (Zohdi and Wriggers, 2008).

45.4 SUMMARY

The measurement of effective properties of particulate composite materials, such as those of Hashin and Shtrikman (1962a, 1962b, 1963) and Hashin (1983), are an effective tool to use in the fast and reliable development of new composite materials. For a thorough analysis of many of such methods, see Torquato (2001), Jikov et al. (1994), Hashin and Shtrikman, (1962b), and Nemat-Nasser and Hori (1999) for solid-mechanics oriented treatments and Zohdi and Wriggers (2008) for computational aspects.

The identification of microstructural parameters, which force the system behavior to match a (desired) target response, is essentially a nonconvex inverse problem where the microstructural parameters are the design variables. For example, a relatively straightforward formulation is to consider inverse problems whereby microstructural parameters are sought, which deliver a desired overall behavior by minimizing a cost function such as

$$\Pi = w_1 \left(\frac{\|\epsilon^* - \epsilon^{*,D}\|}{\|\epsilon^{*,D}\|} \right)^2 + w_2 \left(\frac{\|\mu^* - \mu^{*,D}\|}{\|\mu^{*,D}\|} \right)^2 + \text{local constraints} \quad (45.15)$$

where $\epsilon^{*,D}$ and $\mu^{*,D}$ are desired overall properties and w_1 and w_2 are design weights that indicate the importance of achieving each component of the objective. Specifically, a microstructural design problem can be set up by defining a N -tuple design vector, denoted $\Lambda^{\text{def}}(\Lambda_1, \Lambda_2, \dots, \Lambda_N)$, for example, consist of the following components: (1) the properties of the particles (2) the volume fraction of the particles, and (3) the topological parameters of the particles. For example, ellipsoidal shapes are qualitatively useful since the

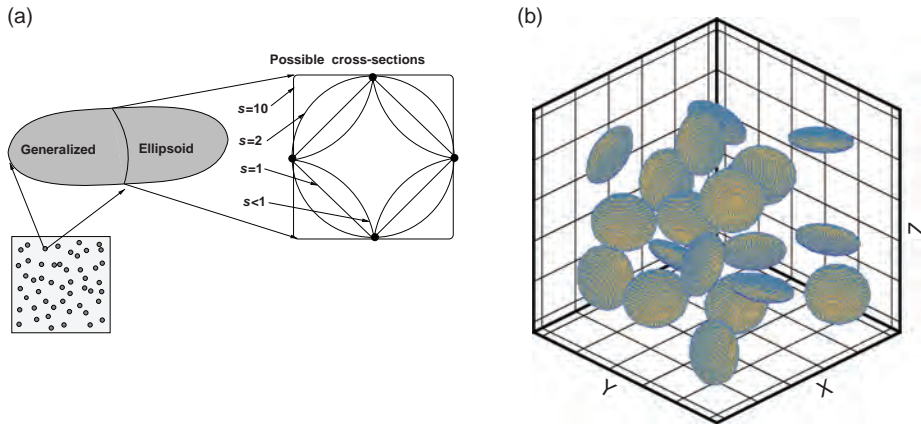


FIGURE 45.5 (a) A generalized ellipsoid and (b) a numerical representation of the nonspherical particulate additives using the finite element method (Zohdi, 2001).

geometry can closely represent a variety of particulate types, for example, platelets when the ellipsoids are oblate or needles (discontinuous fibers) when the ellipsoids are prolate. Such shapes can be generalized by considering the following (Figure 45.5):

$$\left(\frac{|x - x_0|}{r_1}\right)^{s_1} + \left(\frac{|y - y_0|}{r_2}\right)^{s_2} + \left(\frac{|z - z_0|}{r_3}\right)^{s_3} = 1, \quad (45.16)$$

where the s is exponents (Zohdi, 2001, 2003; Zohdi and Wriggers, 2008). Values of $s < 1$ produce nonconvex shapes, whereas $s > 2$ values produce “block-like” shapes. Generally, formulations of such objective functions (II) will possess nonconvex and nondifferentiable dependency on the design variables (especially if there are manufacturing constraints).⁴ The minimization of such objective functions can be achieved using a two step process whereby (1) one determines promising optimal regions in parameter space using (nonderivative) algorithms (e.g., evolutionary “genetic” algorithms, simulated annealing) and then (2) one applies (classical) gradient-based schemes in locally convex regions, if the objective function is smooth enough (since such approaches are generally extremely efficient for the minimization of smooth convex functions). For example, there are a variety of nonderivative algorithms that employ concepts of species evolution, such as reproduction, mutation, and crossover, which are referred to as “genetic” algorithms. Implementation of such algorithms typically involves a randomly generated population of “genetic” strings, each of which represents a specific choice of system parameters. The population of genes undergo “mating sequences,” “offspring production,” and other biologically inspired events in order to find regions of the search space where cost functions are small. These methods date back, at least, to pioneering work John Holland (1975). For general reviews, see Goldberg (1989), Goldberg and Deb (2000) and (Zohdi 2001, 2003; Zohdi and Wriggers, 2008) for specific optimization and genetic algorithms to treat

⁴ Additionally, such objective functions are usually “noisy,” that is, they possess a stochastic nature, which can be attributed to sample-size effects.

inverse problems involving various particulate material systems. Generally, to obtain more detailed information on the response of composite materials, such as the microscale stresses, crack propagation, and material failure; one must resort to numerical techniques, such as the finite element method. However, analytical methods, such as those illustrated here, provide a necessary starting point, in order to minimize computational tests, which typically require significant computation resources. For more on numerical methods for composite material analysis, we refer the reader to Zohdi and Wriggers (2008).

REFERENCES

- Goldberg DE. *Genetic Algorithms in Search, Optimization and Machine Learning*. Addison-Wesley; 1989.
- Goldberg DE, Deb K. Special issue on genetic algorithms. *Computer Methods in Applied Mechanics and Engineering* 2000;186(2–4):121–124.
- Hashin Z. Analysis of composite materials: a survey. *ASME Journal of Applied Mechanics* 1983;50:481–505.
- Hashin Z, Shtrikman S. On some variational principles in anisotropic and nonhomogeneous elasticity. *Journal of the Mechanics and Physics of Solids* 1962a;10:335–342.
- Hashin Z, Shtrikman S. A variational approach to the theory of effective magnetic permeability of multiphase materials. *Journal of applied Physics* 1962b;33(10):3125–3131.
- Hashin Z, Shtrikman S. A variational approach to the theory of the elastic behaviour of multiphase materials. *Journal of the Mechanics and Physics of Solids* 1963;11:127–140.
- Hector LG, Schultz HL. The dielectric constant of air at radiofrequencies. *Journal of Applied Physics* 1936;7:133–136.
- Hill R. The elastic behaviour of a crystalline aggregate. *Proceedings of the Physical Society* 1952; A65:349–354.
- Holland JH. *Adaptation in Natural and Artificial Systems*. Ann Arbor (MI): University of Michigan Press; 1975.
- Jikov VV, Kozlov SM, Olenik OA. *Homogenization of Differential Operators and Integral Functionals*. Springer-Verlag; 1994.
- Maxwell JC. On the dynamical theory of gases. *Philosophical Transactions of the Royal Society of London* 1867;157:49.
- Maxwell JC. *A Treatise on Electricity and Magnetism*. 3rd ed. Oxford: Clarendon Press; 1873.
- Nemat-Nasser, S, Hori M. *Micromechanics: Overall Properties of Heterogeneous Solids*. 2nd ed. Amsterdam: Elsevier; 1999.
- Rayleigh JW. On the influence of obstacles arranged in rectangular order upon properties of a medium. *Philosophical Magazine* 1892;32:481–491.
- Reuss A. Berechnung der Fließgrenze von Mischkristallen auf Grund der Plastizitätsbedingung für Einkristalle. *Zeitschrift für Angewandte Mathematik und Mechanik* 1929;9:49–58.
- Torquato S. *Random Heterogeneous Materials: Microstructure and Macroscopic Properties*. New York: Springer-Verlag; 2001.
- Voigt W. Über die Beziehung zwischen den beiden Elastizitätskonstanten isotroper Körper. *Wiedemanns Annalen* 1889;38:573–587.
- Wiener O. Zur Theorie der Refraktionskonstanten. *Berichte über die Verhandlungen der Königlich-Sächsischen Gesellschaft der Wissenschaften zu Leipzig. Mathematisch-Physische Klasse* 1910;62:256–277.

- Zohdi TI. *Electromagnetic properties of multiphase dielectrics. A primer on modeling, theory and computation*. Springer-Verlag; 2012.
- Zohdi TI. Computational optimization of vortex manufacturing of advanced materials. *Computer Methods in Applied Mechanics and Engineering* 2001;190(46–47):6231–6256.
- Zohdi TI. Genetic optimization of statistically uncertain microheterogeneous solids. *Philosophical Transactions of the Royal Society: Mathematical, Physical & Engineering Sciences* 2003;361 (1806):1021–1043.
- Zohdi ME. *Statistical Quality Control. Mechanical Engineers' Handbook; manufacturing and management*. 3rd ed. Kutz M, editor. Wiley Publishers; 2006.
- Zohdi TI. Simulation of coupled microscale multiphysical-fields in particulate-doped dielectrics with staggered adaptive FDTD. *Computer Methods in Applied Mechanics and Engineering* 2010;199:79–101.
- Zohdi TI, Wriggers P. *Introduction to Computational Micromechanics*. Second Reprinting. Springer-Verlag; 2008.

PART V

INSTRUMENTATION

INPUT AND OUTPUT CHARACTERISTICS

ADAM C. BELL

- 47.1 Introduction
- 47.2 Familiar examples of input–output interactions
 - 47.2.1 Power exchange
 - 47.2.2 Energy exchange
 - 47.2.3 A human example
- 47.3 Energy, power, impedance
 - 47.3.1 Definitions and analogies
 - 47.3.2 Impedance and admittance
 - 47.3.3 Combining impedances and/or admittances
 - 47.3.4 Computing impedance or admittance at an input or output
 - 47.3.5 Transforming or gyrating impedances
 - 47.3.6 Source equivalents: Thévenin and Norton
- 47.4 Operating point of static systems
 - 47.4.1 Exchange of real power
 - 47.4.2 Operating points in an exchange of power or energy
 - 47.4.3 Input and output impedance at the operating point
 - 47.4.4 Operating point and load for maximum transfer of power
 - 47.4.5 An unstable energy exchange: tension-testing machine
 - 47.4.6 Fatigue in bolted assemblies
 - 47.4.7 Operating point for nonlinear characteristics
 - 47.4.8 Graphical determination of output impedance for nonlinear systems
- 47.5 Transforming the operating point
 - 47.5.1 Transducer-matched impedances
 - 47.5.2 Impedance requirements for mixed systems
- 47.6 Measurement systems
 - 47.6.1 Interaction in instrument systems
 - 47.6.2 Dynamic interactions in instrument systems
 - 47.6.3 Null instruments
- 47.7 Distributed systems in brief
 - 47.7.1 Impedance of a distributed system
- 47.8 Concluding remarks
- References

47.1 INTRODUCTION

Everyone is familiar with the interaction of devices connected to form a system, although they may not think of their observations in those terms. Familiar examples include the following:

1. dimming of the headlights while starting a car,
2. slowdown of an electric mixer lowered into heavy batter,
3. freezing a showerer by starting the dishwasher,
4. speedup of a vacuum cleaner when the hose plugs,
5. two-minute wait for a fever thermometer to rise,
6. special connectors required for TV antennas,
7. speedup of a fan in the window with the wind against it,
8. shifting of an automatic transmission on a hill.

These effects happen because one part of a system loads another. Most mechanical engineers would guess that weighing an automobile by placing a bathroom-type scale under its wheels one at a time and summing the four measurements will yield a higher result than would be obtained if the scale was flush with the floor. Most electrical engineers understand that loading a potentiometer's wiper with too low a resistance makes its dial nonlinear for voltage division. Instrumentation engineers know that a heavy accelerometer mounted on a thin panel will not measure the true natural frequencies of the panel. Audiophiles are aware that loudspeaker impedances must be matched to amplifier impedance. We have all seen the 75- and 300- Ω markings under the antenna connections on TV sets, and most cable subscribers have seen balun transformers for connecting a coaxial cable to the flat-lead terminals of an older TV.

Every one of these examples involves a desired or an undesirable interaction between a source and a receiver of energy. In every case, there are properties of the source part and the load part of the system that determine the efficiency of the interaction. This chapter deals exclusively with interactions between static and dynamic subsystems intended to function together in a task and with how best to configure and characterize those subsystems.

Consider the analysis of dynamic systems. To create mathematical models of these systems requires that we idealize our view of the physical world. First, the system must be identified and separated from its environment. The environment of a system is the universe outside the free body, control volume, or isolated circuit. The combination of these, which is the system under study and the external sources, provides or removes energy from the system in a known way. Next, in the system itself, we must arrange a restricted set of ideal elements connected in a way that will correctly represent the energy storages and dissipations of the physical system while, at the same time, we need the mathematical handles that explore the system's behavior in its environment. The external environment of the system being modeled must then itself be modeled and connected and is usually represented by special ideal elements called sources.

We expect, as a result of these sources, that the system under study will not alter the important variables in its environment. The water rushing from a kitchen faucet will not normally alter the atmospheric pressure; our electric circuit will not measurably slow the turbines in the local power plant; the penstock will not draw down the level of the reservoir (in a time frame consistent with a study of penstock dynamics, anyway); the

cantilever beam will not distort the wall it is built into; and so on. In this last instance, for example, the wall is a special source of zero displacement and zero rotation no matter what forces and moments are applied.

In this chapter, we consider, instead of the behavior of a single system in a known environment, the interaction between pairs of connected dynamic systems at their interface, often called the driving point. The fundamental currency is, as always, the energy or power exchanged through the interface. In an instrumentation or a control system, the objective of the energy exchange might be information transmission, but this is not considered here (we would like information exchanges to take place at the lowest possible energy costs, but the second law of thermodynamics rules out a free transmission).

As always, energy factors into two variables, such as voltage and current in electrical systems, and we are concerned with the behavior of these in the energetic interaction. The major difference in this perspective is that the system supplying energy cannot do so at a fixed value. Neither the source nor the system receiving energy can fix its values for a changing demand without a change in the value of a supply variable. The two subsystems are in an equilibrium with each other and are forced by their connection to have the same value of both of the appropriate energy variables. We concern ourselves with determining and controlling the value of these energy variables at the interface where, obviously, only one is determined by each of the interacting systems.

47.2 FAMILIAR EXAMPLES¹ OF INPUT–OUTPUT INTERACTIONS

47.2.1 Power Exchange

In the real world, pure sources and sinks are difficult to find. They are idealized, convenient constructs or approximations that give our system analyses independent forcing functions. We commonly think of an automobile storage battery as a source of 12.6 V independent of the needed current, and yet we have all observed dimming headlights while starting an engine. Clearly, the voltage of this battery is a function of the current demanded by its load. Similarly, we cannot charge the battery unless our alternator provides more than 12.6 V, and the charging rate depends on the overvoltage supplied. Thus, when the current demanded or supplied to a battery approaches its limits, we must consider that the battery really looks like an ideal 12.6-V source in series with a small resistance. The voltage at the battery terminals is a function of the current demanded and is not independent of the system loading or charging it in the interaction. This small internal resistance is termed the output impedance (or input impedance or driving-point impedance) of the battery.

If we measure the voltage on this battery with a voltmeter, we should draw so little current that the voltage we see is truly the source voltage without any loss in the internal resistance. The power delivered from the battery to the voltmeter is negligible (but not zero) because the current is so small. Alternately, if we do a short-circuit test of the battery, its terminal voltage should fall to zero while we measure the very large current that results. Again, the power delivered to the ammeter is negligible because, although the current is very large, the voltage is vanishingly small.

¹Many of the examples in this chapter are drawn from Chapter 6 of a manuscript of unpublished notes, “Dynamic Systems and Measurements,” by C. L. Nachtigal, used in the School of Mechanical Engineering, Purdue University, 1978.

At these two extremes, the power delivered is essentially zero, so clearly at some intermediate load the power delivered will be a maximum. We will show later that this occurs when the load resistance is equal to the internal resistance of the battery (a point at which batteries are usually not designed to operate). The discussion above illustrates a simple concept: impedances should be matched to maximize power or energy transfer but should be maximally mismatched for making a measurement without loading the system in which the measurement is to be made. We will return to the details of this statement later.

47.2.2 Energy Exchange

Interactions between systems are not restricted to resistive behavior, nor is the concept of impedance matching restricted to real, as opposed to reactive, impedances. Consider a pair of billiard balls on a frictionless table (to avoid the complexities of spin), and consider that their impact is governed by a coefficient of restitution, ϵ . Before impact, only one ball is moving, but afterward both may be. The initial and final energies are as follows:

$$\begin{aligned}\text{Initial energy} &= \frac{1}{2}M_1v_{1i}^2 \\ \text{Final energy} &= \frac{1}{2}M_1v_{1f}^2 + \frac{1}{2}M_2v_{2f}^2\end{aligned}\quad (47.1)$$

where the subscript 1 refers to the striker and 2 to the struck ball, M is mass, v is velocity, and the subscripts i and f refer to initial and final conditions, respectively.

Because no external forces act on this system of two balls during their interaction, the total momentum of the system is conserved:

$$M_1v_{1i} + M_2v_{2i} = M_1v_{1f} + M_2v_{2f} \quad (47.2)$$

or

$$v_{1i} + \mathbf{m}v_{2i} = v_{1f} + \mathbf{m}v_{2f},$$

where $\mathbf{m} = M_2/M_1$. The second equation, required to solve for the final velocities, derives from impulse and momentum considerations for the balls considered one at a time. Because no external forces act on either ball during their interaction except those exerted by the other ball, the impulses,² or integrals of the force acting over the time of interaction on the two, are equal. (See “impact” in virtually any dynamics text.) From this, it can be shown that the initial and final velocities must be related:

$$\epsilon(v_{1i} - v_{2i}) = (v_{1f} - v_{2f}), \quad (47.3)$$

where $v_{2i} = 0$ in this case and the coefficient of restitution ϵ is a number between 0 and 1. A 0 corresponds to a plastic impact while a 1 corresponds to a perfectly elastic impact. Equations (47.2) and (47.3) can be solved for the final velocities of the two balls:

$$v_{1f} = \frac{1 - \mathbf{m}\epsilon}{1 + \mathbf{m}}v_{1i} \quad \text{and} \quad v_{2f} = \frac{1 + \epsilon}{1 + \mathbf{m}}v_{1i}. \quad (47.4)$$

² Impulse = $\int_{t=0}^t \text{Force } dt$, where Force is the vector sum of all the forces acting over the period of interaction, t .

Now assume that one ball strikes the other squarely³ and that the coefficient of restitution ϵ is unity (perfectly elastic impact). Consider the following three cases:

1. The two balls have equal mass, so $\mathbf{m} = 1$, and $\epsilon = 1$. Then the striking ball, M_1 , will stop, and the struck ball, M_2 , will move away from the impact with exactly the initial velocity of the striking ball. All the initial energy is transferred.
2. The struck ball is more massive than the striking ball, $\mathbf{m} > 1$, $\epsilon = 1$. Then, the striker will rebound along its initial path, and the struck ball will move away with less than the initial velocity of the striker. The initial energy is shared between the balls.
3. The striker is the more massive of the two, $\mathbf{m} < 1$, $\epsilon = 1$. Then the striker, M_1 , will follow at reduced velocity behind the struck ball after their impact, and the struck ball will move away faster than the initial velocity of the striker (because it has less mass). Again, the initial energy is shared between the balls.

Thus, the initial energy is conserved in all of these transactions. But, the energy can be transferred completely from one ball to the other *if and only if* the two balls have the same mass.

If these balls were made of clay so that the impact was perfectly plastic (no rebound whatsoever), then $\epsilon = 0$, so the striker and struck balls would move off together at the same velocity after impact no matter what the masses of the two balls. They would be effectively stuck together. The final momentum of the pair would equal the initial momentum of the striker because, on a frictionless surface, there are no external forces acting, but energy could not be conserved because of the losses in plastic deformation during the impact. The final velocities for the same three cases are

$$v_f = \frac{1}{1 + \mathbf{m}} v_i. \quad (47.5)$$

Since the task at hand, however, is to transfer kinetic (KE) from the first ball to the second, we are interested in maximizing the energy in the second ball after impact with respect to the energy in the first ball before impact:

$$\frac{\text{KE}_{(M_2, \text{after})}}{\text{KE}_{(M_1, \text{before})}} = \frac{\frac{1}{2} M_2 (v_{2f})^2}{\frac{1}{2} M_1 (v_{1i})^2} = \frac{M_2 (1/(1 + \mathbf{m}))^2 (v_{1i})^2}{M_1 (v_{1i})^2} = \frac{\mathbf{m}}{(1 + \mathbf{m})^2}. \quad (47.6)$$

This takes on a maximum value of $1/4$ when $\mathbf{m} = 1$ and falls off rapidly as \mathbf{m} departs from 1.

Thus, after the impact of two clay balls of equal mass, one-fourth of the initial energy remains in the striker, one-fourth is transferred to the struck ball, and one-half of the initial energy of the striker is lost in the impact. If the struck ball is either larger or smaller than the striker, however, then a greater fraction of the initial energy is dissipated in the impact and a smaller fraction is transferred to the second ball. The reader should reflect on how this influences the severity of automobile accidents between vehicles of different sizes.

47.2.3 A Human Example

Those in good health can try the following experiment. Run up a long flight of stairs one at a time and record the elapsed time. After a rest, try again, but run the stairs two at a

³ Referred to in dynamics as *direct central impact*.

time. Still later, try a third time, but run three steps at a time. Most runners will find that their best time is recorded for two steps at a time.

In the first test, the runner's legs are velocity limited: too much work is expended by simply moving legs and feet, and the forces required are too low to use the full power of the legs effectively. In the third test, although the runner's legs do not have to move very quickly, they are on the upper edge of their force capabilities for continued three-step jumps; the forces required are too high and the runner could, at lower forces, move his or her legs much faster. In the intermediate case, there is a match between the task and the force-velocity characteristics of the runner's legs.

Bicycle riders assure this match with a variable-speed transmission that they adjust so they can crank the pedals at approximately 60 RPM. We will later look at other means of ensuring the match between source capabilities and load requirements when neither of them is changeable, but the answer is always a transformer or gyrator of some type (a gear ratio in this case).

47.3 ENERGY, POWER, IMPEDANCE

47.3.1 Definitions and Analogies

Energy is the fundamental currency in the interactions between elements of a physical system no matter how the elements are defined. In engineering systems, it is convenient to describe these transactions in terms of a complementary pair of variables whose product is the power or flow rate of the energy in the transaction. These product pairs are familiar to all engineers: voltage \times current = power, force \times displacement = energy, torque \times angular velocity = power, pressure \times flow = power, and pressure \times time rate of change of volume exchanged = power. Some are less familiar: flux linkage \times current = energy, charge \times voltage = energy, and absolute temperature \times entropy flux = thermal power. Paynter's (1960) tetrahedron of state shows how these are related (Figure 47.1). Typically, one of these factors is extensive, a flux or flow, such as current, velocity, volume flow rate, or angular velocity. The other is intensive, a potential or effort,⁴ such as voltage, force, pressure, or torque. Thus, $\mathcal{P} = \text{extensive} \times \text{intensive}$ for any of these domains of physical activity.

This factoring is quite independent of the analogies between the factors of power in different domains, for which any arbitrary selection is acceptable. In essence, velocity is

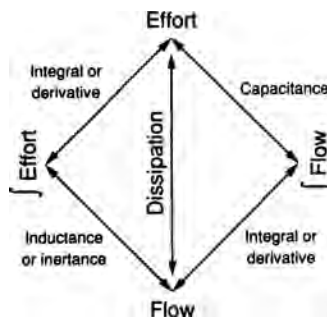


FIGURE 47.1 H. M. Paynter's tetrahedron of state.

⁴This is Paynter's terminology, used with reference to his "Bond Graphs."

not like voltage or force-like current, just as velocity is not like current or force-like voltage. It is convenient, however, before defining impedance and working with it to choose an analogy so that generalizations can be made across the domains of engineering activity. There are two standard ways to do this: the Firestone analogy (Firestone, 1932/1933) and the mobility analogy. Electrical engineers are most familiar with the Firestone analogy, whereas mechanical engineers are probably more comfortable with the mobility analogy. The results derived in this chapter are independent of the analogy chosen. To avoid confusion, both will be introduced, but only the mobility analogy will be used in this chapter.

The Firestone analogy gives circuit-like properties to mechanical systems: all systems consist of nodes like a circuit and only of lumped elements considered to be two-terminal or four-terminal device. For masses and tanks of liquid, one of the terminals must be understood to be ground, the inertial reference frame, or atmosphere. Then, one of the energy variables is measured across the terminals of the element and the other passes through the element. In a circuit, voltage is across and current passes through. For a spring, however, velocity difference is across and the force passes through. Thus, this analogy linked voltage to velocity, angular velocity, and pressure as across variables and linked current to force, torque, and flow rate as through variables. Clearly, across \times through = power.

The mobility analogy, in contrast, considers the complementary power variables to consist of a potential and a flux, an intrinsic and extrinsic variable. The potentials, or efforts, are voltage, force, torque, and pressure, while the fluxes, or flows, are current, velocity, angular velocity, and fluid flow rate.

47.3.2 Impedance and Admittance

Impedance, in the most general sense, is the relationship between the factors of power. Because only the constitutive relationships for the dissipative elements are expressed directly in terms of the power variables, $\Delta V_R = R \times i_R$, for example, while the equations for the energy storage elements are expressed in terms of the derivative of one of the power variables⁵ with respect to the other, $i_C = C(dV_C/dt)$ for example, these are most conveniently expressed in Laplace transform terms. Impedances are really self-transfer functions at a point in a system. Because the concept was probably defined first for electrical systems, that definition is most standardized: electrical impedance $Z_{\text{electrical}}$ is defined as the rate of change of voltage with current:

$$Z_{\text{electrical}} = \frac{d(\text{voltage})}{d(\text{current})} = \frac{d(\text{effort})}{d(\text{flow})}. \quad (47.7)$$

By analogy, impedance can be similarly defined for the other engineering domains:

$$Z_{\text{translation}} = \frac{d(\text{force})}{d(\text{velocity})}. \quad (47.8)$$

$$Z_{\text{rotation}} = \frac{d(\text{torque})}{d(\text{angular velocity})}. \quad (47.9)$$

$$Z_{\text{fluid}} = \frac{d(\text{pressure})}{d(\text{flow rate})}. \quad (47.10)$$

⁵ See Figure 47.1 again. Capacitance is a relationship between the integral of the flow and the effort, which is the same as saying that capacitance relates the flow to the derivative of the effort.

TABLE 47.1 Impedances of Lumped Linear Elements

Domain	Kinetic Storage	Dissipation	Potential Storage
Translational	Mass: M	Damping: b	Spring: k/s
Rotational	Inertia: J	Damping: B	Torsion spring: k_f/s
Electrical	Inductance: L	Resistance: R	Capacitance: $1/Cs$
Fluid	Inertance: I	Fluid resistance: R	Fluid capacitance: $1/Cs$

Table 47.1 is an impedance table using these definitions of the fundamental lumped linear elements. Note that these are derived from the Laplace transforms of the constitutive equations for these elements; they are the transfer functions of the elements and are expressed in terms of the Laplace operator s . The familiar $F = M \cdot a$, for example, becomes, in power-variable terms, $F = M(dv/dt)$; it transforms as $F(s) = Ms v(s)$, so

$$(Z_{\text{translation}})_{\text{mass}} = \frac{dF_{\text{mass}}}{dv_{\text{mass}}} = Ms. \quad (47.11)$$

Because these involve the Laplace operator s , they can be manipulated algebraically to derive combined impedances. The reciprocal of the impedance, the admittance, is also useful. Formally, admittance is defined as

$$\text{Admittance: } Y = \frac{1}{Z} = \frac{d(\text{flow})}{d(\text{effort})}. \quad (47.12)$$

47.3.3 Combining Impedances and/or Admittances

Elements in series are those for which the flow variable is common to both elements and the efforts sum. Elements in parallel are those for which the effort variable is common to both elements and the flows sum. By analogy to electrical resistors, we can deduce that the impedance sum for series elements and the admittance sum for parallel elements form the combined impedance or admittance of the elements:

Impedances in series:

$$\begin{aligned} Z_1 + Z_2 &= Z_{\text{total}} \quad (\text{common flow}) \\ \frac{1}{Y_1} + \frac{1}{Y_2} &= \frac{1}{Y_{\text{total}}} \end{aligned} \quad (47.13)$$

Impedances in parallel:

$$\begin{aligned} Y_1 + Y_2 &= Y_{\text{total}} \quad (\text{common effort}) \\ \frac{1}{Z_1} + \frac{1}{Z_2} &= \frac{1}{Z_{\text{total}}} \end{aligned} \quad (47.14)$$

When applying these relationships to electrical or fluid elements, there is rarely any confusion about what constitutes series and parallel. In the mobility analogy, however, a pair of springs connected end to end are in parallel because they experience a common force, regardless of the topological appearance, whereas springs connected side by side are in

series because they experience a common velocity difference.⁶ For a pair of springs end to end, the total admittance is

$$\frac{s}{k_{\text{total}}} = \frac{s}{k_1} + \frac{s}{k_2}$$

so the impedance is

$$\frac{k_{\text{total}}}{s} = \frac{k_1 k_2}{s(k_1 + k_2)}.$$

For the same springs side by side, the total impedance is

$$\frac{k_{\text{total}}}{s} = \frac{k_1 + k_2}{s}.$$

47.3.4 Computing Impedance or Admittance at an Input or Output

There are basically two ways in which an input or an output admittance can be computed. The first, and most direct, is to compute the transfer function between the effort and the flow at the driving point and take the derivative with respect to the flow. For a mechanical rotational system, for example, torque as a function of angular velocity is expressed and differentiated with respect to angular velocity. This method must be used if the system being considered is nonlinear because the derivative must be taken at an operating point. If the system is linear, then the ratio of flow or effort will suffice; in the rotational system, the impedance is simply the ratio (torque/speed).

The second method takes the impedances of the elements one at a time and combines them. This approach is particularly useful for linear (or linearized) systems. The question then arises of determining the impedance of any sources in the subsystem being considered. Flow sources, such as current sources, velocity sources, angular velocity sources, and fluid flow sources, all have the relationship flow = constant. Their impedance is therefore infinite ($Z_{\text{flow source}} = \infty$) because any change in effort results in zero change in flow. Effort sources, such as voltage sources, force sources, torque sources, and pressure sources, will provide any flow to maintain the effort required; the change in effort for a change in flow remains zero, so their impedance is $Z_{\text{effort source}} = 0$. An effort source therefore represents a short circuit from an impedance point of view; it connects together two nodes that were separate. A flow source represents a null element; since its impedance is infinite, it represents an open circuit. Flow sources are simply removed in impedance calculations.

An example will distinguish between these two approaches. Figure 47.2 shows, on the left, a simple circuit disconnected at a driving point from its load. The load is of no consequence in this calculation; we simply require the driving-point impedance of the circuit. In the first approach, we derive the voltage at the driving point as a function of the current leaving those terminals and the source, V_s and I_s to obtain the relationship

$$V_o = \frac{R_3}{R_1 + R_3} V_s + \left(\frac{R_1 R_2 + R_2 R_3 + R_1 R_3}{R_1 + R_3} \right) (I_s - i_o). \quad (47.15)$$

⁶For many, the appeal of the Firestone analogy is that springs are equivalent to inductors, and there can be no ambiguity about series and parallel connections. End to end is series.

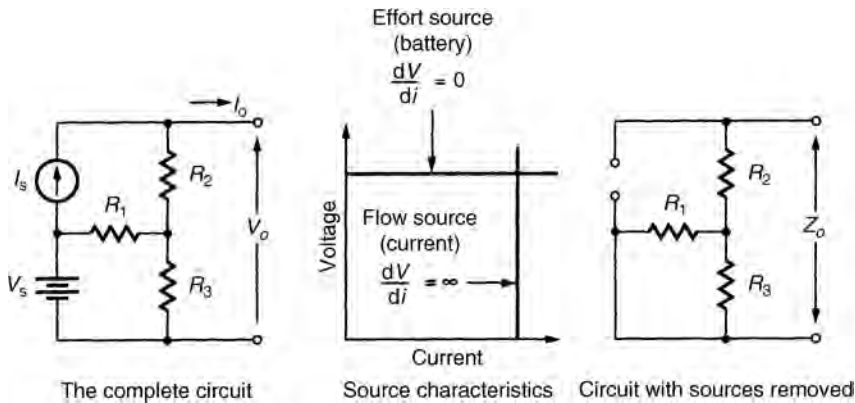


FIGURE 47.2 A simple circuit as a source.

Clearly, the negative derivative of the voltage (V_o) with respect to the current (i_o) is given by

$$-\frac{dV_o}{di_o} = Z_o = \frac{R_1 R_2 + R_2 R_3 + R_1 R_3}{R_1 + R_3}. \quad (47.16)$$

In the second method, the voltage source (V_s) can be set to zero, a short circuit, without affecting the impedances, and the current source (I_s) can be removed to yield the circuit on the right in Figure 47.2. Then, the impedances need only be combined as R_2 in series with the parallel pair R_1 and R_3 . Thus,

$$Z_o = R_2 + \frac{R_1 R_3}{R_1 + R_3} = \frac{R_1 R_2 + R_2 R_3 + R_1 R_3}{R_1 + R_3} \quad \text{as before.} \quad (47.17)$$



47.3.5 Transforming or Gyrating Impedances

Ideal transformers, transducers, and gyrators play an important part in dynamic systems and an equally important part in obtaining optimal performance from source–load combinations. They share several vital features: all are two-port (or four-terminal) devices, all are energetically conservative, and all are considered lumped elements. Each of the many types requires two equations for its description, always of the same form. Table 47.2 lists many of the more common linear two-ports, and Figure 47.3 illustrates them.

All of these transducing devices alter the effort–flow relationships of elements connected at their far end. Consider Figure 47.4, which shows a simple model of a front wheel of an automobile suspension. Because the spring and damper are mounted inboard of the wheel, their effectiveness is reduced by the mechanical disadvantage of the suspension arm. What is the impedance of the spring and shock absorber at $(\cdot)_1$ as viewed from the wheel at $(\cdot)_2$?

Because the spring and shock absorber share a common velocity (both ends share nodes or points of common velocity), their impedances add to give the impedance of the pair at their point of attachment, $(\cdot)_1$, to the lever:

TABLE 47.2 Ideal Linear Lumped Two-Ports

Domain to Domain	Name	Figure Numbers	Governing Equations (Left to Right in Figures)
Translation–translation	Lever, L = distance to function	3a	$F_1 = \frac{L_2}{L_1} F_2$ $v_1 = \frac{L_1}{L_2} v_2$
Rotation—rotation	Gears, N = number of teeth	3b	$\tau_1 = \frac{N_1}{N_2} \tau_2$ $\omega_1 = \frac{N_2}{N_1} \omega_2$
Rotation–translation	Crank ($\theta \ll 1$) or rack and pinion	3c 3d	$\begin{cases} \tau = RF \\ \omega = \frac{1}{R} v \end{cases}$
	 Screw (p = length/radius)	 3e	$\begin{cases} \tau = pF \\ \omega = \frac{1}{p} v \end{cases}$
Electrical–electrical	Transformer	3f	$V_1 = \frac{N_1}{N_2} V_2$ $i_1 = \frac{N_2}{N_1} i_2$
Electrical–rotation	Permanent magnet dc motor	3g	$V_a = K_b \omega$ $i_a = \frac{1}{K_m} \tau$
Electrical–translation	Voice coil	3h	$V_a = K_c v$ $i_a = \frac{1}{K_c} F$
Fluid–translation	Piston area = A	3j	$P = \frac{1}{A} F$ $Q = Av$
Rotation–fluid	Fixed-displacement pump-motor, displacement/radius = D	3k	$\tau = Dp$ $\omega = \frac{1}{D} Q$

$$(Z_1)_{\text{total}} = Z_{\text{spring}} + Z_{\text{damper}} = b + \frac{k}{s} = \frac{bs + k}{s}. \quad (47.18)$$

At the $(\cdot)_2$ end of the lever, the force, from Table 47.2, is $F_2 = (L_1/L_2)F_1$, but from the definition of Z for linear elements, $F_1 = (Z_1)_{\text{total}} v_1$. So we get the following:

$$F_2 = \frac{L_1}{L_2} \left(\frac{bs + k}{s} \right) v_1. \quad (47.19)$$

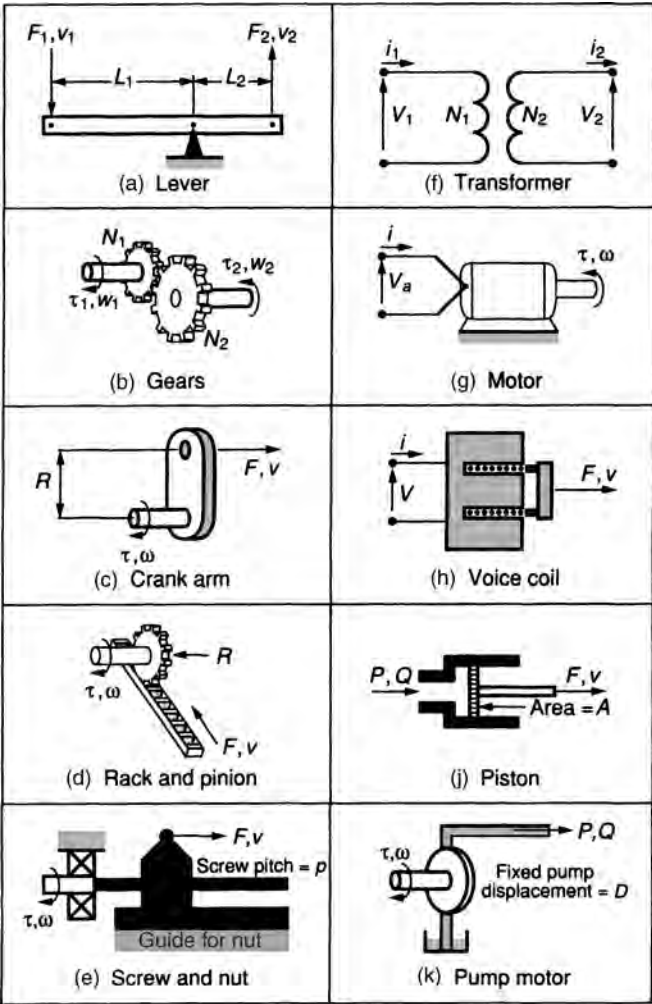


FIGURE 47.3 The ideal linear lumped two-ports: transformers and gyrators.

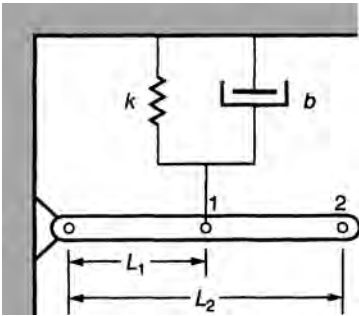


FIGURE 47.4 An abstraction of an automobile suspension linkage.

The second equation for the lever is $v_1 = (L_1/L_2)v_2$, and substituting this into Equation (47.19) yields

$$F_2 = \frac{L_1}{L_2} \left(\frac{bs + k}{s} \right) \left(\frac{L_1}{L_2} \right) v_2 = \left(\frac{L_1}{L_2} \right)^2 \left(\frac{bs + k}{s} \right) v_2 \quad (47.20)$$

$$(Z_2)_{\text{total}} = \left(\frac{L_1}{L_2} \right)^2 (Z_1)_{\text{total}}. \quad (47.21)$$

Thus, the impedance of the suspension, observed from the wheel end of the lever arm, is multiplied by the square of the lever ratio. This general result applies to all transduction elements.

47.3.6 Source Equivalents: Thévenin and Norton

Thévenin's and Norton's theorems were both originally developed for electric circuits. They are duals; that is, while Thévenin uses a series voltage source, Norton uses a parallel current source to construct an equivalent to a real subsystem being considered as a source. Both of these theorems are completely generalizable to any system, provided that the appropriate sources are used. In this chapter, the mobility analogy has been adapted so that Thévenin equivalents are constructed using sources of voltage, force, pressure, or torque, and Norton equivalents are constructed using sources of current, velocity, flow, or angular velocity. In the development below, these two classes of sources will be referred to as effort and flow sources.

47.3.6.1 Thévenin Equivalent Assume that a subsystem being considered as a source contains within its structure one or more ideal sources. Insofar as can be measured externally at the driving point (the point of connection between any two systems considered to be source and load), any active subsystem, no matter what its load, can be replaced by a new effort source added in series with the original subsystem; all the original internal sources are set to zero. The value of the new source effort variable is the output that would appear at the driving point if the load were disconnected from the original subsystem. Setting an effort source to zero is equivalent to connecting its nodes together; setting a flow source to zero is equivalent to removing it from the system.

The output impedance of this Thévenin equivalent is clearly the impedance looking in at the driving point; it is the derivative of the driving-point effort variable with respect to the driving-point flow variable with the load disconnected. Thévenin's theorem simplifies this calculation because it tells us that the internal sources can be set to zero before the calculation is made, and the system topology is often substantially simplified by this move. We have already seen an example of this [see Equations (47.15)–(47.17)]. The Thévenin equivalent of this circuit is simply a new source determined with the current $i_o = 0$ (see Figure 47.2):

$$V_{\text{Thévenin}} = \frac{R_3}{R_1 + R_3} V_s + \left(\frac{R_1 R_2 + R_2 R_3 + R_1 R_3}{R_1 + R_3} \right) I_s \quad (47.22)$$

in series with a resistance determined with sources V_s and I_s zeroed:

$$R_{\text{Thévenin}} = \frac{R_1 R_2 + R_2 R_3 + R_1 R_3}{R_1 + R_3}. \quad (47.23)$$

47.3.6.2 Norton Equivalent Assume the same subsystem considered above. Insofar as can be measured externally at the driving point, any active subsystem, no matter what its load, can be replaced by a new flow source added in parallel with the original subsystem, all the original sources set to zero. The value of the new source flow variable is the flow that would pass through a short circuit substituted at the driving point for the original load.

Referring again to Equation (47.15), with V_o set to zero, we obtain the following:

$$0 = \frac{R_3}{R_1 + R_3} V_s + \left(\frac{R_1 R_2 + R_2 R_3 + R_1 R_3}{R_1 + R_3} \right) (I_s - i_o). \quad (47.24)$$

We see that i_o for a short circuit would be

$$i_o = I_s + \left(\frac{R_3}{R_1 R_2 + R_2 R_3 + R_1 R_3} \right) V_s \quad (47.25)$$

and, as for the Thévenin equivalent, the circuit impedance is

$$R_{\text{Norton}} = \frac{R_1 R_2 + R_2 R_3 + R_1 R_3}{R_1 + R_3}. \quad (47.26)$$

Note that $R_{\text{Thévenin}} = R_{\text{Norton}}$, always.

47.4 OPERATING POINT OF STATIC SYSTEMS

A static system is a system without energy storage, a system in which there are only sources, transducers, and dissipation elements. Such systems have no transient response: they respond instantly to their inputs algebraically. The relationships among any of their variables are proportionalities—simple static gains. If the inputs to a stable dynamic system are held constant for long enough, it will become stationary; its variables will not change with time provided only that there is sufficient dissipation in the system to damp out any oscillations. Such a system is not static; it is at steady state. There is no exchange of energy among its energy storage elements, and the dissipative elements completely determine the state of the system.

47.4.1 Exchange of Real Power

If one system is supplying real power to another system in steady-state operation, then for the purposes of a static analysis, energy storage elements can be ignored. Both the source and the load can be considered to be purely resistive. If the source is separated from the load at the point of interest (at least conceptually), then the characteristics of source and load can be measured or computed. The load will be a power absorber—an electrical fluid resistance or a mechanical damper—and its characteristics can be represented as a line in

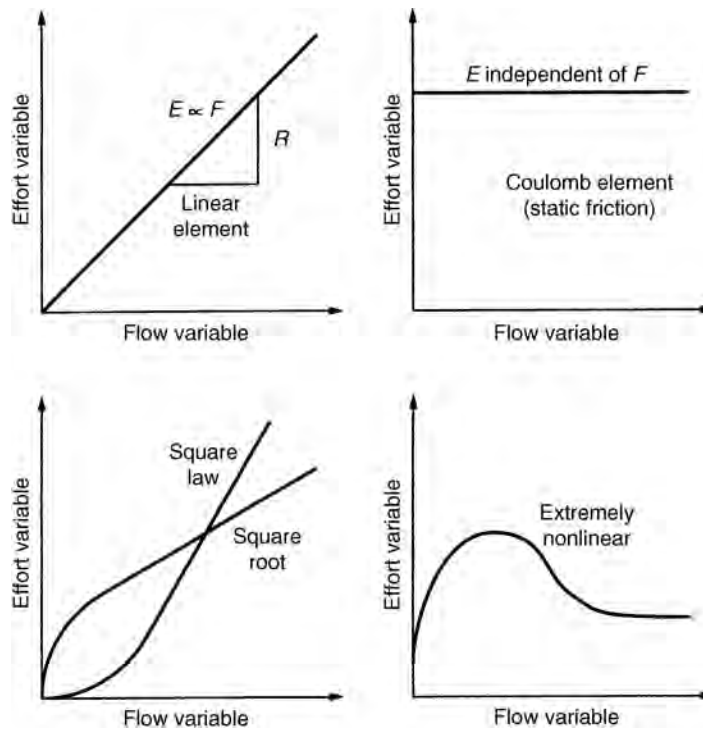


FIGURE 47.5 Common dissipative load characteristics.

a power plane coordinate system: voltage versus current, force versus velocity, torque versus angular velocity, or pressure versus flow. There is no requirement that this line be straight, and except that the measurement might be more difficult, there is no necessity that this line be single valued or that it start at the origin. Figure 47.5 shows a selection of common dissipative load characteristics.

Similarly, when the source portion of the system is loaded with a variable dissipation, the line representing its output characteristics can be plotted on the same coordinates as the load. Such characteristics are often given for pumps, servomotors, transistors, hydraulic valves, electrical supplies, and fans. Again, there is no requirement that this line be straight or that it be single valued, but it is very unusual for it to pass through the origin. If the source or the load is not constant in time or can be controlled, then a family of these characteristics will be required for variations in the parameters of the source or load. Figure 47.6 shows a selection of these sources. With very few exceptions, real sources cannot operate at the maximum values of their power variables simultaneously: a battery cannot deliver its maximum voltage and current at the same time. In more general terms, this means that in spite of local variations, real source characteristics tend to droop from left to right in the power plane; their average slope is negative.

47.4.2 Operating Points in an Exchange of Power or Energy

When resistive source and load characteristics are plotted on the same coordinates, they intersect at least once. The coordinate values at that point, or at those points if there are several, are the values of the power variables at which that combination of source and load

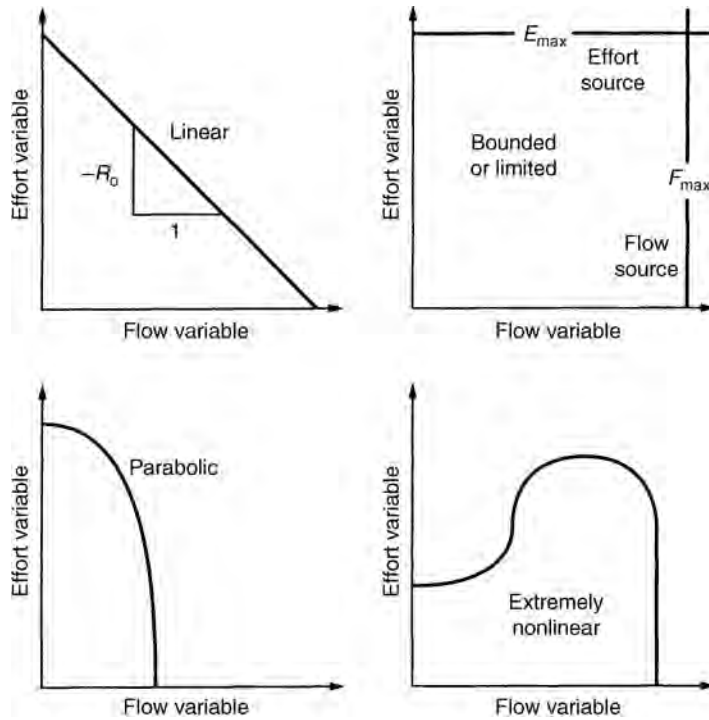


FIGURE 47.6 Load source variations.

must operate if they are connected. This is called an operating point. From a computational point of view, the source causes one of the power variables given the other and the load causes the other given the source. They must operate at the same point in the power plane to satisfy continuity (common flows) and compatibility (common efforts) conditions.

When there are multiple intersections, all are possible operating points, but not all will be stable operating points; for example, any disturbance from equilibrium might result in a transition to another operating point. The condition for a stable intersection is best seen graphically in Figure 47.7. For a stable intersection, as shown on the left, it is required

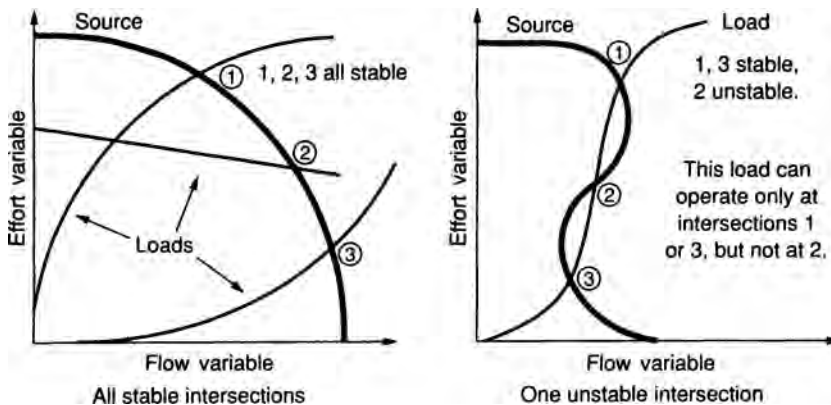


FIGURE 47.7 Stable and unstable operating point intersections.

that a small perturbation of the load, which increases its demand for power, be countered by a shortage of power from the supply side of the system and a small perturbation of the load, which decreases its demand for power, be met with an excess from the source. In either case, the load will be driven back to the intersection by the excess or deficit in the source capability. A reversal of these conditions is an unstable operating point because disturbances will be driven further in the direction of their initial departure.

At the unstable intersection in Figure 47.7 (2, on the right-hand side), a slight increase in the flow demand of the load will result in an overwhelming increase in the supply flow available to drive the load, which will then cause a traversal to point 1 in the figure. Similarly, if the effort decreases slightly from operating point 2, then the source will be starved compared to the demand of the load at that effort level, so the effort will fall until point 3 is reached.

47.4.3 Input and Output Impedance at the Operating Point

Lines of constant power are hyperbolas in the power plane, with increasing values of the power at increasing distance from the origin. The usual sign conventions imply that sources deliver power in the first and third quadrants while loads absorb power in those quadrants. Conversely, sources absorb power in the second and fourth quadrants and loads return it. The output impedance of a source is defined as minus the slope of the output characteristic. For nonlinear characteristics, the output impedance at any point is defined as minus the slope at that point. For loads, the input impedance is defined as the slope of the load line, but for nonlinear characteristics, there are two possibilities of importance: the slope of the line at a point (the incremental or local input impedance) and the slope of the chord to the point from the origin (the chordal impedance). Figure 47.8 summarizes these features of the power plane.

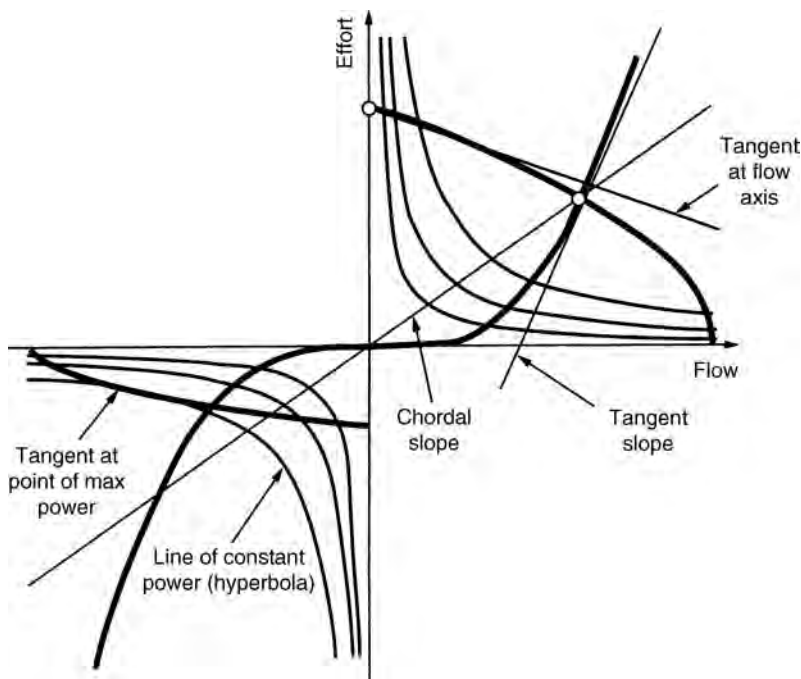


FIGURE 47.8 Definitions in the power plane.

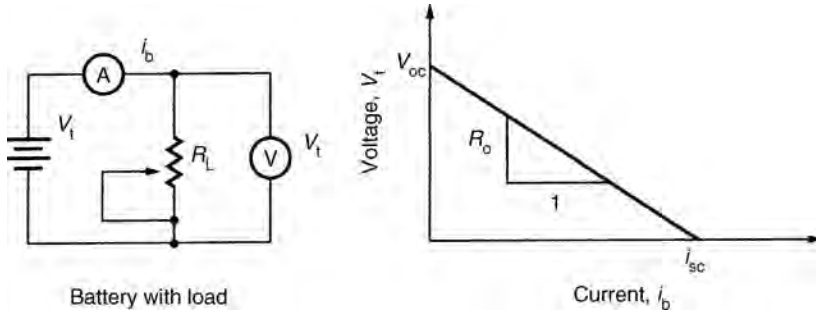


FIGURE 47.9 A battery with load.

47.4.4 Operating Point and Load for Maximum Transfer of Power

Consider a battery with the voltage–current characteristic shown in Figure 47.9. The maximum unloaded terminal voltage (open circuit) of the battery is V_{oc} volts and the short-circuit current is i_{sc} amperes. The equation of the line shown is

$$V_t = V_{oc} - R_o i_b$$

or

$$i_b = i_{sc} - \frac{V_t}{R_o}, \quad (47.27)$$

where V_t is the terminal voltage and i_b is the battery current.

The output impedance $Z_o = -dV_t/di_b = R_o = V_{oc}/i_{sc}$ (a pure resistance in this case). If the battery is loaded by a resistor across its terminals (R_L), the terminal voltage must be

$$V_t = R_L i_b. \quad (47.28)$$

Solving Equations (47.27) and (47.28) simultaneously for V_t and i_b yields the following operating point coordinates:

$$V_t = \frac{V_{oc}}{1 + R_o/R_L} \quad i_b = \frac{V_{oc}}{R_o + R_L}. \quad (47.29)$$

The output power at the operating point is [from (47.29)]

$$\mathcal{P} = V_t i_b = \frac{(V_{oc})^2 R_L}{(R_o + R_L)^2}. \quad (47.30)$$

Clearly, the output power depends on the load resistance. If R_L is zero or infinite, no power is drawn from the battery. To measure V_{oc} , we would want a voltmeter with an infinite input impedance, and to measure i_{sc} , we would want an ammeter with zero resistance. In practice, we would use a voltmeter with an input resistance very large compared to R_o and an ammeter with a resistance very small compared to R_o .

If our objective is to deliver power, then a best value of R_L is that for which the derivative $d\mathcal{P}/dR_L = 0$. This value is the point at which $R_L = R_o$. Alternately, the maximum

power output of the battery for any load occurs at the current (i_b) that maximizes $V_t i_b$. Equation (47.27) can be restated as

$$V_t = V_{oc} \left(1 - \frac{i_b}{i_{sc}} \right) \quad (47.31)$$

so that

$$\mathcal{P} = V_t i_b = V_{oc} \left(1 - \frac{i_b}{i_{sc}} \right) i_b, \quad (47.32)$$

which is maximized at $\mathcal{P} = V_{oc} i_{sc} / 4$ when $i_b = i_{sc} / 2$.

For the battery characteristic, the operating point $i_b = i_{sc} / 2$ yields $V_t = V_{oc} / 2$ [substitution in Equation (47.29)]. A loading resistor characteristic must pass through this point to draw maximum power from the battery, so that the load resistance must be $R_L = V_{oc} / i_{sc} = R_0$. At this operating point, the equivalent resistor within the battery (representing the internal losses in the battery) is dissipating exactly as much power as is being delivered to the load.

If we want maximum power delivery, impedance should match the load to the source, but if we want to minimize power delivery from a source, then impedance mismatching is the key. Impedance matching assures that the source and load will divide the power equally; all other impedances will result in less power transfer.

47.4.5 An Unstable Energy Exchange: Tension-Testing Machine

Although tensile studies of material properties require only a simple test apparatus, it is not simple to interpret the data from such tests. The problem is that the tensile test machine and the specimen can interact (Bell and Ramalingam, 1974) in an unstable way. Almost any desired stress–strain curve can be obtained in a given material by a suitable choice of the test machine’s elastic compliance compared to the specimen.

A tensile test involves the interaction between two springs, one that represents the specimen and the other combined with a velocity source that represents the testing machine. Figure 47.10 shows this simple model. Although the testing machine is linearly elastic and does not yield, the test specimen is not elastic. It undergoes a large plastic deformation in a typical load–elongation test. Normally in such a test, the specimen is to be elongated at a constant cross-head velocity (v). The test machine, however, is not a velocity source as is commonly supposed; that source is really in series with a spring (K)

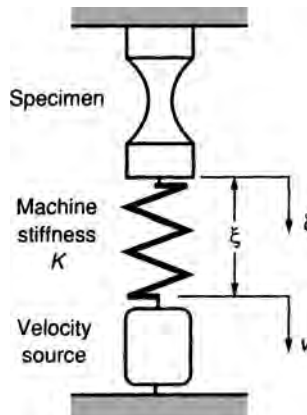


FIGURE 47.10 Simple model of a tensile-testing machine.

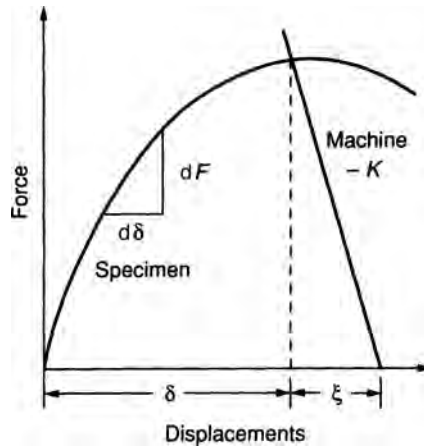


FIGURE 47.11 Components of the force–displacement curve.

representing the elastic deformations of the testing machine structure between the source of the motion and the jaws of the machine.

In the course of a test, the specimen undergoes an elongation (δ) comprised of both elastic and plastic displacements. The test machine undergoes only an elastic displacement (ξ) given by F/K , where F is the load applied to the specimen by the machine at a given cross-head displacement (y), which is really the sum of both the specimen (δ) and the machine (ξ) displacements. Figure 47.11 shows the components of this force–displacement situation.

The cross-head velocity of the test machine is made up of two components as well:

$$v = \frac{dy}{dt} = \frac{d\delta}{dt} + \frac{d\xi}{dt}. \quad (47.33)$$

The machine displacement (ξ), however, is a function of the force being applied, so it is therefore a function of the elongation of the specimen (δ). Equation (47.33) must be rewritten:

$$v = \frac{d\delta}{dt} + \frac{d\xi}{d\delta} \frac{d\delta}{dt} = \frac{d\delta}{dt} \left(1 + \frac{d\xi}{d\delta} \right). \quad (47.34)$$

Since $\xi = F/K$, presuming linearity for the machine, we get the following derivative:

$$\frac{d\xi}{d\delta} = \frac{1}{K} \frac{dF}{d\delta}, \quad (47.35)$$

where the term $dF/d\delta$ is the slope of the force–elongation curve of the specimen. The slope, in other words, is the driving-point stiffness at any point along the test curve. If these last two equations are combined, the specimen's elongation rate is found in terms of the cross-head velocity (v), which is normally constant, the machine stiffness (K), and the driving-point stiffness of the machine ($dF/d\delta$):

$$\frac{d\delta}{dt} = \frac{v}{1 + \frac{1}{K} \frac{dF}{d\delta}}. \quad (47.36)$$

As a test proceeds, however, the specimen starts to neck down, its stiffness begins to decrease, and then it becomes negative. When the driving-point stiffness ($dF/d\delta$) passes

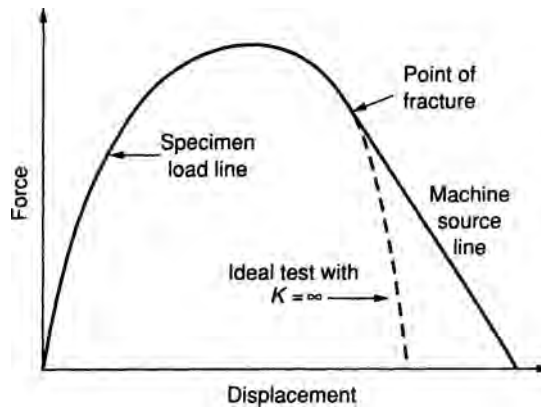


FIGURE 47.12 The point of instability in a tensile test.

through zero, the specimen has reached the maximum force. The stiffness thereafter decreases until it equals $-K$ and at that point the specimen elongation rate ($d\delta/dt$) becomes infinite for any cross-head rate (v), and the system is mechanically unstable. This instability point in a tensile test is the point at which the machine load line is tangential to the load–elongation curve of the material being tested. At that point, the specimen breaks and the energy stored in the machine structure dumps into the specimen at this unstable intersection.

Clearly, percent elongation at failure specifications is not very meaningful because it depends on the test machine stiffness. Figure 47.12 shows that for accurate measurements the stiffness of the test machine must be many times as large as the stiffness of the specimen near failure or the machine will dump its energy into the specimen and force a premature failure.

47.4.6 Fatigue in Bolted Assemblies

The mechanical engineering reader will perhaps recall that preloading a bolt stretches the bolt and compresses the part being bolted, but it is the relative stiffness of the bolt and part that determines what fraction of external loads applied to separate the part from the bolt will be felt by the bolt.⁷ If the objective is to relieve the bolt of these loads, as it is in the head bolts of an automobile engine, then the designer tries to make the part much stiffer than the bolt; he or she mismatches the stiffness of the bolt and the part by specifying a hard head gasket. If the stiffness of the bolt and the part was the same, then they would share the external load equally—the bolt tension would increase by half the applied load while the part compression would decrease by half the applied load. If the gasket was very soft compared to the bolts, then the bolts would see virtually all of the applied load.

47.4.7 Operating Point for Nonlinear Characteristics

Let us continue to use the battery as an example of a linear source. It should be obvious that the maximum power point for the battery is independent of the load it must drive, but the load characteristic, however nonlinear, must pass through this point for maximum power transfer. Figure 47.13 illustrates this. It is not the slope of the load impedance that

⁷ Refer to any text on machine design under the indexed heading “fatigue in bolts.”

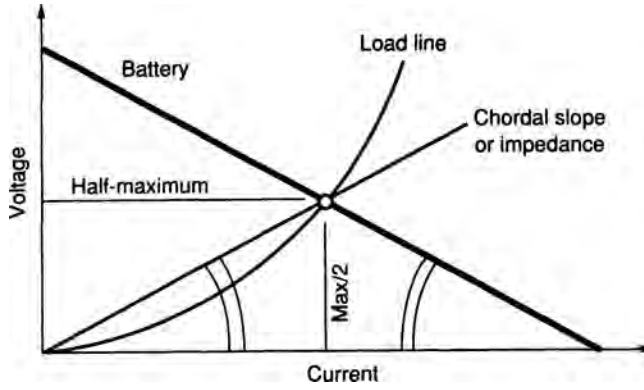


FIGURE 47.13 Chordal impedance matching.

must match the source impedance, it is the chordal slope, the slope of a line from the origin of the power plane to the maximum power point of the source.

An orifice supplied from a constant upstream pressure and loaded at its output is a good example of a nonlinear source. For a short, sharp-edged orifice, the orifice equation⁸ applies:

$$Q_o = C_d A_o \sqrt{\frac{2}{\rho} \Delta P} = C_d A_o \sqrt{\frac{2}{\rho} (P_{up} - P_{down})}, \quad (47.37)$$

where Q_o is the orifice flow, C_d is the discharge coefficient of the orifice, A_o is the orifice area, ρ is the density of the fluid, and $\Delta P = P_{up} - P_{down}$ is the pressure drop across the orifice. If the upstream pressure is kept constant, then P_{down} may be considered the output pressure, and Q_o may be considered the output flow from this orifice, a characteristic typical of many hydraulic valves. Where is the maximum power point on this characteristic?

Hydraulic power \mathcal{P} is the product $Q \cdot P$. It is a maximum when it is stationary with respect to either Q or P . If we nondimensionalize Equation (47.37) with respect to the maximum flow through the orifice when $P_{down} = 0$, then the orifice equation becomes the following:⁹

$$Q^* = \frac{Q_o}{Q_{max}} = \sqrt{1 - \frac{P_{down}}{P_{up}}} = \sqrt{1 - P^*} \quad (47.38)$$

$$\mathcal{P}^* = \frac{\mathcal{P}}{\mathcal{P}_{max}} = P^* Q^* = P^* \sqrt{1 - P^*} \quad (47.39)$$

$$\frac{d\mathcal{P}^*}{dP^*} = \sqrt{1 - P^*} - \frac{P^*}{2\sqrt{1 - P^*}} = 0. \quad (47.40)$$

⁸ This is derived from Bernoulli's equation. The discharge coefficient (C_d) corrects for viscous effects not considered and for the existence of a vena contracta or convergence in the flow through the orifice, which makes the area of the jet smaller than the orifice itself. For most oils $C_d \approx 0.62$.

⁹ Where bold letters will be used to indicate the nondimensional forms.

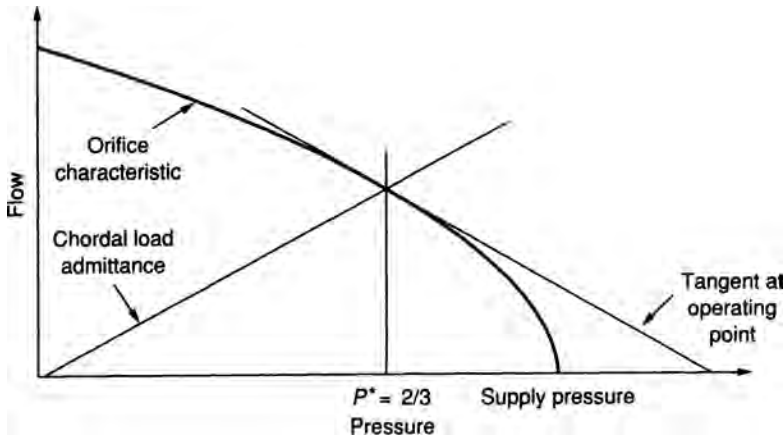


FIGURE 47.14 Nonlinear impedance matching.

for which $P^* = 2/3$, and by substitution into (47.38), the flow at that point is $Q^* = 1/\sqrt{3}$, and the maximum power delivered is $(\sqrt{2}/3)P_{\max}Q_{\max}$. The output admittance of the orifice is the slope of the curve at any operating point:

$$Z_o = \frac{dQ^*}{dP^*} = \frac{1}{2\sqrt{1-P^*}} = \frac{\sqrt{3}}{2} \Big|_{P^*=2/3}. \quad (47.41)$$

The load for which maximum power will be delivered, whether it has a linear or nonlinear characteristic, must pass through the maximum power point. Its chordal admittance must be

$$Z_{\text{chordal}} = \frac{Q_{\text{op}}^*}{P_{\text{op}}^*} = \frac{1/\sqrt{3}}{2/3} = \frac{\sqrt{3}}{2}, \quad (47.42)$$

which is exactly the slope in Equation (47.41). Figure 47.14 shows this graphically.

Note that for this power plane, the ordinate is “flow” and abscissa is “pressure” so that the slopes shown are admittances rather than impedances. While in other figures, the abscissa has always been effort and the ordinate flow, it is conventional in hydraulic systems to show these figures the other way probably because pressure is usually the independent variable in hydraulic system characteristics: pressure is controlled and flow is measured.

47.4.8 Graphical Determination of Output Impedance for Nonlinear Systems

A three-way hydraulic valve supplies or drains its load through a pair of variable orifices, one connecting the supply to the load and the other connecting the drain or tank to the load. These orifices are operated (typically on a single moving part of the valve) in a push-pull fashion; that is, as one opens, the other closes. If both orifices are partially open together when the valve is centered, the three-way valve is open centered. If one is wide open just as the other closes, the valve has no overlap.

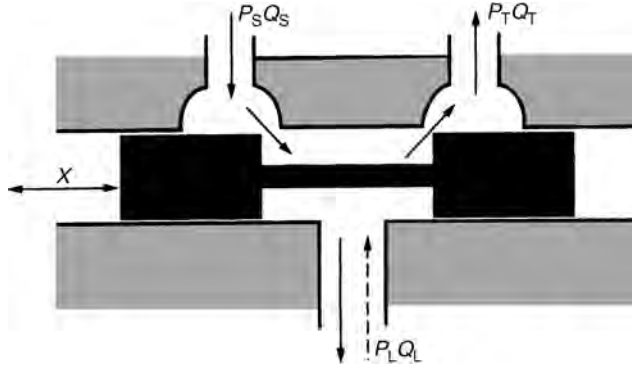


FIGURE 47.15 A three-way valve geometry and output.

Suppose it is required to find the P - Q characteristics of the load port for flows both into and out of the valve. The system is shown in Figure 47.15. Thus for the upstream and downstream orifices respectively, Equation (47.37) becomes

$$Q_u = C_d A_u \sqrt{\frac{2}{\rho} (P_S - P_L)} \quad \text{and} \quad Q_d = C_d A_d \sqrt{\frac{2}{\rho} (P_L - P_T)}. \quad (47.43)$$

The load flow (Q_L) is the difference between Q_u , and Q_d . Also, the tank pressure (P_T) can be taken as zero. These equations combined become

$$Q_L = C_d A_u \sqrt{\frac{2}{\rho} (P_S - P_L)} - C_d A_d \sqrt{\frac{2}{\rho} P_L} \quad (47.44)$$

It is much more convenient to work with Equation (47.44) in dimensionless form. If we assume that the maximum upstream and downstream areas are the same and that they are truly push-pull, then we can nondimensionalize the area using A_{\max} . The discharge coefficient (C_d) is a constant for conventional valve geometries. The supply pressure (P_S) is a convenient term for the pressure nondimensionalization. Flows can be nondimensionalized with respect to a maximum flow that would pass through either orifice at full area A_{\max} with P_S acting across it:

$$Q_{\max} = C_d A_{\max} \sqrt{\frac{2}{\rho} P_S}. \quad (47.45)$$

If we set $P_L/P_S = \mathbf{P}$ and $Q_L/Q_S = \mathbf{Q}$ and express the upstream and downstream orifice sizes as push-pull fractions of A_{\max} , $0.5(1+x)A_{\max}$ upstream and $0.5(1-x)A_{\max}$ downstream, where $-1 \leq x \leq 1$ and the valve is centered for $x=0$, then Equation (47.44) combined with (47.45) yields

$$\mathbf{Q} = \mathbf{Q}_u - \mathbf{Q}_d = 0.5(1+x)\sqrt{1-\mathbf{P}} - 0.5(1-x)\sqrt{\mathbf{P}}, \quad (47.46)$$

which is one of those unfortunate equations in which the radical cannot be eliminated by squaring both sides. While Equation (47.46) can be readily plotted, \mathbf{Q} versus \mathbf{P} with x (the valve stroke) as a parameter, it is instructive to construct it instead from its parts.

The term $0.5(1+x)\sqrt{1-\mathbf{P}}$ is the characteristic family for the upstream orifice, and the term $0.5(1-x)\sqrt{\mathbf{P}}$ is the characteristic family for the downstream orifice. The first of these are parabolas to the left (on their sides because they are roots), starting at $\mathbf{Q}_u = 0.5(1+x)$, $\mathbf{P}=0$ and ending at $\mathbf{Q}_u=0$, $\mathbf{P}=1$ (there is no flow when the downstream

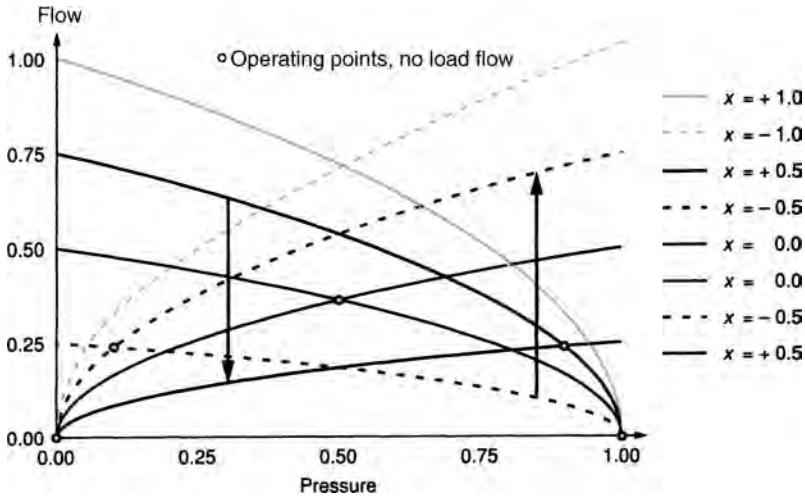


FIGURE 47.16 Flow versus pressure drop characteristics for the valve orifices.

pressure equals the upstream pressure). The second term starts at $Q_d = 0$, $P = 0$ and rises to the right to the points $Q_d = 0.5(1 - x)$, $P = 1$. All this is shown in Figure 47.16. If there is no load flow, then the curves for the upstream orifice show its output characteristic while those for the downstream orifice represent the only load. The intersections predict the operating pressures for $Q = 0$ as the valve is stroked, $-1 \leq x \leq 1$.

If there is a load flow (Q), then continuity must be served. This requires that

$$Q = Q_u - Q_d \rightarrow Q_u = Q + Q_d. \quad (47.47)$$

In Figure 47.16, Q is simply a vertical bar between the curves for Q_u and Q_d whose length is the load flow. Thus, for any load pressure value on the abscissa, the load flow is determined as the vertical distance between the input orifice curve and the output orifice curve. The load flow is positive if the upstream orifice curve is above the downstream orifice curve but negative otherwise. These data can be picked off and plotted as the Q - P characteristic of the valve, as shown in Figure 47.17.

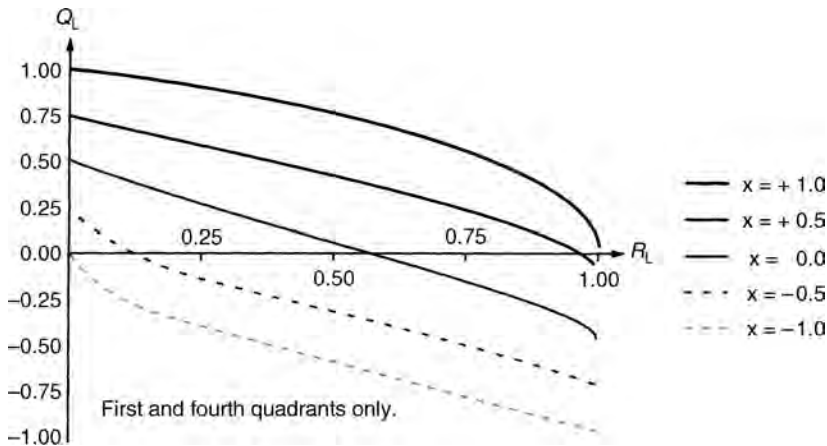


FIGURE 47.17 The output characteristic of the valve.

Figure 47.17 illustrates that an operating point on an input–output characteristic forces compatibility and continuity, but as long as those are preserved, a second energy exchange as at the $P_L - Q_L$ port of this valve can still be accommodated. This approach can be extremely useful in systems such as air-conditioning ducting where the fan characteristic is known only graphically, and then its output impedance at some point along the ducting must be determined. Exactly, the same procedure is used.

47.5 TRANSFORMING THE OPERATING POINT

It is often not among the system designer's options to choose either the output impedance of the power source in the system or the input impedance of the load that must drive. The only recourse at that point is to insert a transformer, transducer, or gyrator in the system if it does not already contain one or to vary the modulus of the two-port element if it does. In the old tube-type audio systems, the output impedance of the push–pull power tubes exceeded $1000\ \Omega$ while the input impedance of the speakers available then was 4, 8, or $16\ \Omega$. To match the amplifier to the load, each channel had a large transformer for speaker connections with taps having turns ratios of $\sqrt{1000/4}$, $\sqrt{1000/8}$, and $\sqrt{1000/16}$. With this arrangement, maximum power transfer was assured down to the lowest frequencies for which the transformers were designed.

47.5.1 Transducer-Matched Impedances

Suppose a permanent direct current (dc) magnet servomotor is to drive a screw that, in turn, drives a mass, perhaps a machine-tool table. If the objective is to minimize the move time from stationary start to full stop, then the optimal trajectory for the servo, assuming equal and constant acceleration and deceleration, is well known: maximum acceleration to either half of the move distance or maximum velocity, whichever comes first, followed by maximum deceleration to the finish. These trajectories are illustrated in Figure 47.18.

The motor operates in two modes: at maximum acceleration (maximum torque), which is set by the maximum short-term current permitted by the coercivity of the motor magnets and the capacity of the commutation, and at maximum speed, if that is reached, set by the maximum voltage available or by whatever the commutation allows. Often, servo designers accomplish these two modes by using an overvoltage (two to three times rating) during acceleration and deceleration, which runs the power amplifier as a current source,

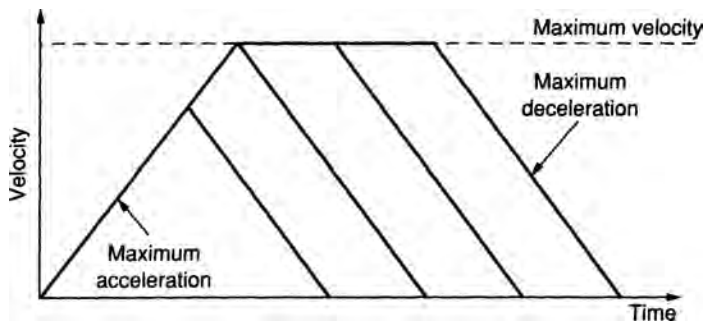


FIGURE 47.18 Optimal trajectories for point-to-point move.

and then as maximum speed is attained, switching the amplifier voltage limit to the motor rating, which runs the amplifier as a voltage source. This achieves a constant acceleration and a constant top speed.¹⁰

During the acceleration or deceleration phase of the trajectory, the electrical torque available is accelerating two inertias: the motor itself and the load. Since the motor is acting as an electromagnetic torque source while constant current is supplied, $\tau_{\text{motor}} = K_m i_{\text{armature}}$, the mechanical output impedance viewed at the motor shaft is the rotor inertia between the output shaft and the electromagnetic torque source:

$$Z_o \left| \begin{array}{l} \text{motor shaft with} \\ \text{constant armature current} \end{array} \right. = J_{\text{armature}} s. \quad (47.48)$$

During acceleration and deceleration, the optimal load impedance (Z_{load}) will be equal to the motor inertia, and the available electromagnetic torque will be shared equally by the motor armature and the load.

The load given in this example is primarily massive, so with reflection through the screw, the load inertia is computed as the load inertia times the square of the transducer ratio (p for the screw; see Table 47.2):

$$Z_{\text{load}} = p^2 M_{\text{load}} s. \quad (47.49)$$

To the extent possible, the pitch of the screw or the inertia of the motor should be chosen to achieve this match. Failing both of those options, a gear box should be placed between the screw and the motor to accomplish the match required.

47.5.2 Impedance Requirements for Mixed Systems

When a source characteristic is primarily real or static (so that the source impedance is resistive) and the load is reactive or dynamic (dominated by energy storage elements), then impedance matching in the strictest sense is impossible, and the concept of passing the load line through the source characteristic at the maximum power point does not make sense. How then does one match a static source to a dynamic load or the reverse?

Figure 47.19 shows an electrohydraulic position servo driving a mass load with negligible damping losses, and Figure 47.20 shows the pressure flow characteristics of the

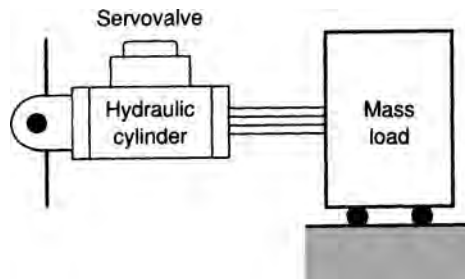


FIGURE 47.19 Electrohydraulic position servo.

¹⁰ Actually, designers rarely use maximum acceleration or maximum velocity because such full-scale values leave no overhead for control. A servo running in saturation is an open-loop system.

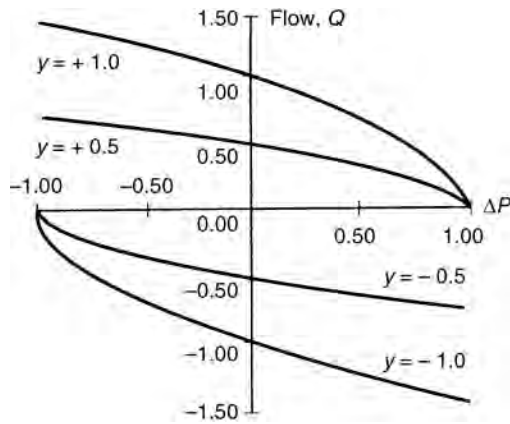


FIGURE 47.20 Servovalve characteristic.

servovalve: a family of parabolas in the power plane used for hydraulic systems. The transducer between this hydraulic power plane and the force–velocity power plane in which the load operates is a piston of area A . There are two equations:

$$P = \frac{1}{A}F \quad \text{and} \quad Q = Av. \quad (47.50)$$

With these equations, a load in F – v coordinates can be transformed to a load in P – Q coordinates for superposition on the source characteristics.

The impedance of the mass load, however, is simply $Z_{\text{load}} = (M)(s)$, but this cannot be plotted in an F – v coordinate system because it is the slope of a straight line in energetic coordinates: F versus dv/dt . To match load to source, we need the force–velocity relationship for the load. A servo designer normally has a frequency response in mind for his or her application or can determine one from a linearization of the system. Suppose, in this example, that the valve operating between a supply pressure (P_s) and a maximum load flow (Q_{max}) must drive the mass load (M) through the piston of the area (A), with a frequency response flat in position to a frequency of ω (radians per second), at an amplitude of D_{max} . With these conditions, the force–velocity relationship for the load can be found as follows.

If y is the displacement, \dot{y} is the velocity and \ddot{y} is the acceleration, then the most taxing demand on the servo will be $y = D_{\text{max}} \sin(\omega t)$, where t is time. Then $\dot{y} = D_{\text{max}} \omega \cos(\omega t)$ and $\ddot{y} = -D_{\text{max}} \omega^2 \sin(\omega t)$. Given that $F = M\ddot{y}$ is the load equation, however, the load can now be expressed parametrically as a pair of equations:

$$v = \dot{y} = D_{\text{max}} \omega \cos(\omega t) \quad \text{and} \quad F = -MD_{\text{max}} \omega^2 \sin(\omega t). \quad (47.51)$$

If these are cross-plotted in the force–velocity plane with time as a parameter, the plot traces out an ellipse as time goes from zero through multiples of $2\pi/\omega$. Transforming these to the P – Q plane requires application of the piston equations (47.50) to yield

$$Q = AD_{\text{max}} \omega \cos(\omega t) \quad \text{and} \quad P = -\frac{MD_{\text{max}} \omega^2}{A} \sin(\omega t). \quad (47.52)$$

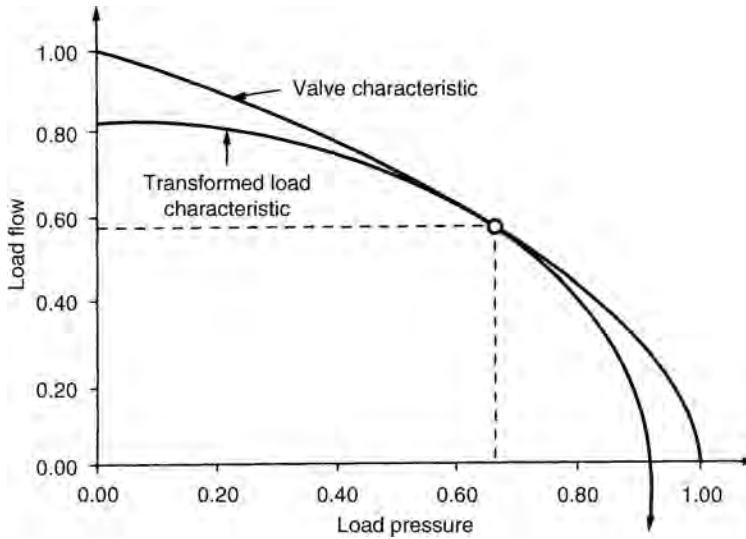


FIGURE 47.21 Matching power requirements for a dynamic load.

If the valve is to drive the mass through the piston around the trajectory, $y = D_{\max} \sin(\omega t)$, then the valve output characteristic for the maximum valve stroke must entirely enclose the elliptical load characteristic derived in Equation (47.52). Furthermore, if the valve and load are to be perfectly matched, then the valve characteristic and the load ellipse must be tangent at the maximum power point for the valve: $P = 2/3 P_s$, $Q = (1/\sqrt{3}) Q_{\max}$. This is shown in Figure 47.21.

Note that this requirement means that neither maximum valve output pressure nor maximum valve flow will ever be reached while the load executes its maximal sinusoid. If $P_s = -MD_{\max} \omega^2 / A$ and $Q_{\max} = AD_{\max} \omega$, the valve sizing would have provided inadequate power for reaching any but those points on the trajectory, which would therefore have followed the valve characteristic instead. The correct matching relationships are those which satisfy the following equations:

$$\frac{1}{\sqrt{3}} Q_{\max} = AD_{\max} \omega \cos(\omega t) \quad \text{and} \quad \frac{2}{3} P_s = \frac{MD_{\max} \omega^2}{A} \sin(\omega t) \quad (47.53)$$

at the appropriate time. These equations relate D_{\max} , ω , M , Q_{\max} , P_s , and A , so any five of these determine t and the sixth. If the load mass, supply pressure, piston area, peak frequency, and maximum amplitude are all known, for example, Equation (47.53) size the valve by determining its maximum required flow. If the valve has been selected, these equations will size the piston.

In the example above, the load was purely massive. Most real loads are dissipative as well. The procedure outlined above, however, need only be modified by adding the damping term to the force equation:

$$F = M\ddot{y} + B\dot{y} = -MD_{\max} \omega^2 \sin(\omega t) + BD_{\max} \omega \cos(\omega t), \quad (47.54)$$

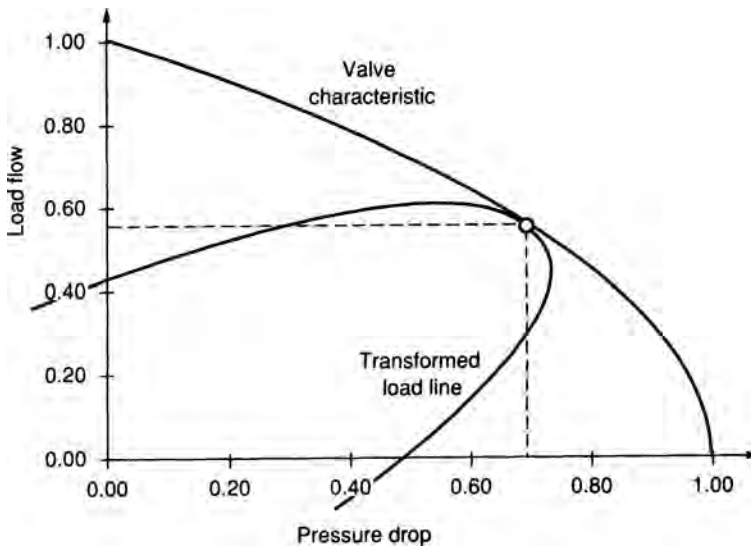


FIGURE 47.22 Matching power requirements for a mass load with dissipation.

where B is the damping coefficient. This has the effect of tipping the axis of the elliptical load line up to the right, but the rest of the development follows as before with the added term. Figure 47.22 shows the result. If the load mass were to be negligible, then the elliptical transformed load line in Figure 47.22 would collapse to a line passing through the maximum power point of the valve characteristic, as discussed previously.

47.6 MEASUREMENT SYSTEMS

Measurement extracts information about the state of a measured system, usually in the form of one of the factors in the measurement domain whose product is power or energy flux, such as voltage or current, pressure or flow, and so on. A measurement usually extracts some energy from the system being measured, if only in a transient sense; each link in the chain from measurement to display involves a further exchange of energy or power. If we measure voltage in a circuit, for example, we must draw some current, at least instantaneously, so that the power is dissipated by effective resistances, or energy is stored capacitively, inductively, or both.

At the measurement interface, we wish to disturb the measured system as little as possible by extracting as little energy or power as can be managed. It is usually our objective to pass power or energy along the chain of elements that form the measurement system, and there is a best combination of the energy variables to attain the optimum transfer. This chapter has dealt with these issues: the maximization of energy transfer within the system where we want it and the minimization of energy theft at the measurement interface where we do not.

47.6.1 Interaction in Instrument Systems

The generalized instrument consists of a number of interconnected parts with both abstract and physical embodiments. An orifice flow-metering system might consist of the following chain: an orifice plate converts the flow to a pressure drop; a diaphragm converts the pressure

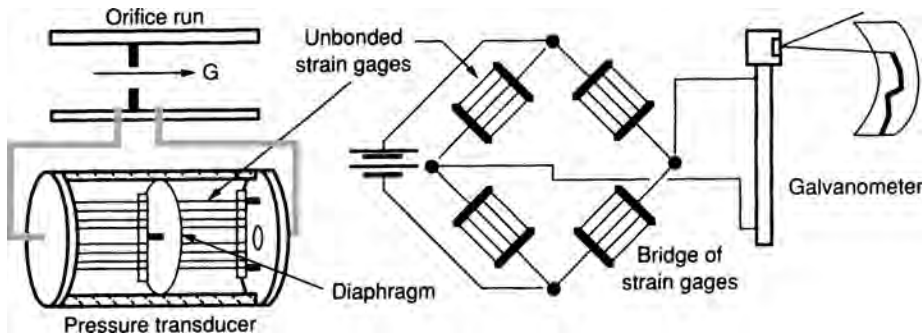


FIGURE 47.23 The chain of a flow measurement.

drop to a force; a spring (perhaps an unbonded strain-gage bridge) converts the force to a displacement; the strain gages convert the displacement to a resistance change; a bridge arrangement converts the resistance change to a differential voltage; and a galvanometer converts the voltage to a trace on paper. Figure 47.23 illustrates this chain.

In setting up this chain, the orifice is sized to minimize the pressure drop resulting from our flow measurement, and the diaphragm must be sized for minimum pumping volume during transients in pressure, or it will alter the flow reading in a transient sense. The diaphragm, however, has an output stiffness; the force it transmits decreases with increasing displacement, and it is driving the unbonded gages, converting force to displacement. The combined stiffness of the unbonded strain gages, which will be linear springs, must equal the average stiffness of the diaphragm¹¹ to ensure that the strain energy transfer is maximized for a given input energy from the fluid.

The displacement–strain relationship is a definition. Further along the chain, however, this strain is converted to a differential voltage whose magnitude depends on the input impedance of the galvanometer, which is loading the bridge. For bridges constructed with the same resistance in each arm, whether the resistors are active gages or not, the input and output impedances are the same: the resistance of one arm, usually the gage resistance. If we ignore dynamic considerations (such as galvanometer damping), the optimal galvanometer will therefore have the same resistance for maximum deflection per unit strain as one gage in the bridge.

At the measurement interface, therefore, the objective is to mismatch the impedances between the measurement system and the measured system as much as possible. There is more to this than the selections of voltage-measuring devices with higher impedance than the system in which voltage is being measured and current-measuring devices with lower impedance than the circuits in which they are placed. The measurement and control engineer must always be aware of dynamic loading as well.

The output impedance of any piezoelectric device, for example, is almost purely capacitive and is typically only a few picofarads for small devices. The input impedance of most oscilloscopes is a parallel combination of 1 M Ω and 100 pF. Either or both of

¹¹ If the diaphragm were rigidly supported in the center, then the force on the support would be the pressure times the effective area of the diaphragm. If the support were not rigid, this output force would depend nonlinearly on the support displacement. The plot's negative slope of force transmitted vs. displacement is the output stiffness of the diaphragm, and pressure \times area is the force source in this model.

those would load a piezoelectric device to near uselessness, even though for most other purposes they are high impedances. The resistance would reduce the charge on the crystal much too quickly, and the capacitance would steal charge and reduce the voltage output drastically. An attempt to measure the pressure in a small volume with a transducer that has a large swept volume itself would meet with the same failure: the displacement of the transducer would alter the volume in which the pressure was measured. Holography has become popular in the study of the vibration of thin plates and shells because it does not load the structure by adding mass as an array of accelerometers would.

47.6.2 Dynamic Interactions in Instrument Systems

It is not only the steady-state loading of a measurement that is of concern; under many circumstances, the unsuspected dynamics of the instrument being used will lead to erroneous results. Consider the simple measurement interface shown in Figure 47.24. The readout instrument, an oscilloscope for example, has both resistance and capacitance. The load impedance is therefore the parallel combination of these:

$$Z_{\text{inst}} = \frac{R_i(1/C_i s)}{R_i + 1/C_i s} = \frac{R_i}{R_i C_i s + 1} = \frac{R_i}{\tau_i s + 1} \quad \text{with} \quad \tau_i = R_i C_i. \quad (47.55)$$

Thus, the load depends on the frequency of the voltage being measured. This might not concern us in the sense that we can predict it and compensate for the known phase shift, which this will induce, but when the interaction of the two systems is considered, the problem becomes more obvious. The total impedance loading the measurement of V_s includes the source impedance Z_o . Suppose this is purely resistive (R_o):

$$Z_{\text{total}} = R_o + Z_{\text{inst}} = \frac{R_o R_i C_i s + (R_o + R_i)}{R_i C_i s + 1}. \quad (47.56)$$

The readout instrument is sensitive only to the voltage it sees, which has been reduced by the voltage drop in R_o . In general, the measured voltage, V_{meas} , is the voltage across the instrument's input resistor and capacitor, R_i and C_i :

$$\frac{V_{\text{meas}}}{V_s} = \frac{Z_{\text{inst}}}{Z_{\text{total}}} = \frac{R_i}{R_o R_i C_i s + (R_o + R_i)} = \frac{1}{R_o C_i s + (R_o/R_i + 1)}. \quad (47.57)$$

Unless we know the output impedance at the point of measurement, in this case R_o , we do not even know the time constant or break frequency germane to the measurement. In the event that Z_o is also complex (has reactive terms), this situation is more complicated, and if $Z_o = L_o s + R_o$, for example, the system could even be oscillatory.

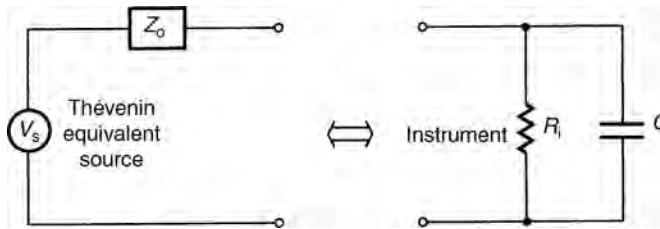


FIGURE 47.24 A measurement with a dynamic instrument.

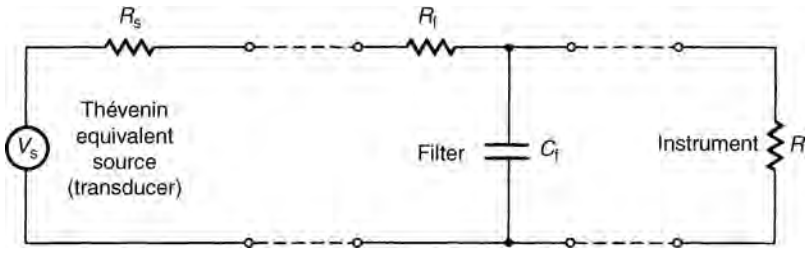


FIGURE 47.25 Schematic of cascaded instrument system.

Sometimes, frequency-dependent impedances are intentionally introduced into a measurement system, most commonly in the form of passive filters. Figure 47.25¹² shows a first-order, low-pass filter being driven by a source with a nonzero output impedance (R_s) and being loaded by a readout instrument with finite input impedance (R_i).

The loading problem is potentially present at the interface of each stage indicated by the broken lines and terminals. What is not so obvious is that the dynamics of this filter are quite sensitive to its source and load. An unwary designer might choose the filter time constant ($\tau_f = R_f C_f$) with little regard for the values of R_f and C_f and essentially design the filter under the simple assumption that source and load were ideal.

To the instrument, however, the filter is part of the source impedance, so the true source impedance combines R_s , R_f , and C_f , and the Thévenin equivalent source voltage (V_T) is no longer simply V_s . The Thévenin voltage (V_T) and the new output impedance (Z_T) with the filter included become

$$V_T = \frac{V_s}{(R_s + R_f)C_f s + 1} \quad \text{and} \quad Z_T = \frac{R_s + R_f}{(R_s + R_f)C_f s + 1}. \quad (47.58)$$

Figure 47.26 shows the new equivalent circuit for the cascaded instrument system.

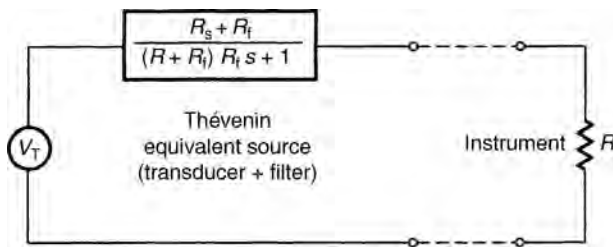


FIGURE 47.26 Thévenin equivalent of cascade system.

¹² This example is drawn entirely from Ref. (Nachtigal, 1978), with the permission of Dr. C. L. Nachtigal. In this and the previous example, Figures 47.24 and 47.25, the source voltage is referred to as a Thévenin equivalent source. In single-loop circuits the source voltage is by definition the Thévenin voltage, since removing one element from the loop causes it to become open circuited. If the source is a sensor, for example, the magnitude of the voltage is governed by the value of the sensed variable and, of course, its own design parameters. The Thévenin, or open-circuit voltage, is specified on the sensor data sheet.

The instrument is now loading this system, and the output (V_o) is equal to the voltage across R_i . Therefore, in terms of the original measured voltage (V_s)

$$\frac{V_o}{V_s} = \frac{1/(1 + (R_s + R_f)/R_i)}{[(1 + R_s/R_f)/(1 + (R_s + R_f)/R_i)]R_f C_f s + 1} = \frac{K_f}{\tau s + 1} \quad (47.59)$$

where

$$K_f = \frac{1}{1 + (R_s/R_i) + (R_f/R_i)} = \frac{1}{1 + \beta} \quad (47.60)$$

and

$$\tau = \left(\frac{1 + (R_s/R_f)}{1 + (R_s/R_i) + (R_f/R_i)} \right) R_f C_f = \left(\frac{1 + \alpha}{1 + \beta} \right) R_f C_f. \quad (47.61)$$

Our filter has its expected characteristics, $\tau = \tau_f = R_f C_f$ and $K_f = 1$, only if both α and β approach zero, a circumstance that requires $R_s \rightarrow 0$ and $R_i \rightarrow \infty$ while R_f remains finite. As the instrument resistance (R_i) decreases, the static gain and time constant decrease as well, which means that the break frequency of the filter increases from the designed value because the instrument provides another path for the discharge of the capacitor in the filter. As the transducer resistance (R_s) increases from zero, the static sensitivity (K) again decreases, but this time the break frequency decreases as well.

This analysis shows us that source loading by a filter and filter loading by a readout instrument can cause significant changes in both the designed filter gain and break frequency. Only if α and β in Equations (47.60) and (47.61) remain equal does the filter break frequency remain unscathed by nonideal source and load impedances. This condition requires that

$$\frac{R_s}{R_f} = \frac{R_s + R_f}{R_i}$$

or

$$R_f^2 + R_s R_f - R_s R_i = 0 \quad (47.62)$$

Dividing the second of Equations (47.62) by R_s^2 and solving it for R_f/R_s yield

$$\frac{R_f}{R_s} = \frac{1}{2} \left(-1 \pm \sqrt{1 + 4 \frac{R_i}{R_s}} \right) \cong \sqrt{\frac{R_i}{R_s}} \quad (47.63)$$

since any realistic measurement system has $R_i/R_s \gg 1$, and since the resistances must be positive, the negative solution is discarded. This approximation is equivalent to saying that R_f should be chosen to be the geometric mean of the estimated or known values of R_i and R_s ; that is, $R_f = \sqrt{R_i R_s}$ presuming only that $R_i \gg R_s$. For this choice of filter resistance, the resulting gain is given by

$$K = \frac{1}{1 + \sqrt{R_s/R_i} + R_s/R_i} = \frac{1}{1 + (R_s/R_i)^{3/2}} \cong 1 \quad \text{for } R_i \gg R_s. \quad (47.64)$$

If the filter fails these conditions, it must either be carefully designed for the task or it should be an active¹³ filter with a high input impedance and low output impedance.

47.6.3 Null Instruments

Many instruments are servos, active systems designed to oppose the variable they measure so as to decrease their demand on the complementary energy or power variable to zero. In a steady-state sense, these instruments draw no power or energy from the measurement interface because these energetic requirements are provided by the instrument's power supply. These instruments therefore have infinite or zero input impedance in steady state. Examples abound: slidewire potentiometers for strain readout and thermocouple readout both measure voltage by servoing to zero current. Servo accelerometers avoid the problems with temperature inherent in the elasticity variations of metals and crystals by servoing the motion of the proof mass to zero and measuring the force required to do it. The differential pressure-cells used in the process control industry measure differential pressure across a diaphragm or capsule while preventing the diaphragm or capsule (separating the two pressures) from moving. They thus avoid having to worry about the nonlinear elastic characteristics of the capsule and seals in the system.¹⁴ Their output is either the current in a voice coil or the pressure in a bellows necessary to oppose the motion. In all of these cases, however, there is a transient displacement until the servo zeros it out, and some energy must be lost to transfer the information required by the servo in the instrument. For this reason, null instruments are no better at measuring stored energy, the voltage on a small capacitor for example, than passive but high-impedance instruments, particularly if the energy consumed to reach null is not extremely low.

47.7 DISTRIBUTED SYSTEMS IN BRIEF

While a detailed study of the input–output relationships for distributed systems is beyond the scope of this chapter, a brief discussion can tie these into the concepts already covered. All of the systems discussed to this point have been lumped, a label that implies physical dimensions are not of importance; the system parts can be considered as point objects. Considering a tank, for example, to be a point does not mean that the tank has no dimensions; it merely means that the internal pressure is considered to be the same everywhere within it; conditions in its interior are absolutely uniform. When studying lumped circuits, we are not concerned with the dimensions of the circuit elements or with their distances from each other on the circuit board. In reality, the nodes of the circuit have lengths, but they are ignored for the purposes of analyzing the lumped circuit.

In the mechanical domain, we consider that masses are rigid and behave as if the forces acting were applied at a point, and if we are interested in the distributed properties of an object, we consider only its moment of inertia. In each of these examples, changes in the physical variables of our model are assumed to propagate instantaneously, even though it is well known that all have finite propagation velocities. A system element must be considered to be distributed to have properties that vary with a physical dimension if that

¹³ Incorporating amplification, usually an operational amplifier.

¹⁴ Dry friction would nonetheless be fatal to the instrument because it would be fatal to its servo and would lead to dynamic instabilities of the limit-cycle variety.

assumption is not true. This occurs whenever the dimensions of the object are large compared to the characteristic size of the events occurring.

Mechanical disturbances of all kinds propagate at the speed of sound in the medium involved, and electromagnetic disturbances propagate at speeds near the speed of light, depending again on the medium. A hammer blow to the end of a long, slender bar, for example, induces a strain pulse at the struck end which travels into the bar (informing the interior of the event) at the speed of sound in compression in the bar material ($c = \sqrt{E/\rho}$, where E is Young's modulus of the material, ρ is its mass density, and c is the propagation speed of compressive or tensile events). If the pulse duration is short, its physical length approaches that of the bar or may be shorter than the bar. Our simple lumped models of the bar's behavior [$F = M\ddot{y}$ and $\int F dt = Mv(t) - Mv(0)$], which treat it as a solid rigid object, are incorrect. Similarly, if a small explosion, a spark, for example, is initiated in a tank, the pressure in the tank will not remain uniform throughout. Instead, pressure waves will propagate within the tank at the speed of sound until damping takes its toll; we can no longer consider the tank to be a simple lumped capacitor, at least in the time scale of the spark event.

47.7.1 Impedance of a Distributed System

Imagine a long, slender tank of water, a trough, open at the top, and perform the following *thought* experiment. If one end wall was moved inward suddenly, the level of water at that end would rise higher up than was required by the change in tank volume because the remaining water further along the tank would not change level instantaneously. Then, this wave coming down off the tank wall would travel to the other end where it would slosh up the far wall and return. This wave would continue to slosh back and forth at decreasing amplitude as viscosity took its toll, until finally the surface would be calm again at a new, slightly higher level. Each time the wave reached an end, it would be returned in kind, that is, with the same sign.

Now suppose that the far end of the tank opened into a large lake, so large that no level changes would take place when an end wall was moved, and again move the remaining end wall inward quickly. Then, when the first wave reached the opening to the lake, it would leave the tank and be lost in the lake. But that would involve water leaving the tank in excess of the volume change in the tank, and a negative wave would return from the open end to signal the new, lower level required by the loss. The closed end would reflect this wave as a rarefaction, and when that returned to the open end, lake water would spill back in as an upward wave. Eventually these alternating processes would return the level in the tank to that of the lake.

A closed end returns a wave of like kind, and an open end returns a wave of opposite kind. If there are no losses as the waves hit the ends of the tank, a wave of strength $+1$ is reflected with strength $+1$ from a closed end and with a strength -1 from an open end. This implies that there is an end condition somewhere between *closed* and *open* from which a wave will not be reflected at all. A suitably constructed porous wall, in this example, would simply absorb the wave completely by accepting and dissipating all of its energy. The impedance of this wave-matched wall is the wave impedance of the channel and is a characteristic of it, depending on the inductive and capacitive properties of the medium.

The 75- and 300- Ω markings on the antenna connections of a television receiver imply two things: First the input impedances of those terminals are resistive at 75 and 300 Ω , respectively, and, second, the coaxial cable and flat-lead antenna wiring are really

waveguides whose wave impedances are 75 and 300 Ω . By matching the impedance of the cable at the receiver terminals, we are assured that all the incoming *wave energy* will be absorbed by the receiver and none will be reflected back up the cable to the antenna and thus lost.

47.8 CONCLUDING REMARKS

This chapter has demonstrated an alternate viewpoint for the interaction of systems with each other. Control engineers are quite accustomed to transfer functions: relationships in the frequency (s or Laplace) domain between a variable at one point in a system and another at some other point, most often between inputs to and outputs from a controlled system. This chapter has dealt with a special class of these relationships between the complementary variables of power at a single point in a system. These special transfer functions are called driving-point impedances or admittances, and they determine how one subsystem will load or be loaded by another.

Admittances are the reciprocals of impedances, and both are unique properties of a system. The Laplace operator (s) expresses these properties as a polynomial ratio. The denominator polynomial (when set to zero, it becomes the characteristic equation of the system) is always the same, and the numerator polynomial is a function of the location of the point considered in the system. It was also shown that driving-point impedances are not a function of the controlled variables on any ideal sources the system contains. Instead, all effort sources may be replaced by solid connections, and all flow sources may be removed before the driving-point impedance is computed.

When two systems are connected together at a driving point, port, or pair of terminals, usually so that one can pass energy or information to the other, then there is a favorable relationship between the impedances of the two systems that depends on the objective of the connection. When it is desired to pass energy or power from one system to the other, then the output impedance of the driving system should match the input impedance of the driven system. If neither the driver nor the driven are adjustable, then a transducer, gyrator, or transformer is used to match them by selecting the modulus to achieve the match. Any impedance seen through a transformer, for example, appears to be increased or diminished by the square of the transformer ratio. In a chain of subsystems, it is not necessary to install the transformer at the driving point under consideration; the correct ratio can be determined no matter where it is placed within the chain because the square of the modulus will always appear in one of the driving-point impedances.

If the interconnection represents a measurement interface, then the most favorable relationship between the driving-point impedances is the largest possible mismatch consistent with obtaining the measurement. The ideal instruments for measuring efforts have infinite input impedance and the ideal instruments for measuring flows have infinite input admittances. Instruments that measure the integral of flows, such things as volume, charge, and displacements, should have very low compliance (should displace easily, have low volume themselves, or have small capacitances), while instruments that measure the integral of efforts, such things as flux linkage or momentum, must have low mass or inductance.

The operating point of a pair of coupled systems is at the intersection of their input and output characteristics in the power or energy plane. If one of these, for example the source or output characteristic, exists in the power plane, that is, is static, but the other is

energetic (i.e., dynamic: massive, inductive, capacitive, etc.), then the source characteristic must enclose the trajectory of the load characteristic at the highest frequency of interest, and ideally, the source characteristic and load trajectory should be tangent at the maximum power point or should be made tangent there by suitable choice of system parameters.

The key issue in this chapter is this: whenever two dynamic systems are connected, an interaction occurs. If the connection is to meet its objectives, the nature of this interaction must be explored and controlled.

REFERENCES

- Bell AC, Ramalingam S. Design and application of a tensile testing stage for the SEM. *Journal of Engineering Materials and Technology* 1974;96:157–162.
- Firestone FA. A new analogy between mechanical and electrical systems. *Journal of the Acoustical Society of American* 1932/1933;4:249–267.
- Paynter HM. *Analysis and Design of Engineering Systems*. Cambridge (MA): MIT Press; 1960.

48

BRIDGE TRANSDUCERS

PATRICK L. WALTER

- 48.1 Terminology
- 48.2 Flexural devices in measurement systems
 - 48.2.1 Cantilever beams
 - 48.2.2 Bourdon tubes
 - 48.2.3 Clamped diaphragms
 - 48.2.4 Error contributions from the flexure properties
- 48.3 The resistance strain gage
 - 48.3.1 Strain gage types and fabrication
 - 48.3.2 Gage factor
 - 48.3.3 Mechanical aspects of gage operation
 - 48.3.4 Electrical aspects of gage operation
 - 48.3.5 Technical societies and strain gage manufacturers
- 48.4 The Wheatstone bridge
 - 48.4.1 Bridge equations
 - 48.4.2 Lead wire effects
 - 48.4.3 Temperature compensation
- 48.5 Resistance bridge balance methods
- 48.6 Resistance bridge transducer measurement system calibration
 - 48.6.1 Static calibration
 - 48.6.2 Dynamic calibration
 - 48.6.3 Electrical substitution techniques
- 48.7 Resistance bridge transducer measurement system considerations
 - 48.7.1 Bridge excitation
 - 48.7.2 Signal amplification
 - 48.7.3 Slip rings
 - 48.7.4 Noise considerations
- 48.8 AC impedance bridge transducers
 - 48.8.1 Inductive bridges
 - 48.8.2 Capacitive bridges

References

Further readings

48.1 TERMINOLOGY

A telemetry system responding to a measurand consists of four basic parts—the transducer, the transmitting system, the receiving system, and the data output or display system:

Telemetry: The transmission of information about a measurand.

Measurand: The object of a measurement. The process to be defined.

Transducer: A component in the telemetry system which provides information about a process and, as a by-product, transfers energy from the process. Typical bridge transducers convert physical quantities such as force, pressure, displacement, velocity, acceleration, temperature, and humidity into electrical quantities for input to the transmitting system.

Transmitting System: The transmitting system typically consists of some or all of the following devices: cable, amplifier, subcarrier oscillator, filter, analog-to-digital (A/D) converter, transmitter, and antenna.

Receiving System: The receiving system typically consists of some or all of the following devices: antenna, preamplifier, multicoupler, receiver, tape or disc recorder, discriminator, decommutator, digital-to-analog (D/A) converter, and output filter.

Data Output and Display System: The data output and display system typically consist of some or all of the following devices: oscilloscope, analog meter, digital meter, graphic display, and digital printer. Either these devices may be connected directly to the output of the receiving system or a computer may process the data from the receiving system before display.

48.2 FLEXURAL DEVICES IN MEASUREMENT SYSTEMS

Bridge transducers depend on a measurand to directly modify some electrical or magnetic property of a conductive element. For example, the thermal coefficient of impedance can result in a change in impedance of a conductive element proportional to temperature (e.g., resistance thermometer). Similarly, hygroscopic materials can have their impedance change in a deterministic fashion due to humidity (e.g., humidity sensor). Most bridge transducers, however, depend on the displacement of a flexure to vary the impedance of a conductive element, resulting in an electrical signal proportional to the measurand. Advantage is taken of either the strain pattern on the surface of the flexure or the motion of this surface. Among the gamut of flexure elements associated with bridge transducers are cantilever beams, Bourdon tubes, and clamped diaphragms.

48.2.1 Cantilever Beams

Cantilever beams are routinely designed into bridge transducers. Strain near the clamped end of the beam can be correlated to displacement of the free beam end, force or torque applied to the free beam end, dynamic pressure associated with fluid flow acting over the beam surface, and so on. The compliance of a cantilever beam is defined as

$$\frac{y}{F} = \frac{L^3}{3EI} \quad (48.1)$$

where y is deflection of the beam free end, F is the force applied to this end, L is the beam length, E is the modulus of elasticity of the beam material, and I is the beam area moment of inertia. A compliant flexure will result in a bridge transducer with a large electrical signal output. Equation (48.1) indicates a compliant flexure design can be achieved by a long, thin, and narrow beam of low-modulus material. The penalty attached to such a design in application could be a transducer which is bulky, displays undesirable response to physical inputs orthogonal to its sensing direction, and has poor dynamic response.

48.2.2 Bourdon Tubes

Bourdon tubes are one of the most widely used flexures for sensing pressure. The original patent for this device was granted to Eugene Bourdon in 1852. Bourdon tubes are hollow tubes that are twisted or curved along their length. The application of pressure deforms the tube wall which, depending on tube shape, causes it to untwist or unwind. Motion of the tube is typically used to modify the alternating current (AC) impedance of bridge transducers. Bourdon tubes can be integrated into transducers to achieve extremely high accuracies and have been manufactured from perfectly elastic materials such as quartz. Transducers employing Bourdon tubes tend to be physically large and easily damaged by environmental inputs such as acceleration. In addition, the tubes themselves afford poor frequency response to time-varying pressure.

48.2.3 Clamped Diaphragms

Clamped diaphragms are another flexure used to transform a measurand into a strain or displacement proportional to applied pressure. A small, flat, circular diaphragm can be made simply, and it can be placed flush against surfaces whose flow dynamics are being studied. This type of diaphragm is typically designed to deflect in according with theory associated with clamped circular plates. Corrugated diaphragms provide extensibility over a greater linear operating range than do flat diaphragms. A catenary diaphragm consists of a flexurally weak seal diaphragm bearing against a thin cylinder whose motion is measured. The compliance of a flat, clamped circular diaphragm is defined as

$$\frac{y}{P} = \frac{3R_0^4(1 - \nu^2)}{16t^3E} \quad (48.2)$$

where y is the deflection of the center of the diaphragm, P is the applied pressure, R_0 is the diaphragm radius, ν is Poisson's ratio, t is the diaphragm thickness, and E is the modulus of elasticity of the diaphragm material. Somewhat analogous to the cantilever beam, a compliant diaphragm will have a large radius, be thin, and be made of a low-modulus material. Equation (48.2) holds for deflections no greater than t .

Figure 48.1 shows the radial and tangential strain distribution in a flat, clamped, circular diaphragm. The radial and tangential strains at the center of the diaphragm are identical. The tangential strain decreases to zero at the periphery while the radial strain becomes negative. Figure 48.2 describes a strain gage pattern designed to take advantage of this strain distribution. The central sensing elements measure the higher tangential strain while the radial sensing elements measure the high radial strains near the periphery. Resistance strain gages are discussed beginning in Section 48.3.

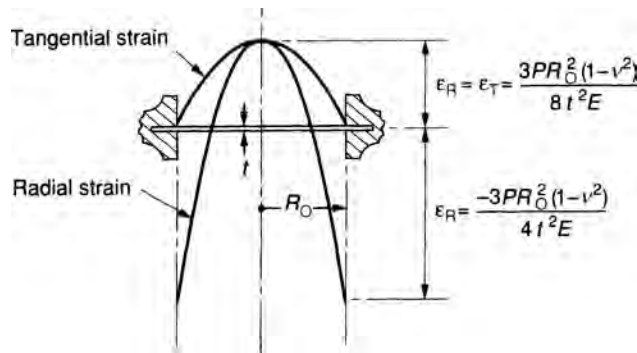


FIGURE 48.1 Radial and tangential strain distribution in a flat, clamped, circular diaphragm. (Courtesy of Measurements Group, Inc., Raleigh, NC.)

48.2.4 Error Contributions from the Flexure Properties

When flexures are designed for bridge transducers, the final transducer may have to possess an accuracy over its operating temperature range of from a few to a fractional percent. Knowledge of the inelasticities and metallurgical behaviors of flexural elements must be considered in transducer design. Metals under a constant load experience a minute deformation with time, called creep. Differences between the loading and the unloading curve of a flexure, due to energy absorbed by the material as internal friction, introduce another effect, known as hysteresis. The modulus of elasticity of materials changes with temperature. Corrosion resistance, machinability, magnetics, fatigue effects,

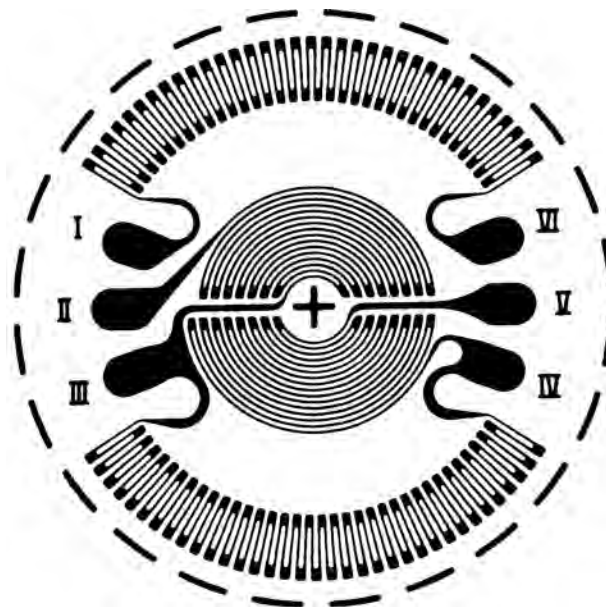


FIGURE 48.2 Micromeasurements' "JB" pattern strain gage for circular diaphragm pressure transducers. (Courtesy of Measurements Group, Inc., Raleigh, NC.)

thermal conductivity, and thermal expansion are other properties of flexural materials to consider in design application. The 300 series stainless steels are useful flexural materials due to their corrosion resistance, desirable low-temperature properties, and good creep properties at elevated temperatures. Inconel is a good flexural material in corrosive salt-water environments. These materials and others are discussed in an extremely good article on transducer flexures in Chapter 11 of Stein (1964).

48.3 THE RESISTANCE STRAIN GAGE

Strain gages are used to measure the strain pattern on the surface of the flexure in bridge transducers. In 1938, Simmons, at the California Institute of Technology, and Ruge, at the Massachusetts Institute of Technology, discovered independently that fine wire bonded directly to a surface being studied would respond to surface strain. Ruge's original gage was made by unwinding a constantan wire-wound vitrified resistor, gluing a portion of this wire with Duco cement to a celluloid bar, attaching brass shim stock as terminals, and interfacing the completed assembly to a galvanometer. The first strain gage manufacturer established in the United States was the Baldwin Lima Hamilton (BLH) Corporation (now part of the Vishay Measurements Group). BLH gages operating on the principle discovered by Simmons and Ruge were designated the SR-4 gage to include the initials of both men. The evolution of the bonded resistance wire strain gage occurred during the early 1940s. The first practical bridge transducer load cell was built by Baldwin Lima Hamilton in 1941. By the mid-1950s, Baldwin Lima Hamilton remained the only major strain gage manufacturer, and the foil strain gage was beginning to appear. Subsequent work by W. P. Mason and R. N. Thurston (reported in the *Journal of the Acoustical Society of America*, vol. 29, 1957) resulted in the introduction of the commercial semiconductor strain gage. Continued maturation of the bonded resistance strain gage has enabled a transducer industry centered around this technology to develop.

Other sources provide the derivation of all the equations dealing with ensuing topics in this chapter (Stein, 1953; Perry and Lissner, 1962; Dove and Adams, 1964; Window and Hollister, 1982).

48.3.1 Strain Gage Types and Fabrication

Paper-backed wire strain gages typically consist of a grid of resistance wire to which a paper backing has been attached with nitrocellulose cement. The wire is manufactured by drawing the selected alloy through progressive forming dies. To protect it during handling and assembly, the wire is usually sandwiched between two thin layers of paper. Typical grid wires are 0.02 mm in diameter.

Foil strain gages are essentially small printed circuits. Artwork for a master gage pattern is first prepared. This pattern is then photographically reduced, and multiple images are placed on a photographic plate. A sheet of foil (typically 0.003–0.005 mm thick) of the appropriate alloy has a light-sensitive emulsion applied, is exposed to the photographic plate, and then undergoes a development process. Chemical etching removes all but the grid material. The resultant grid cross section is square as opposed to round for wire. Advantages inherent in foil gages include better strain transmission due to improved bonding of the grid to the backing, a better thermal path for dissipation of electrically

generated heat, and a grid that can more readily be configured to minimize sensitivity to transverse strains.

The total combination of wire and foil gages span grid lengths from 0.2 mm to more than 250 mm. Foil gages satisfy the smaller of these requirements. Both wire and foil gages have associated with them a variety of ohmic values, such as 120, 175, 350, 1000, and so on. Standard values which have historically evolved are 120 and 350 Ω . These values are carry-overs from impedance-matching requirements for galvanometers which were formally used for strain recording. Figure 48.3 displays numerous configurations of wire and foil strain gages.

In the manufacture of bridge transducers using metallic strain gages, vacuum deposition of the gages is an alternate technique to bonding individual gages to the transducer flexure. The flexure is coated with aluminum oxide and then the metal gages are selectively deposited. This process yields the closest match of thermal and electrical characteristics for each bridge element.

The manufacture of semiconductor strain gages starts with a single high-purity silicon crystal. Atoms such as phosphorus (n type) or boron (p type) are doped into the material to lower its resistivity. The parent crystal is sliced into wafers before the dopant is added in a furnace at high temperature ($>1000^{\circ}\text{C}$). The wafer is masked and etched to produce a suitable grid pattern (usually either a straight element or U-shape). This grid can remain unbacked or can be mounted on a suitable carrier.

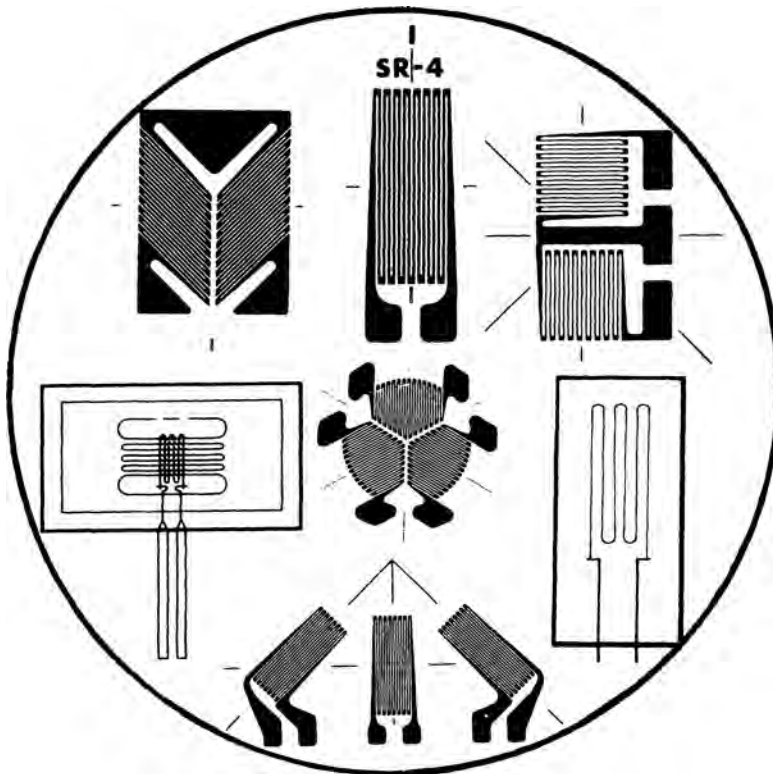


FIGURE 48.3 Numerous configurations of wire and foil strain gages—not to scale. (Courtesy of BLH Electronics, now Vishay Measurements Group.)

Alternately, in diffused semiconductor transducers, the transducer flexure itself may be made of silicon. Its surface can be passivated, etched, and doped to form gage elements integral to the flexure. Similar to vacuum deposition of metal gages, diffused semiconductor transducers offer a more nearly optimum match of thermal and electrical properties for each bridge element. Problems with slippage associated with the bond of a gage carrier or backing are nonexistent.

Bridge transducers using semiconductor gages typically possess poorer thermal and linearity specifications than those using metal gages. However, the sensitivity of semiconductor gages to surface strain is much greater than metal gages. This allows them to be used in transducers providing more signal output (typically 100–500 mV versus 30 mV) and on stiffer flexures, resulting in smaller transducer size, higher frequency response, and increased ruggedness. Although not strictly correct, by convention, it has become equivalent to refer to semiconductor-based transducers as either piezoresistive or solid-state transducers.

48.3.2 Gage Factor

The gage factor F for a strain gage is defined as

$$F = \frac{\Delta R}{R\epsilon} \quad (48.3)$$

where R is resistance and ϵ is strain equal to $\Delta l/l$ (Δl is the change in length of l_0). Equation (48.3) may be redefined as

$$F = 1 + 2\nu + \frac{d\rho}{\rho\epsilon} \quad (48.4)$$

where ν is Poisson's ratio and ρ is the resistivity of the grid material. Most metal gages have a gage factor between 2 and 4.5. For strain gages made from a semiconductor material, the change in resistivity with applied stress is the dominant factor and values as high as 170 are possible. Table 48.1 lists properties and gage factors for various grid materials.

TABLE 48.1 Grid Material Composition, Trade Name, Properties, and Gage Factor

Composition	Trade Name	Properties	Gage Factor
1. Copper–nickel (57%–43%)	Constantan	Strain sensitivity relatively independent of level and temperature; used to 200°C; high resistivity applicable to small grids; measures strains to 20% in annealed form	2.0
2. Nickel–chromium–iron–molybdenum (36%–8%–55.5%–0.5%)	Isoelastic	High gage factor; high fatigue life; used to 200°C; high temperature coefficient of resistance; nonlinear at strain levels above 5%	3.5
3. Nickel–chromium (80%–20%), nickel–chromium (75%–20%) plus iron and aluminum	Nichrome V Karma	Good fatigue life; stable; high resistance applicable to small grids; used to 400°C	2.2

Note: Nickel alloy gages are susceptible to magnetic fields.

48.3.3 Mechanical Aspects of Gage Operation

To build effective bridge transducers, one must be aware of the interaction between the gage and the surface of the transducer flexure to which it is mounted. Mechanical aspects of this interaction include the influences of temperature, backing material, size, orientation, transverse sensitivity, distance from the surface, bonding and installation, and gage frequency response.

48.3.3.1 Temperature A qualitative discussion of temperature effects on bonded strain gages indicates the effects to be attributable to three principal causes: (1) The transducer flexure to which the gage is attached expands or contracts, (2) the strain gage resistance changes with temperature, and (3) the strain gage grid expands or contracts. With some gages (particularly semiconductors), the change of gage factor with temperature is also extremely significant. These temperature effects are accounted for by temperature-strain calibration; self-temperature-compensated gages where combined effects 1, 2, and 3 above are minimized over a given temperature range for a given combination of grid and flexure material; and a dummy gage integrated into a bridge circuit (discussed later) to electrically subtract temperature-induced strain.

48.3.3.2 Backing Material The purpose of the backing material used in constructing strain gages is to provide support, dimensional stability, and mechanical protection for the grid element. The backing material of the gage element(s) acts as a spring in parallel with the flexure to which it is attached and can potentially modify flexure mechanical behavior. In addition, the temperature operating range of the gage can be constrained by its backing material. Most backings are epoxies or glass fiber-reinforced epoxies. Some gages are encapsulated for chemical and mechanical protection as well as extended fatigue life. For high-temperature applications, some gages have strippable backings for mounting with ceramic adhesives. Still other metal gages can be welded. The frequency response of welded gages, due to uncertainties in dynamic response, is a subject area still requiring investigation.

48.3.3.3 Size The major factors to be considered in determining the size of strain gage to use are available space for gage mounting, strain gradient at the test location, and character of the material under test. The strain gage must be small enough to be compatible with mounting location and concentrated strain field. It must be large enough so that, on metals with large grain size, it measures average strain as opposed to local effects. Grid elements greater than 3 mm generally have greater fatigue resistance.

48.3.3.4 Transverse Sensitivity and Orientation Strain gage transverse sensitivity and mounting orientation are concurrent considerations. Transverse sensitivity in strain gages is important due to the fact that part of the geometry of the gage grid is oriented in directions other than parallel to the principal gage sensing direction. Values of transverse sensitivities are provided with individual gages but typically vary between fractional and several percent. The position of the strain gage axis relative to the numerically larger principal strain on the surface to which it is mounted will have an influence on indicated strain.

48.3.3.5 Distance from the Surface The grid element of a strain gage is separated from the transducer flexure by its backing material and cement. The grid then responds to strain at a location removed from the flexure surface. The strain on flexures such as thin plates in bending can vary considerably from that measured by the strain gage.

48.3.3.6 Bonding Adhesives Resistance strain gage performance is entirely dependent on the bond attaching it to the transducer flexure. The grid element must have the strain transmitted to it undiminished by the bonding adhesive. The elimination of this bond is one of the principal advantages of vacuum-deposited metallic and diffused semiconductor bridge transducers. Typical adhesives are as follows:

Epoxy Adhesives: Epoxy adhesives are useful over a temperature range of -270 to $+320^{\circ}\text{C}$. The two classes are either room-temperature curing or thermal setting type; both are available with various organic fillers to optimize performance for individual test requirements.

Phenolic Adhesives: Bakelite, or phenolic adhesive, requires high bonding pressure and long curing cycles. It is used in some transducer applications because of long-term stability under load. The maximum operating temperature for static loads is 180°C .

Polyimide Adhesives: Polyimide adhesives are used to install gages backed by polyimide carriers or high-temperature epoxies. They are a one-part thermal setting resin and are used from -200 to $+400^{\circ}\text{C}$.

Ceramic cements (applicable from -270 to $+550^{\circ}\text{C}$) and welding are other mounting techniques.

48.3.3.7 Frequency Response The frequency response of bridge transducers cannot be addressed without considering the frequency response of the strain gage as well. It is assumed that the transducer is used in such a manner that mounting variables do not influence its frequency response.

Piping in front of pressure transducer diaphragms and mounting blocks under accelerometers are two examples of variables which can violate this assumption. Transducers, particularly those which measure force, pressure, and acceleration, typically are dynamically modeled as single-degree-of-freedom systems characterized by a linear second-order differential equation with constant mass, damping, and stiffness coefficients. In reality, transducers possess multiple resonant frequencies associated with their flexure and their case. Figure 48.4 presents the actual frequency response of a bridge-type accelerometer. The response indicates this single-degree-of-freedom model to be adequate through the first major transducer resonance. Such devices have a frequency response usable (constant within 4% referenced to their AC response) to one-fifth of the value of this major resonance. The strain gage itself acts as a spatial averaging device whose frequency response is a function of both its gage length and the sound velocity of the material on which it is mounted. Walter (1980) discusses this relationship from which Figure 48.5 is extracted. Figure 48.5 contains curves for three different length gages. Its abscissa must be multiplied by a specific sound velocity. For most bridge transducers, the structural resonance of the flexure constrains its frequency response.

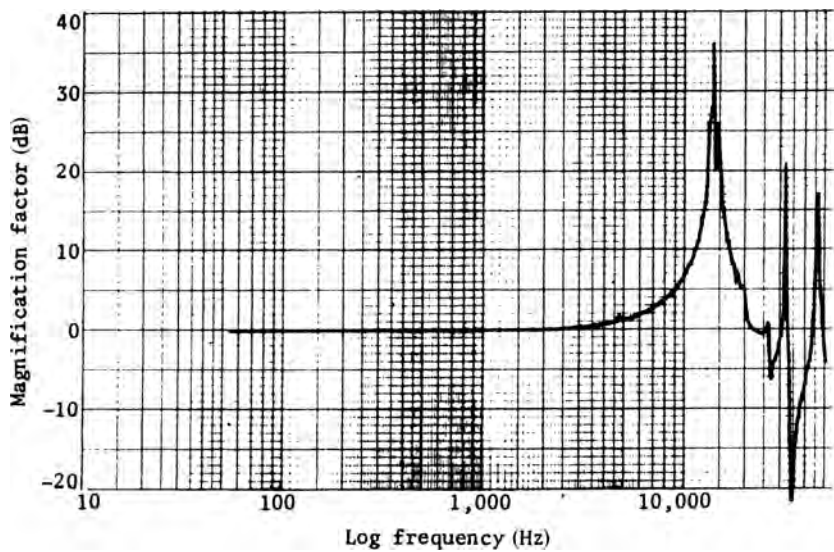


FIGURE 48.4 Magnitude of transfer function of piezoresistive accelerometer.

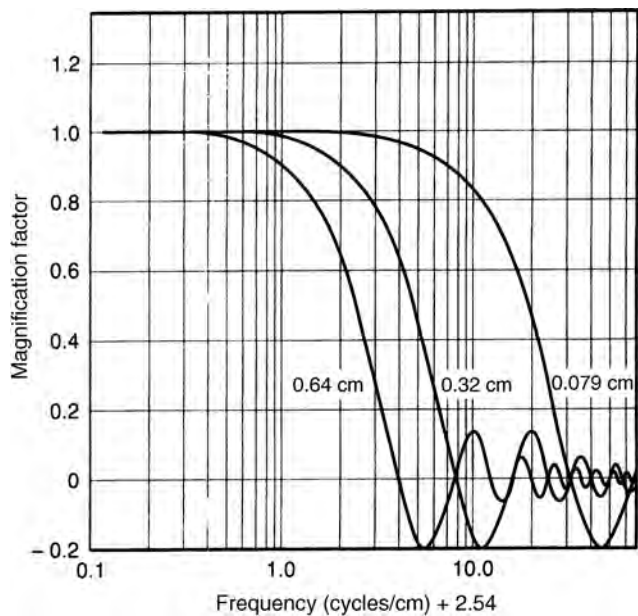


FIGURE 48.5 Transfer function for strain gages of varying lengths when analyzed as spatial averaging transducers. (Reprinted with permission from *ISA Transactions*, Vol. 19, Issue 3. Copyright 1980 by the Instrument Society of America.)

48.3.4 Electrical Aspects of Gage Operation

The resistance strain gage, which manifests a change in resistance proportional to strain, must form part of an electrical circuit such that a current passed through the gage transforms this change in resistance into a current, voltage, or power change to be measured. The electrical aspects of gage operation to be considered include current in the gage, resistance to ground, and shielding.

Strain gages are seldom damaged by excitation voltages in excess of proper values but performance degrades. The voltage applied to a strain gage bridge creates a power loss in each arm, which must be dissipated in the form of heat. By its basic design, all of the power input to the bridge is dissipated in the bridge with none available to the output circuit. The sensing grid of every strain gage then operates at a higher temperature than the transducer flexure to which it is bonded. The heat generated within the gage must be transferred by conduction to the flexure. Heat flow into the flexure causes a temperature rise, which is a function of its heat sink capacity and gage power level. The optimum excitation level for strain gage applications is a function of the strain gage grid area, gage resistance, heat sink properties of the mounting surface, environmental operating temperature range of the gage installation, required operational specifications, and installation and wiring techniques. Rigid operating requirements for precision transducers require performance verification of the optimum excitation level. Zero shift versus load and stability under load at the maximum operating temperature are the performance tests most sensitive to excessive excitation voltage.

Table 48.2 and Figures 48.6–48.8 allow a first approximation at optimizing bridge excitation levels. Table 48.2 defines the suitability of various structural materials for providing an adequate heat sink for gage mounting dependent on both accuracy requirements and static or dynamic measurements. Figures 48.6–48.8 define the recommended excitation voltage for specific gages as a function of the power density capability of the heat sink and gage grid area.

Resistance to ground is an important parameter in strain gage mounting since insulation leakage paths produce shunting of the gage resistance between the gage and the metal structure to which it is bonded, producing false compressive strain readings. The ingress of fluids typically leads to this breakdown in resistance-to-ground value and can also change the mechanical properties of the adhesive. A minimum gage-to-mounting-surface resistance-to-ground value of 50 M Ω is recommended.

Since signals of interest from strain gage bridges are typically on the order of a few millivolts, shielding of the bridge from stray pickup is important. Gage leads should also be shielded and proper grounding procedures followed. Stray pickup may be introduced by 60-Hz line voltage associated with other electronic equipment, electrical noise from motors, radio frequency interference, and so on. Note that shielding materials for electrical fields are different from those for magnetic fields. Nickel alloy strain gages are particularly susceptible to magnetic fields.

48.3.5 Technical Societies and Strain Gage Manufacturers

In concluding a discussion of the resistance strain gage, it is appropriate to identify some of the technical societies dealing with strain gages and some of the manufacturers of strain gages. In 1956, to accelerate the development of the resistance strain

TABLE 48.2 Suitability of Various Materials as Heat Sink for Strain Gage Mounting

Accuracy Requirements	Excellent, Heavy Aluminum or Copper Specimen	Good, Thick Steel	Fair, Thin Stainless Steel or Titanium	Poor, Filled Plastic Such as Fiberglass/Epoxy	Very Poor, Unfilled Plastic Such as Acrylic or Polystyrene
<i>Static</i>					
High	2-5 <i>3.1-7.8</i>	1-2 <i>1.6-3.1</i>	0.5-1 <i>0.78-1.6</i>	0.1-0.2 <i>0.16-0.31</i>	0.01-0.02 <i>0.016-0.031</i>
Moderate	5-10 <i>7.8-16</i>	2-5 <i>3.1-7.8</i>	1-2 <i>1.6-3.1</i>	0.2-0.5 <i>0.31-0.78</i>	0.02-0.05 <i>0.031-0.078</i>
Low	10-20 <i>16-31</i>	5-10 <i>7.8-16</i>	2-5 <i>3.1-7.8</i>	0.5-1 <i>0.78-1.6</i>	0.05-0.1 <i>0.078-0.16</i>
<i>Dynamic</i>					
High	5-10 <i>7.8-16</i>	5-10 <i>7.8-16</i>	2-5 <i>3.1-7.8</i>	0.5-1 <i>0.78-1.6</i>	0.01-0.05 <i>0.016-0.078</i>
Moderate	10-20 <i>16-31</i>	10-20 <i>16-31</i>	5-10 <i>7.8-16</i>	1-2 <i>1.6-3.1</i>	0.05-0.2 <i>0.078-0.31</i>
Low	20-50 <i>31-78</i>	20-50 <i>31-78</i>	10-20 <i>16-31</i>	2-5 <i>3.1-7.8</i>	0.2-0.5 <i>0.31-0.78</i>

Note: Units are W/in.² on top, kW/m² in italics underneath.
Source: Courtesy of Measurement Group, Inc., Raleigh, NC.

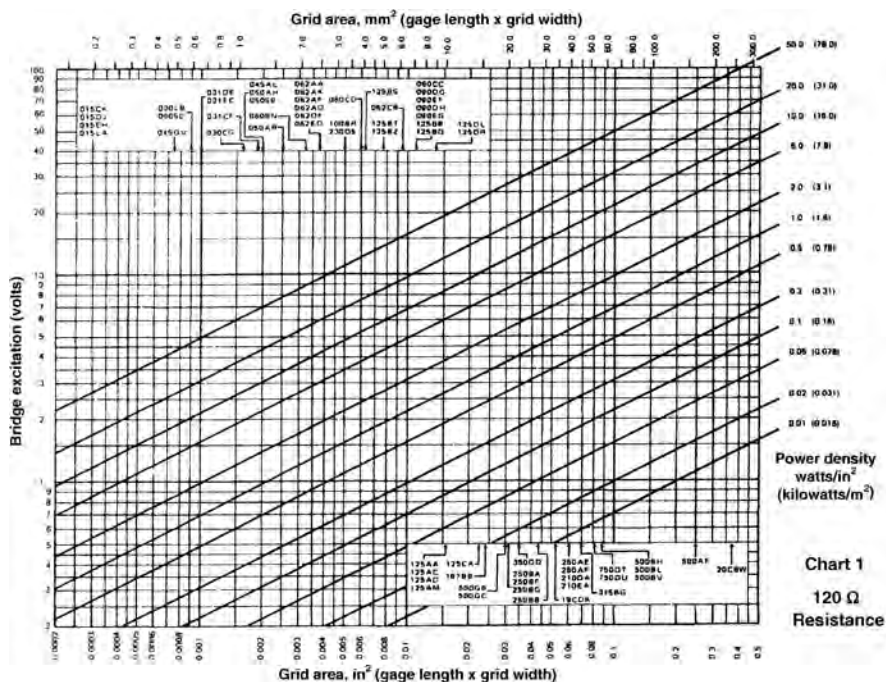


FIGURE 48.6 Bridge excitation versus grid area for various power densities and 120-Ω gages. (Courtesy of Measurements Group, Inc., Raleigh, NC.)

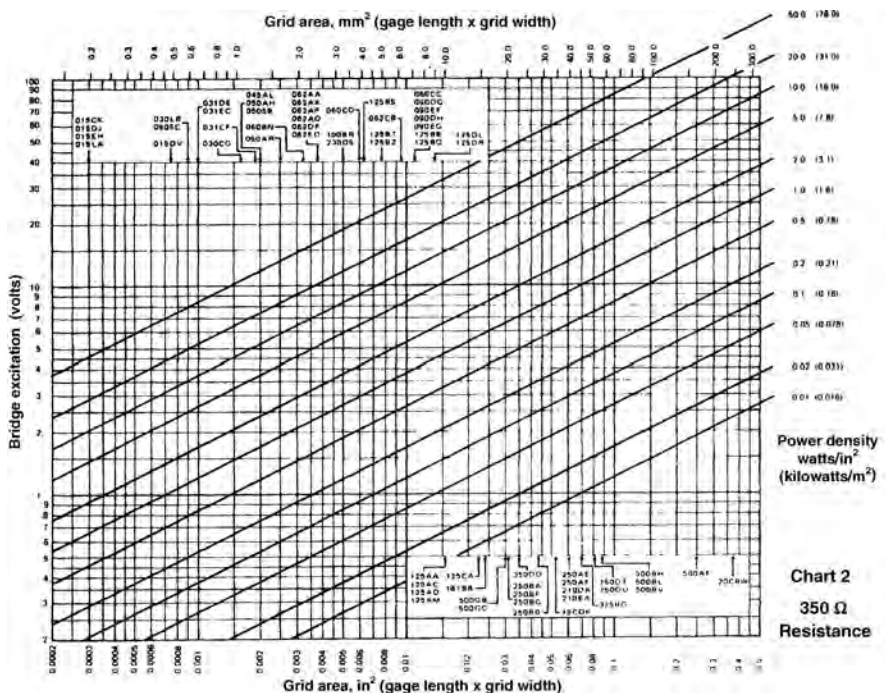


FIGURE 48.7 Bridge excitation versus grid area for various power densities and 350-Ω gages. (Courtesy of Measurements Group, Inc., Raleigh, NC.)

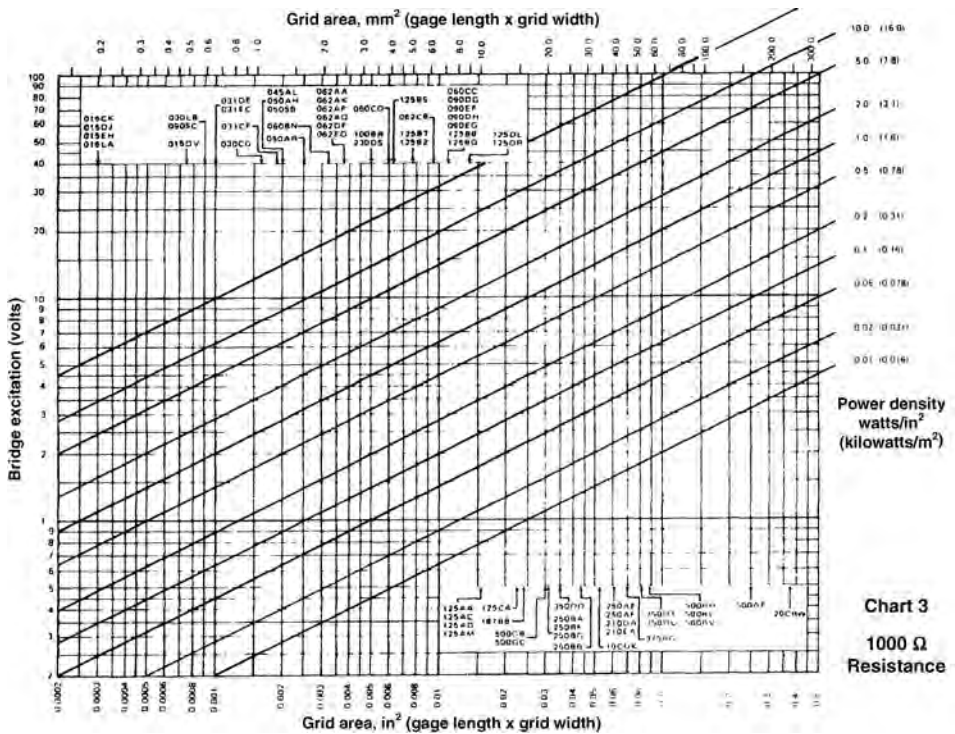


FIGURE 48.8 Bridge excitation versus grid area for various power densities and 1000- Ω gages. (Courtesy of Measurements Group, Inc., Raleigh, NC.)

gage, BLH Electronics established a users' group to accomplish this purpose and to further the state of the art in strain gage technology in general. This users' group was formed primarily of various aircraft companies in the western United States and is entitled the Western Regional Strain Gage Committee (WRSGC). For 15 years, the WRSGC was an autonomous organization financed by BLH Electronics. Since 1971, WRSGC has operated under the auspices of the Technical Committee on Strain Gages (TCSG) of the Society for Experimental Mechanics (SEM). The SEM is the premier organization in the United States involved with strain gages and experimental mechanics in general. The SEM (formerly Society for Experimental Stress Analysis) was founded by William Murray at the Massachusetts Institute of Technology. Publications of this society include *Experimental Mechanics* and *Experimental Techniques*. A similar European organization is the Joint British Committee for Stress Analysis, whose publication is *The Journal of Strain Analysis for Engineering Design*.

By 2004, essentially the entire strain gage manufacturing capability in the United States was consolidated by Vishay Measurements Group, corporate headquarters at Vishay Intertechnology, Inc., 63 Lincoln Highway, Malvern, Pennsylvania 19355-2120. Its associated website has an extensive array of in-depth articles on strain gage technology.

48.4 THE WHEATSTONE BRIDGE

Best transducer performance can be achieved by minimizing the strain level in the transducer flexure. Lower strains allow increased safety without mechanical overload protection. Effective overload stops are usually troublesome to design and an added expense to make. Reduced strain levels almost always produce an improvement in linearity accompanied by a reduction in the hysteresis originating in the transducer flexure material.

Small strains result in small impedance changes in resistive strain gage elements. Electromechanical transducers use a Wheatstone bridge circuit to detect a small change in impedance to a high degree of accuracy.

48.4.1 Bridge Equations

The circuit most often used with strain gages is a four-arm bridge with a constant-voltage power supply. Figure 48.9 shows a basic bridge configuration. The supply voltage E_{ex} can be either AC or AC, but for now we assume it is AC so equations can be written in terms of resistance R rather than a complex impedance. The condition for a balanced bridge with e_0 equal to zero is

$$\frac{R_1}{R_2} = \frac{R_4}{R_3} \quad (48.5)$$

Next, an expression is presented for e_0 due to *small* changes in R_1 , R_2 , R_3 , and R_4 :

$$e_0 = \left[-\frac{R_3 dR_4}{(R_3 + R_4)^2} + \frac{R_4 dR_3}{(R_3 + R_4)^2} - \frac{R_1 dR_2}{(R_1 + R_2)^2} + \frac{R_2 dR_1}{(R_1 + R_2)^2} \right] E_{\text{ex}} \quad (48.6)$$

In many cases, the bridge circuit is made up of equal resistances. Substituting for individual resistances, a strain gage resistance R , and using the definition of the gage factor from Equation (48.3), Equation (48.6) becomes

$$e_0 = \frac{FE_{\text{ex}}}{4} (-\epsilon_4 + \epsilon_3 - \epsilon_2 + \epsilon_1) \quad (48.7)$$

The unbalance of the bridge is seen to be proportional to the sum of the strain (or resistance changes) in opposite arms and to the difference of strain (or resistance changes) in adjacent arms.

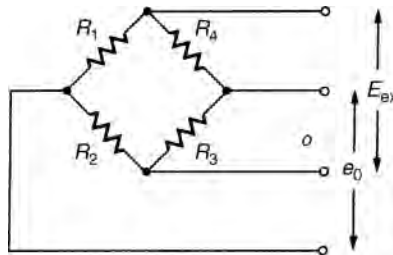


FIGURE 48.9 Four-arm bridge with constant-voltage (E_{ex}) power supply.

Equations (48.6) and (48.7) indicate one technique to compensate strain gage circuits to minimize the influence of temperature-induced strain. This was referred to in Section 48.3 as the dummy gage method.

Assume that we have a bridge circuit with one active arm and arbitrarily let this arm be number 4. Equation (48.7) becomes

$$e_0 = \frac{FE_{\text{ex}}}{4}(-\epsilon_4) \quad (48.8)$$

Arm 4 responds to the total strain induced in it, which is comprised of both thermal (t) and mechanical (m) strain

$$\epsilon_4 = \epsilon_m + \epsilon_t \quad (48.9)$$

A problem arises if it is desired to isolate the mechanical strain component. One solution is to take another strain gage (the dummy gage) and mount it on a strain-isolated piece of the same material as that on which gage 4 is mounted. If placed in the same thermal environment as gage 4, the output from the dummy gage becomes simply ϵ_t . If the dummy gage is wired in an adjacent bridge arm to 4 (1 or 3), Equation (48.7) becomes

$$e_0 = \frac{FE_{\text{ex}}}{4}(-\epsilon_m - \epsilon_t + \epsilon_t) \quad (48.10)$$

Equation (48.10) indicates that thermal strain effects are canceled. Similarly, in Figure 48.2, four gages were shown mounted on a transducer diaphragm. Equation (48.7) indicates that thermal strain effects from this circuit should be canceled.

In reality, perfect temperature compensation is not achieved since no two strain gages from a lot track one another identically. However, compensation adequate for many applications can be accomplished.

The biggest thermal problem with bridge transducers occurs in transient situations, such as explosive or combustion environments. Here, due to individual physical locations, gages in a bridge are not in the same time-varying temperature, and compensation cannot be achieved. The only technique which can be used in this situation is either to cool the transducer by circulating water or gas around it or to delay the thermal transient until the measurement is complete.

The alternating signs in Equation (48.7) are useful in isolating various strain components when using bridge circuits containing strain gages. Figure 48.10 shows a beam flexure used in an accelerometer. Four gages are mounted on the beam—two on the top and two on the bottom. Notches are placed in the beam to intensify the strain field under the gages. Due to symmetry, the tension gages see the same strain as do the compression gages. If the tension gages occupy two adjacent arms of the bridge and the compression gages the other two, Equation (48.7) indicates that the net bridge output will be zero. However, if the tension and compression gages are in opposite arms, Equation (48.7) indicates that a bending strain signal four times that of an individual gage will be produced with temperature compensation also achieved.

Equation (48.6) presented the generalized form of the bridge equation for four active arms. If only one arm (e.g., arm 4) is active, this equation becomes

$$e_0 = \left[\frac{-R_3 dR_4}{(R_3 + R_4)^2} \right] E_{\text{ex}} \quad (48.11)$$

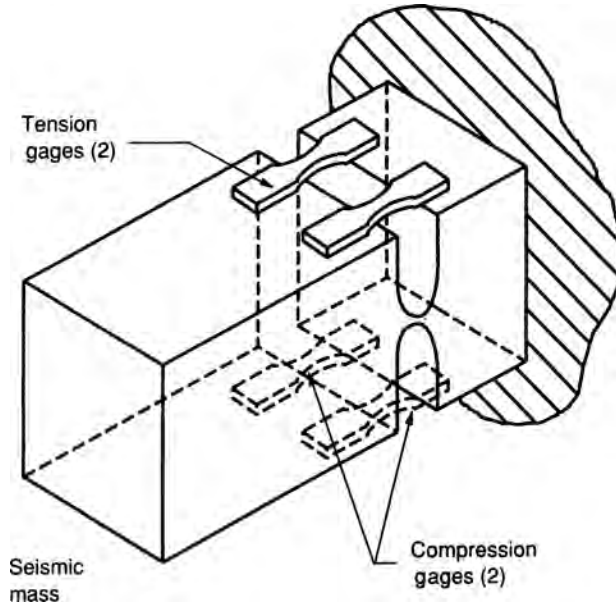


FIGURE 48.10 Strain gaged beam flexure used in accelerometer. (Courtesy of Endevco.)

This equation was specifically presented for *small* changes in resistance, such as those associated with metallic strain gages. If the change in resistance in arm 4 is large, Equation (48.11) is better expressed as

$$e_0 = \frac{(R_4 + \Delta R_4)E_{ex}}{(R_4 + \Delta R_4) + R_3} - \frac{R_4 E_{ex}}{R_4 + R_3} \quad (48.12)$$

For an equal-arm bridge, this becomes

$$e_0 = \frac{\Delta R E_{ex}}{4R + 2\Delta R} = \frac{F E_{ex} \epsilon}{4 + 2F\epsilon} \quad (48.13)$$


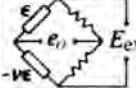

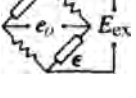
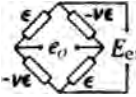
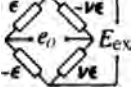
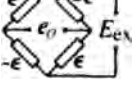
For an equal-arm bridge, Equation (48.11) becomes

$$e_0 = \frac{dR E_{ex}}{4R} = \frac{F E_{ex} \epsilon}{4} \quad (48.14)$$

The difference between Equations (48.14) and (48.13) is that Equation (48.14) describes a linear process while Equation (48.13) describes a nonlinear one. Semiconductor gages, because of their large gage factor, require analysis using Equation (48.13).

Semiconductor gages may be used in constant-voltage four-arm bridge circuits when two or four gages are used in adjacent arms and strained so that their outputs are additive. Analysis of the bridge equations for this situation will show that if gages in adjacent arms are subjected to equal but opposite values of ΔR , the output signal is doubled and the nonlinearity in the bridge output is eliminated. Another approach to eliminating this nonlinearity is to design a circuit where the current through the strain gage remains constant.

TABLE 48.3 Equations for One, Two, and Four Equal-Active-Arm Bridges

Bridge/Strain Arrangement	Description	Output Equation $-e_0/E_{ex}$ (mV/V)
	Single active gage in uniaxial tension or compression	$\frac{e_0}{E_{ex}} = \frac{F\epsilon \times 10^{-3}}{4 + 2F\epsilon \times 10^{-6}}$
	Two active gages in uniaxial stress field—one aligned with maximum principal strain, one “Poisson” gage	$\frac{e_0}{E_{ex}} = \frac{F\epsilon(1 + \nu) \times 10^{-3}}{4 + 2F\epsilon(1 - \nu) \times 10^{-6}}$
	Two active gages with equal and opposite strains—typical of bending-beam arrangement	$\frac{e_0}{E_{ex}} = \frac{F\epsilon}{2} \times 10^{-3}$
	Two active gages with equal strains of same sign—used on opposite sides of column with low temperature gradient (e.g., bending cancellation)	$\frac{e_0}{E_{ex}} = \frac{F\epsilon \times 10^{-3}}{2 + F\epsilon \times 10^{-6}}$
	Four active gages in uniaxial stress field—two aligned with maximum principal strain, two “Poisson” gages (column)	$\frac{e_0}{E_{ex}} = \frac{F\epsilon(1 + \nu) \times 10^{-3}}{2 + F\epsilon(1 - \nu) \times 10^{-6}}$
	Four active gages in uniaxial stress field—two aligned with max. principal strain, two “Poisson” gages (beam)	$\frac{e_0}{E_{ex}} = \frac{F\epsilon(1 + \nu) \times 10^{-3}}{2}$
	Four active gages with pairs subjected to equal and opposite strains (beam in bending or shaft in torsion)	$\frac{e_0}{E_{ex}} = F\epsilon \times 10^{-3}$

Source: Courtesy of Measurements Group Inc., Raleigh, NC.

Table 48.3 provides generalized bridge equations for one, two, and four equal-active-arm bridges of various configurations. The dimensionless bridge output is presented in millivots per volts for a constant-voltage power supply. Strain is presented in microstrain. No small-strain assumption is built into these equations. For large strains with semiconductor gages, F may not be a constant and this correction also has to be built into the equations. In this table, the Poisson gage is one that measures the lateral compressive strain accompanying an axial tension strain. As noted earlier, only for two adjacent active gages with equal and opposite strains or for four active gages with pairs subjected to equal and opposite strains is the bridge output a linear function of strain.

48.4.2 Lead Wire Effects

There has been a historical lack of agreement between manufacturers of strain gages as to color codes and wiring designations. This is particularly true in bridge transducers.

Figures 48.11 and 48.12 are suggested industry standards that have assisted in lessening this confusion. Figure 48.11 covers the situation where all bridge elements are remote from the power supply, whereas Figure 48.12 covers the situation where only one bridge arm is remote from the power supply. The bridge balance network and shunt calibration are discussed in Sections 48.5 and 48.6, respectively. Table 48.4 presents guidelines for multiconductor strain gage cable.

The previous discussion has assumed that the only resistive elements in the circuits are the gages themselves. Resistance of circuit lead wires also is a consideration.

One possible need for remote recording occurs when the bridge power supply and the readout instrumentation are at one location and the bridge transducer is at a remote location. In this situation, the resistance R_L of each lead wire between the bridge and the power supply or readout must be accounted for. Most readout instruments have very high input impedances, so the effect of R_L in series with them can be ignored. The significant effect of lead-wire resistance is to modify the resistance in series with the power supply from R_{bridge} to $R_{\text{bridge}} + 2R_L$. For example, a lead-wire resistance of $3\ \Omega$ and a bridge resistance of $120\ \Omega$ will produce loading effects which, if not corrected, will result in a 5% error in bridge transducer output.

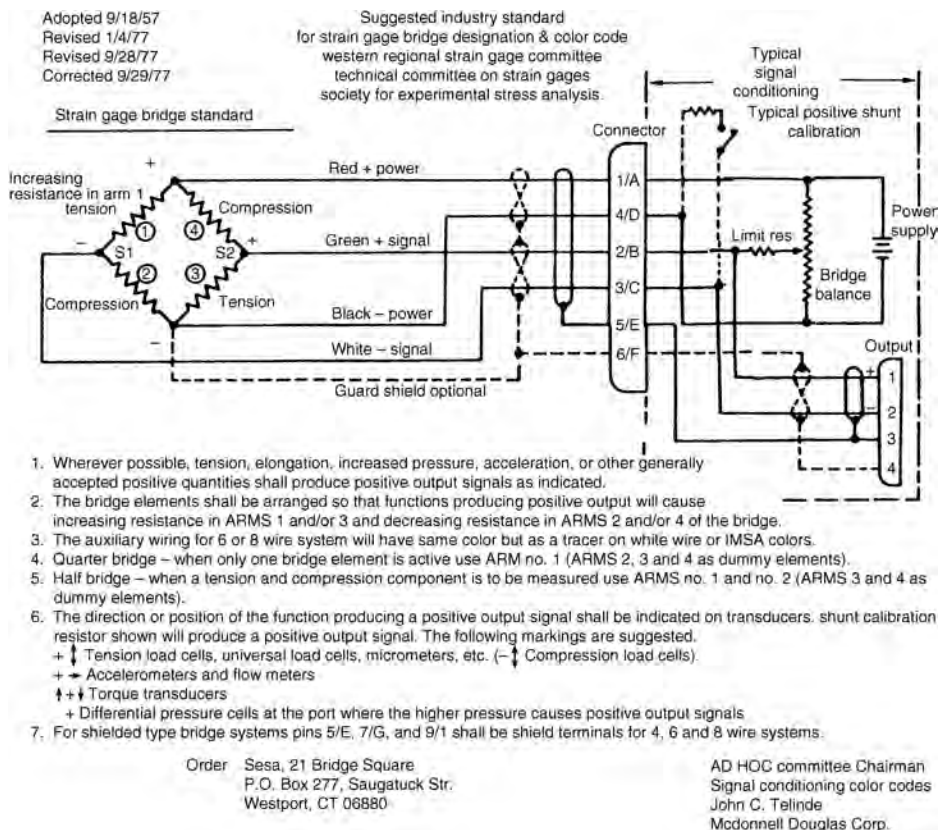


FIGURE 48.11 Color code and wiring designation, four-arm bridge. (Courtesy of Western Regional Strain Gage Committee.)

Adopted 9/17/80
Revised 3/4/81

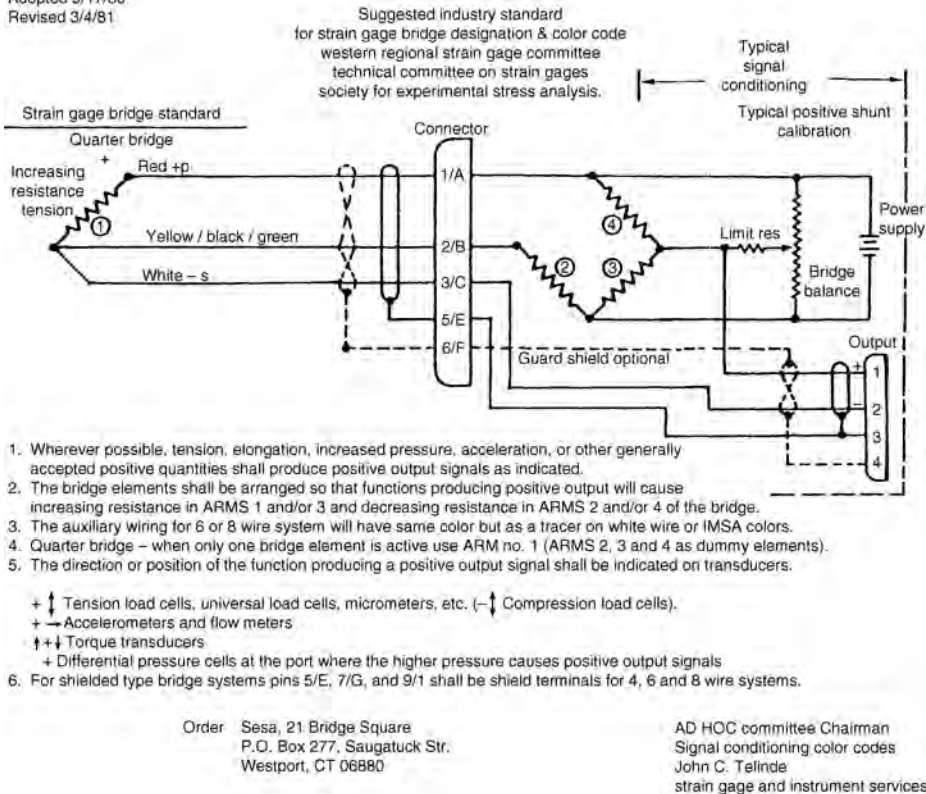


FIGURE 48.12 Color code and wiring designation, single-arm bridge. (Courtesy of Western Regional Strain Gage Committee.)

There are at least three simple techniques to eliminate this error source:

1. The bridge transducer can be calibrated with the long length of cable with which it will operate.
2. The excitation voltage E_{ex} can be measured at the bridge itself instead of at the power supply and appropriate values substituted in Equation (48.6) or equivalent versions of it.
3. The bridge voltage E_{ex} can be determined by measuring the current to the bridge (I_{bridge}) and calculating E_{ex} as the product $I_{bridge} \times R_{bridge}$.

Another possible need for remote recording occurs when two gages (either both active or one active and one for temperature compensation) are at the test site. The other two bridge completion resistors are in parallel with the power supply and located adjacent to it. In Figure 48.9, assume R_3 and R_4 are the two remote active arms. In Equation (48.6), the last two terms would be zero since these arms are not active. In this equation R_3 and R_4 would become, respectively, $R_3 + R_L$ and $R_4 + R_L$. If the strain gages in both arms are identical, Equation (48.6) reduces to

$$e_0 = \frac{FE_{ex}R}{4(R + R_L)}(-\epsilon_4 + \epsilon_3) \quad (48.15)$$

TABLE 48.4 Multiconductor Strain Gage Cable Guideline, Western Regional Strain Gage Committee

A need exists for low-millivolt signal levels to be transmitted by better quality multiple conductor cables of considerable length	
I. Conductors	Three through eight twisted, stranded conductors of tinned copper per ASTM-B-174, AWG 20-16/32, or AWG 18-16/30
II. Color code	Jacket: orange, gray, white, or black Conductors: Per ISA-S37.3, ANSI-MC6.2-1975, and WRSGC/SESA 5-6-1960
III. Insulation	Superior to the PVC materials currently in use. The dielectric material should be nonhygroscopic and approach zero water absorption and zero water permeability. Target jacket thickness of 0.016 in. or less and conductor insulation of 0.012 in. or less. Target resistance values should be constant as practical humid and wet environments and as high as possible (target value of 500 Ω per 1000 ft). The breakdown level of the dielectric materials shall be greater than 150 V AC.
IV. Construction	The cable shield shall be aluminized polyester tape with 100% coverage of all conductors. A 22-AWG drain wire shall be in intimate contact with the shield throughout the entire cable length. The cable shall have as small a diameter as practical and be flexible enough to have a bend radius less than six cable diameters. Overall cable strength sufficiently high to be pulled through conduits.

Source: Courtesy of Society for Experimental Mechanics.

Other situations can be investigated by substituting appropriate values for the resistance in each bridge arm (including lead-wire resistance) in the governing bridge equation. In addition, Section 48.6 will show that shunt calibration is one technique that can be used to compensate the system for the effects of lead-wire resistance.

48.4.3 Temperature Compensation

Before leaving the analysis of the Wheatstone bridge circuit, temperature compensation of bridge-type transducers should receive additional discussion. An ideal transducer would yield an output voltage that is a constant calibration factor times its mechanical input, independent of other environmental factors. Ambient temperature variations are one of the major error sources in precision transducers. Even when using self-temperature-compensated strain gages and taking advantage of the ability of the Wheatstone bridge circuit to subtract in the dummy gage method, some residual error remains. These remaining errors are of two types.

First, the transducer zero output can change with temperature. Unequal mechanical expansion of transducer members can cause this effect. Second, the calibration factor, span, or sensitivity also can change with temperature. This can be caused, for example, by a change in the stiffness of the transducer flexure with temperature.

The following discussion provides one compensation scheme for each type (metallic and semiconductor) of bridge transducer. Temperature Compensation of Bridge Type Transducers, (1951) and TECHNICAL DATA TD4354-1 (1975) are sources of more detailed information. An equal-arm bridge transducer operating with a constant-voltage supply is assumed. Metallic strain gages are discussed first.

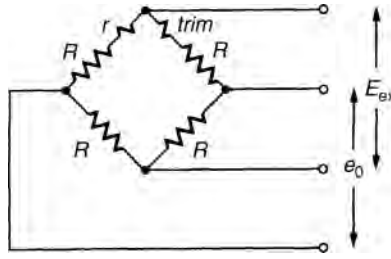


FIGURE 48.13 Transducer bridge compensation for zero shift, metal gages.

Figure 48.13 shows one scheme for compensating for transducer zero shift. A corner of the bridge is brought out to terminals, and a temperature-sensitive resistor, r , is placed in one side of the bridge. Typically, a wire resistor such as Balco, nickel, or copper with a positive temperature coefficient is used.

The transducer must first be temperature calibrated and the change in zero reading for a given temperature range determined. This can be characterized in volts of output change per volt of input. Definitions are

b = output voltage change per degree per input volt,

a = temperature coefficient of resistance of r ,

R = bridge arm resistance,

T = temperature change from reference temperature.

If the bridge supply voltage is E_{ex} and R is changed a small amount by the addition of r , the bridge output is

$$e_0 = \frac{E_{\text{ex}} r}{4R} \quad (48.16)$$

Equation (48.16) can further be expressed as

$$e_0 = \frac{E_{\text{ex}} r_0 (1 + aT)}{4R} \quad (48.17)$$

or

$$e_0 = \text{const} + E_{\text{ex}} bT \quad (48.18)$$

where r_0 is the value of r at the reference temperature. The effect of the constant term is eliminated by a temperature-insensitive trim resistor in an adjacent arm. The above equations indicate that at the reference temperature r_0 should be selected equal to $4Rb/a$. If the transducer is properly designed, b is very small compared to a , keeping the compensating resistor small in value. The compensating resistor should be located in an arm causing a voltage change of opposite sign to the zero drift with increasing temperature.

After zero shift is compensated, the calibration or span factor remains to be compensated. Most metal strain gage transducers give larger outputs with increasing temperature, so the temperature coefficient of the calibration scale factor, K , is positive. The trick in

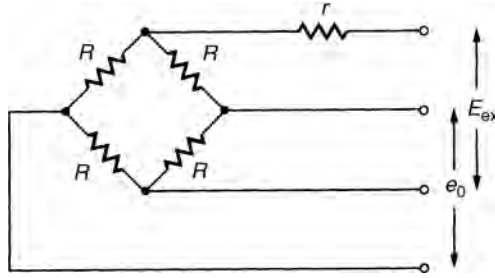


FIGURE 48.14 Transducer bridge compensation for span, metal gages.

span compensation is to hold the transducer supply voltage constant while automatically varying the bridge current, causing it to decrease with increasing temperature. In this discussion, r is identified to be a positive series resistor (Figure 48.14). Definitions are

$$r = r_0(1 + aT),$$

a = positive temperature coefficient of r ,

T = temperature difference from reference temperature,

c = temperature coefficient of the calibration factor K , so

$$K = K_0(1 + cT),$$

E_{ex} = transducer supply voltage.

The voltage on the transducer at the reference temperature is $RE_{ex}/(R + r_0)$ and at temperature T is $RE_{ex}/(R + r)$. The ratio by which it changes is $(R + r_0)/(R + r)$, which is used to correct for the variation in K . This variation is corrected for when $K_0(1 + cT)(R + r_0)/[R + r_0(1 + aT)] = \text{constant}$. The value of r_0 which satisfies this requirement can be shown to be

$$r_0 = \frac{cR}{a - c} \quad (48.19)$$

Note that in span and zero compensation as discussed thus far, the compensating resistors must be at the same temperature as the transducer. Usually, this is accomplished by mounting the resistors inside the transducer.

Figure 48.15 shows one technique for correcting for zero shift due to the temperature in semiconductor bridges. Temperature compensation is performed by adding nontemperature-sensitive resistors in series and parallel to the gage having the highest resistance change with temperature. The objective of this method is to achieve both zero balance and temperature compensation together. Since the compensation resistors are nontemperature sensitive, they can be added wherever convenient in the circuit.

The bridge is first balanced using a series resistor at ambient room temperature. Next, the transducer is cycled over the temperature extremes. A parallel resistor is installed across the gage having the greatest resistance change. The bridge is then rebalanced and the procedure repeated until satisfactory performance is achieved.

Semiconductor bridge transducers are typically compensated for calibration or span factor with a circuit as in Figure 48.14. However, r for this situation is a nontemperature-sensitive resistor. For p-type silicon gages, the strain sensitivity drops with temperature while the resistance rises. The increase in resistance occurs at a greater rate than does the

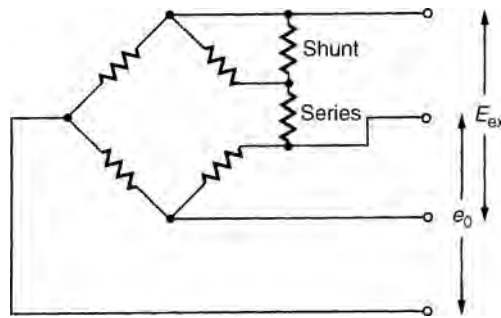


FIGURE 48.15 Transducer bridge compensation for zero shift, semiconductor gages.

decrease in sensitivity. Figure 48.14 shows that the effect of an increase in resistance R , with r constant, is to increase the voltage applied to the bridge, offsetting the decrease in strain sensitivity. Alternately, in Figure 48.14, r can be replaced by a thermistor instead of a fixed dropping resistor. The thermistor is generally a more efficient method of compensation but must be in the same thermal environment as the bridge network.

When balancing Wheatstone bridges, it must be determined that the balancing circuit does not significantly alter the thermal compensation network. Balancing methods are discussed in the following section.

48.5 RESISTANCE BRIDGE BALANCE METHODS

Even when a best attempt is made at matching resistors, the output from a bridge transducer with zero measurand applied is always something other than zero volts. With microprocessors and scanners, this is of little consequence. The initial bridge output can be acquired and stored in the memory of the microprocessor and then subtracted from all subsequent readings. Frequently, however, it is desired to initialize a bridge circuit such that a zero value of measurand corresponds to zero voltage. For example, assume it is desired to acquire a vibration measurement on a space vehicle using a bridge transducer. Assume the channel is to be calibrated for $\pm 20g$ and the accelerometer has a sensitivity of 1 mV/g (g = standard acceleration of gravity). If the data channel range were $\pm 20 \text{ mV}$, and the accelerometer acquiring the measurement had a zero offset of 5 mV , the channel could transmit only in the range of $+15g$ to $-25g$ as opposed to $\pm 20g$. Balancing the bridge would solve this problem.

Equation (48.5) presented the requirement for a balanced bridge. Basically, the resistance ratio of any two adjacent bridge arms must be equal to the resistance ratio of the other two arms. Any bridge-balancing network must then have as its objective the satisfying of this criterion. The two main types of zero balancing methods are those which manipulate one arm of a transducer bridge to bring its output to the desired condition and those which manipulate two adjacent arms of the transducer bridge.

Figure 48.16 presents the most common circuit for manipulating a single bridge arm. A variable resistor R_B is placed across one of the resistors (say R_4) whose value needs to be lessened such that $R_1/R_4 = R_2/R_3$. The effect of R_B in parallel with R_4 is to lessen the value of the bridge arm from R_4 to some new value R_T .

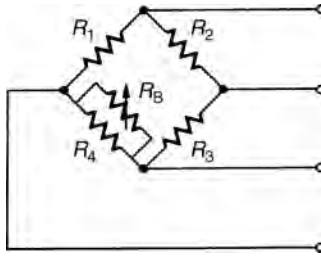


FIGURE 48.16 Circuit for manipulating a single bridge arm.

The overall combination of R_B in parallel with R_4 must be variable over a range at least equal to the maximum possible initial unbalance of the bridge. Selecting this range, other than by trial and error, requires knowledge of the strain gage resistance R , its tolerance in percentage m , and the number of active gages n in the bridge. The range of the balancing circuit should be

$$\frac{2Rmn}{100} \quad (48.20)$$

Note that the presence of the variable resistor R_B desensitizes the bridge network since $\Delta R_4/R_4$ is not equal to $\Delta R_T/R_T$. If the strain gages are initially closely matched, the influence of this effect is small since R_B will remain large and R_T will closely approximate R_4 . For optimum precision, the best method to minimize the influence of the variable resistor is to calibrate the transducer once the bridge is balanced. Of course, if less than four arms of the bridge are active and balancing is performed across a dummy completion resistor, no desensitizing of the bridge occurs.

Two techniques are available to manipulate two adjacent arms in a bridge. Again, the rationale for this manipulation is to satisfy Equation (48.5). The first technique is series manipulation, which assumes the bridge is open such that the variable resistance may be inserted in series with two arms of the bridge. This technique is not applicable to a closed bridge.

Figure 48.17 shows a variable series resistor R_S inserted in one corner of the bridge. The insertion of R_S , which is typically quite small, allows adjustment of the ratio of R_2 to R_3 to achieve balance. Stein (1953) provides the best discussion of bridge balance networks and indicates that minimum bridge desensitization occurs when bridge power is applied across the vertical terminals of Figure 48.17 as opposed to the horizontal.

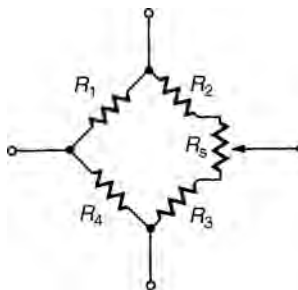


FIGURE 48.17 Circuit for series manipulation of two adjacent bridge arms.

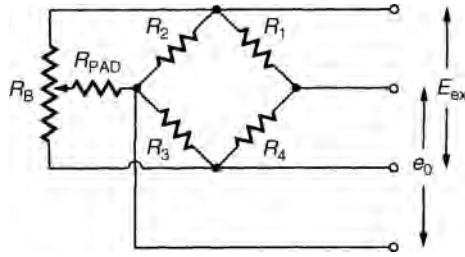


FIGURE 48.18 Circuit for parallel manipulation of two adjacent bridge arms.

The second technique discussed (and the one typically used) is parallel manipulation of two adjacent bridge arms. Figure 48.18 illustrates this technique. The parallel variable resistor R_B allows the ratio of R_2/R_3 to be adjusted. A pad resistor, R_{PAD} , serves simply to keep the individual bridge arms from being shorted out at the end of travel of R_B . Again, the secret is to keep the combination of R_B and R_{PAD} as high as possible to avoid bridge desensitization. If no other guidance is available, start out with a pad resistor about 100 times the bridge resistance and a variable resistor about 20 times the bridge resistance. Again, maximum accuracy is achieved when the bridge transducer is calibrated with the balance network with which it will be used.

As alluded to earlier, the addition of a balance network to a bridge transducer may react unfavorably with temperature compensation resistors placed in the transducer's circuitry by its manufacturer. Temperature compensation can be severely modified by the presence of this balancing network. The prerequisite to insertion of a balance network should be an exact knowledge of the circuit of the transducer. For this reason, and for reasons associated with desensitization of the transducer, balance networks should be avoided unless required.

48.6 RESISTANCE BRIDGE TRANSDUCER MEASUREMENT SYSTEM CALIBRATION

A basic component in any measurement system is the transducer. The measurement system can be as simple as a transducer, a cable from the transducer, and a recorder. Alternately, the measuring system can contain many more of the elements of the transmitting and receiving system defined in Section 48.1. Cables, amplifiers, filters, digitizers, tape recorders, and so on all have the capability, when inserted into a measurement system, to modify both the amplitude and spectral content of the signal from the transducer defining the measurand. The response of these components may also drift with time.

To obtain measurements of the highest possible quality, one must accurately and carefully calibrate the entire measurement system as near to the time of actual measurement as possible. The calibrations may be conducted prior to, immediately after, or even during the time of actual measurement. Such calibration of an entire measurement system is referred to as "end-to-end" calibration. This calibration ordinarily does not replace the evaluation of individual components of the measurement system.

One group concerned with "end-to-end" calibration of measurement systems is the Telemetry Group/Range Commanders Council whose Secretariat is headquartered at

White Sands Missile Range, New Mexico. The Transducer Committee of the Telemetry Group coordinated the writing of Chapter 2 of *End-to-End Test Methods for Telemetry Systems* (1979), entitled "Test Methods for Transducer-Based System Calibrations." The following information is largely extracted from that chapter, which also deals with piezoelectric transducers, servo transducers, capacitive and inductive transducers, and thermoelectric transducers.

A preferred calibration procedure is one in which a known value of the measurand is applied directly to the transducer in the measurement system (transmitting, receiving, and display) in which it will be used. This procedure permits the output display to be read directly in terms of units of the measurand.

48.6.1 Static Calibration

The basic equipment for the static calibration of transducer systems consists of a measurand source supplying accurately known and precisely repeatable values of the measurand and an output-indicating or recording system. The combined errors or uncertainties of the calibration system should be sufficiently smaller than the permissible tolerance of the system performance characteristic under evaluation so as to result in meaningful calibration values. All calibration system components should be periodically checked against standards. Environmental conditions during calibrations should be constant and specified to permit corrections to the data, as required.

The procedures, which will be specified, are based on the assumption that the measuring system is linear. For systems that will ultimately measure dynamic data, linearity is a prerequisite.

The static calibration sequence consists of the following steps:

1. zero-measurand output verification,
2. sensitivity verification,
3. linearity and hysteresis verification,
4. repeatability verification.

Zero-measurand output verification starts with a measurement of system output with zero measurand applied to the transducer. Zero measurand is an important measurement for several reasons:

1. In many measurement systems, the transducers are exposed to zero measurand before the test begins.
2. For many transducers, zero measurand means that no external input to the transducer is required, thus greatly simplifying the test procedure.
3. Measurement system malfunctions, including drift, will frequently appear when the system output is monitored over some reasonable time period with zero measurand applied.
4. In some measurement systems with no external measurand input, the ambient environment will furnish an important reference point for the system calibration. For example, for an absolute-pressure-measuring system, the ambient atmospheric pressure provides this reference. Similarly, for certain accelerometer systems, the measurable attitude of the test vehicle prior to launch

represents a known component of the earth's gravitational field as input to the accelerometer.

For a measurement system with a linear response, the slope of the line characterizing the input measurand versus system output represents the sensitivity of the system. There are a number of straight lines that may be chosen to provide this sensitivity verification. These include the following:

1. Endpoint Line: The straight line between the outputs at the specified upper and lower limits of the range.
2. Best Straight Line: The line midway between the two parallel straight lines closest together that enclose all output versus measurand values on a calibration curve.
3. Terminal Line: The straight line for which the endpoints are 0 and 100% of both measurand and output.
4. Theoretical Slope Line: The straight line connecting the specified points between which a specified theoretical curve has been established.
5. Least-Squares Line: The straight line for which the sum of the squares of the residuals is minimized for all calibration points.

Procedures used in the verification of sensitivity will depend on specific accuracy, calibration time, and expense trade-off choices for each system. For unidirectional transducers, it is typical to calibrate from zero to full scale and back again in 10% of full-scale increments (21 points). For bidirectional transducers, a 21-point calibration cycles the transducer from negative full scale to positive full scale and back again in 20% of full-scale increments.

Data extracted from these calibrations are typically linearity and hysteresis. Linearity is the closeness of a calibration curve to the specified straight line, expressed as the maximum deviation of any calibration point from that line during any one calibration cycle. Hysteresis is the maximum difference in output at any measured value within the specified range when the value is approached first with increasing and then with decreasing measurand.

The reference straight line selected is often the linear least-squares line. This line is based on the following principle: the most probable value of an observed quantity is such that the sum of the squares of the deviations of the observations from this value is a minimum. This is based on the fact that most measurements of physical quantities show a normal distribution with both positive and negative deviations from the mean probable and very large deviations less likely than small deviations. The line can be defined unequivocally in terms of the quantities measured. The line also is statistically significant, and standard deviations can be assigned to estimates of the slope, intercept, and other parameters derived from it.

An additional parameter describing the performance of the measurement system is obtained by repeating the static calibration of the system. A minimum of two, but preferably three, consecutive static calibrations yield data from which the "repeatability" of the system is verified. Repeatability is the ability of the measurement system to reproduce output readings when the same measurand value is applied to it consecutively under the same conditions and in the same direction. It is expressed as the maximum difference between corresponding values from at least two consecutive calibrations. Although there

is no universal agreement as to the particular values selected, a value close to full-scale output is commonly used.

If the bridge transducer will be used to acquire time-varying measurements, the measuring system must be both dynamically and statically calibrated. The dynamic response of any system is described by a frequency response function which is a complex function of frequency. The frequency response function relates system output to system input in the frequency domain. For measurement systems, this frequency response function is typically represented by Bode plots, which are log amplitude and phase versus log frequency.

48.6.2 Dynamic Calibration

Dynamic calibrations are inherently more difficult to perform than are static calibrations and usually require specialized equipment. Dynamic calibrations can be performed using several types of well-defined input signals, such as applications of sinusoids, transients, or broadband noise. The principal requirement that the input must satisfy is that it must contain significant energy at frequencies over the range of the frequency response function of interest.

The dynamic calibration sequence consists of the following steps:

1. dynamic sensitivity determination;
2. dynamic amplitude linearity determination;
3. amplitude–frequency verification;
4. phase–frequency verification.

If the measuring system does not have zero frequency response, its end-to-end calibration is of necessity made by dynamic methods. The simplest approach to dynamically determining system sensitivity is the application of a sinusoidally varying measurand to the transducer. The amplitude of the measurand should be equal to the range of the transducer. For unidirectional transducers, this test involves biasing the transducer to its half-range point. At the test frequency, it is possible to relate the peak amplitude of the system response to the amplitude of the measurand and determine system dynamic sensitivity.

It is further desired to acquire dynamic amplitude linearity by performing tests equivalent to dynamic sensitivity determination at several levels of the measurand (usually levels of 25, 50, 75, and 100% of full scale are adequate). This testing should be performed at several different frequencies over the range of the frequency response function of interest. If the measurement system cannot be verified to be linear, it should not be used to acquire time-varying measurements.

The concept of a measuring system having a unique amplitude–frequency and phase–frequency response is only meaningful for systems which have been verified to be dynamically linear. Amplitude–frequency response tests consist essentially of a series of dynamic sensitivity determinations at a number of frequencies within the bandwidth of the system. Three is the *minimum* number of test frequencies. One test should be performed close to the upper limit of the frequency band where the response has not been affected by the high-frequency roll-off characteristic of the system. The second frequency should be sufficiently higher than the first to provide some indication of the roll-off rate of the system. The third frequency should be about halfway between zero frequency and the first test frequency to verify a flat response to the upper band edge. More

improved definition obviously can be provided by increasing the number of test frequencies.

Phase–frequency response characteristics of a measuring system can often be acquired simultaneously with the amplitude–frequency response. An output recording device is required with two identically responding channels. The system output is recorded on one channel. The second channel records the measurand, which is typically acquired by a previously calibrated monitoring transducer whose amplitude–frequency and phase–frequency response characteristics are well established. A time correlation between the system output and this monitoring transducer can establish measuring system phase–frequency response. For systems measuring signals whose time history is important, a linear phase response with frequency is required. For those signals about which only statistical information is to be acquired (e.g., random vibration), phase response is not an important system characteristic.

With today's technology, frequency response functions can also be characterized by transient or random system excitation. Dual-channel spectrum analyzers can ratio input-to-output measuring system Fourier transforms in near-real time. Recall that the system input stimulus must contain significant signal content at all frequencies of interest.

48.6.3 Electrical Substitution Techniques

If actual values of the measurand cannot be used to calibrate resistance bridge transducers, electrical substitution techniques can be used. Test equipment required includes a precision voltage source, precision resistors or decade box, and a signal generator. The techniques include shunt calibration, series calibration, and bridge substitution. Shunt calibration techniques are discussed first.

Inserting a resistor of known value in parallel with one arm of a strain gage bridge is single-shunt calibration. The calibration resistor is inserted across the arm opposite the strain gage conditioning system. The conditioning system may contain a balance potentiometer, a limit or pad resistor, modulus resistors, and temperature compensation resistors. Standard practice is to insert the shunt resistor between the negative input (excitation) and the negative output (Figure 48.19). This reduces errors caused by shunting some of the bridge-conditioning resistors.

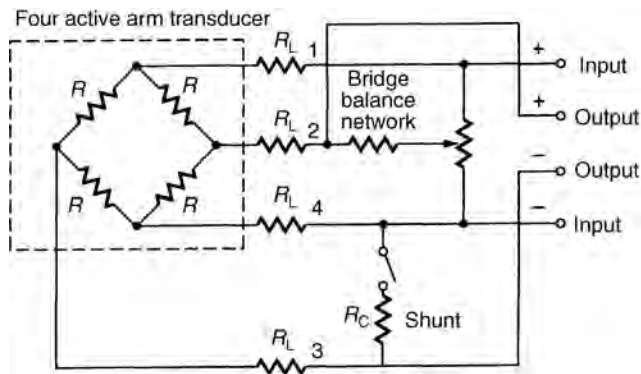


FIGURE 48.19 Single-shunt calibration of bridge transducer.

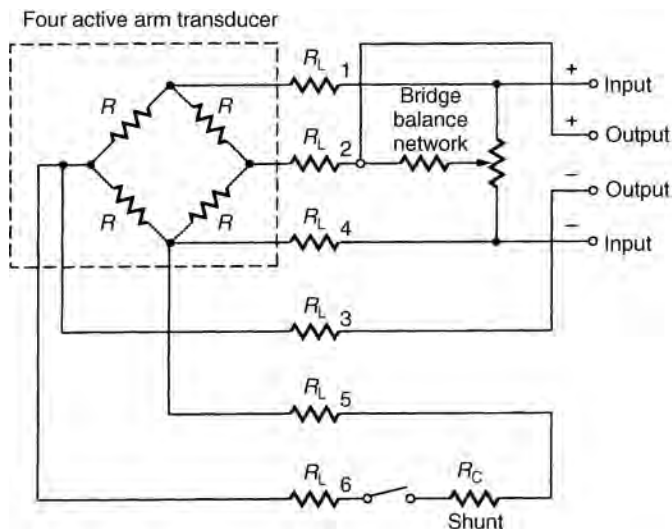


FIGURE 48.20 Remote-location single-shunt calibration of bridge transducer.

The value of the shunt resistor R_C is determined by first applying a value of the measurand to the transducer and monitoring the voltage change at the transducer output terminals (Figure 48.19). With the measurand removed, a decade box is substituted for R_C and its resistance adjusted until a voltage change results with a magnitude equal to that caused by the measurand. For subsequent calibrations, a fixed resistor R_C can be substituted for the decade box. When the switch in series with R_C is closed, it will produce a step voltage through the measuring system of amplitude equal to that produced by the measurand. When shunting one arm of the bridge, the resistance change produced in that arm is $-R^2/(R_C + R)$.

In the calibration laboratory, the small lead length associated with the transducer introduces no error in establishing R_C . The application of the bridge transducer in the field can require significant lengths of cable with significant transmission line resistance. Figure 48.20 illustrates the situation where R_C must be applied remotely. If R_C were applied directly at the bridge, loading errors introduced by transmission lines would be accounted for. If (as in Figure 48.20) R_C is applied at a remote location, the effect of the transmission line resistance $2R_L$ in series with R_C must be considered.

Bipolar shunting is used when the physical loading creates both positive- and negative-going signals and it is desired to create positive and negative calibration outputs. The calibration resistor is alternately inserted across the two arms opposite the bridge-conditioning network. If line resistance is significant, it must be considered as in Figure 48.21.

Series calibration of bridge transducers is considerably different from shunt calibration. Figure 48.22 describes this process. Series calibration consists of two distinct calibration phases.

In the zero calibration phase the Zero Cal switch is moved downward so excitation is removed from the bridge. The sensitivity resistor (R_{SENS}) is concurrently placed across the bridge input terminals, simulating the power supply impedance to result in the same overall system impedance encountered in the data circuit. Zero bridge transducer output is recorded.

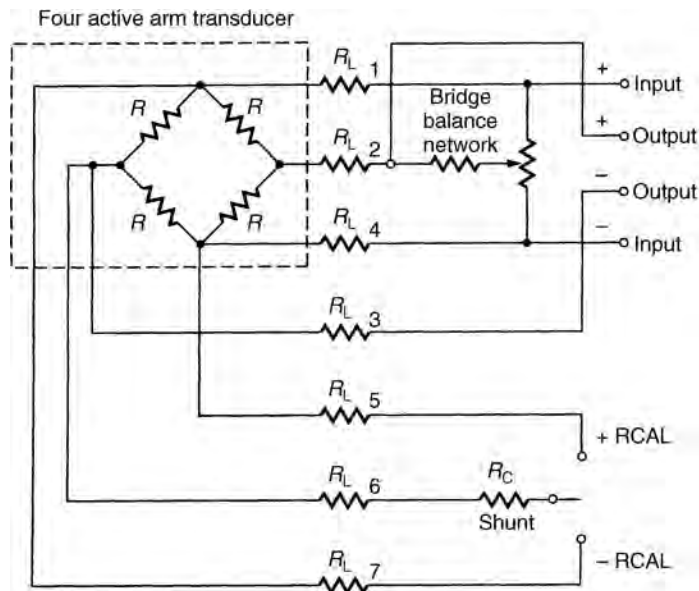


FIGURE 48.21 Remote-location double-shunt calibration of bridge transducer.

The next calibration phase is the series phase. In the series calibration mode, the two switches in Figure 48.22 are closed with the Zero Cal switch back in its original position, introducing R_{C1} and R_{C2} into the circuit. This removes power from one corner of the bridge and puts a calibration resistor R_{C1} in series with the sensitivity resistor and one side of the bridge output. The second calibration resistor becomes intermediate between the R_{SENS} -to- R_{C1} connection and excitation return. This second resistor, R_{C2} , is selected to maintain the approximate equivalent bridge impedance across the excitation. The calibration circuit then electrically simulates the bridge transducer. The value of R_{C1} is determined experimentally, corresponding to some measurand equivalent.

Series calibration overcomes a serious shortcoming of shunt calibration. During application of a shunt resistor, the transducer can still respond to mechanical input. The calibration step is superimposed upon any mechanically induced signal present. If the mechanical input is static and of sufficient magnitude, overranging will invalidate the

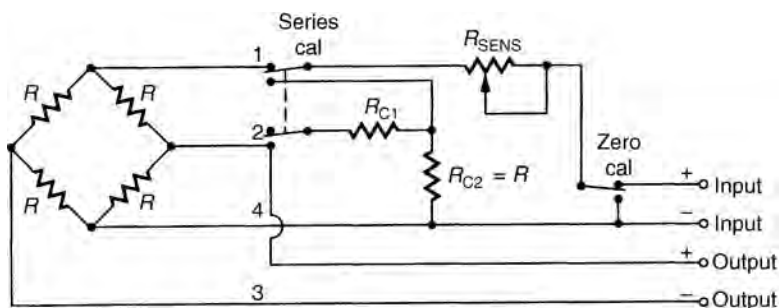


FIGURE 48.22 Series calibration of bridge transducer.

calibration step. If the mechanical input is dynamic, it may be impossible to accurately measure the magnitude of the calibration step. The magnitude of the series calibration step is significantly more independent of this mechanical input. As in all calibration, transmission line resistance must be considered where significant. Similarly, a change in sensitivity resistance modifies the effect of the series calibration resistance. However, the typical error incurred is negligible.

The final electrical substitution technique discussed is bridge substitution. This technique involves substitution of a model for the bridge transducer itself. Figure 48.23 represents a typical low-level bridge system.

An accurate bridge transducer model has the same terminal impedance as the transducer and provides a fast and simple method of generating a static and dynamic output equivalent to that generated by the transducer for a given physical load. It also provides a convenient method for verifying the calibration resistor's measurand equivalency for shunt and series systems. The two types of bridge transducer models employed for system calibrations are the shunt resistor adapter and the shunt resistor bridge.

Figure 48.24 describes the shunt resistor adapter, which is simple, inexpensive to construct, and an exact model since it is used in conjunction with the actual transducer. The adapter is inserted between the transducer and the rest of the measurement system. It performs three primary functions:

1. It supplies the stimulus for performance of system end-to-end calibrations. Shunting the arms of a transducer bridge with the appropriate resistors produces an unbalance in the bridge equivalent to that produced by a given measurand. The adapter provides a convenient method of applying these shunt resistors directly to the bridge with negligible line loss (S_1 and S_2).
2. It performs a system frequency response test. A convenient system frequency response can be performed by selecting the appropriate shunt resistor and sweeping the adapter's AC power supply over the desired range. Figure 48.25 shows a typical oscilloscope display of the results.
3. It provides a convenient check of the system's calibration resistors (R_c) and equivalents. The system's R_c equivalent will differ from the values established by the laboratory calibration as a function of line resistance, calibration resistor tolerance, and so on. Since the adapter shunt resistors are precision resistors that are applied directly to the bridge, the equivalency of the adapter shunt resistors will not be affected by lead resistance and other variables.

Although the shunt resistor adapter model is a very powerful and simple calibration tool, it has two undesirable characteristics. The least desirable characteristic is that the system calibration and calibration resistor equivalents generated by the adapter are incremental values superimposed on the transducer output resulting from the physical stimulus acting at the time of the test. Also, the adapter does not provide a fixed independent reference since it is used in conjunction with the transducer.

The undesirable features of the shunt resistor adapter can be eliminated by replacing the actual transducer with a bridge model (Figure 48.26). Since the shunt resistor bridge (bridge model plus shunt resistor adapter) is a stable, complete model of the transducer, it can be used to perform an absolute end-to-end system calibration and can be a valuable tool in troubleshooting.

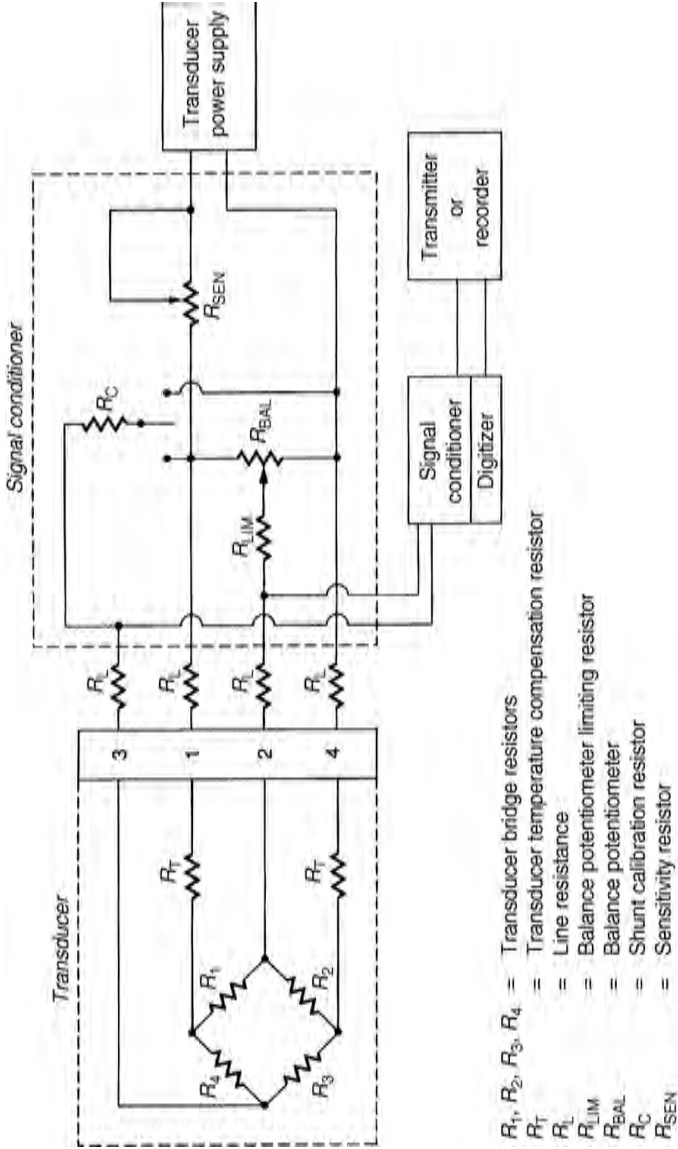


FIGURE 48.23 Typical low-level bridge system.

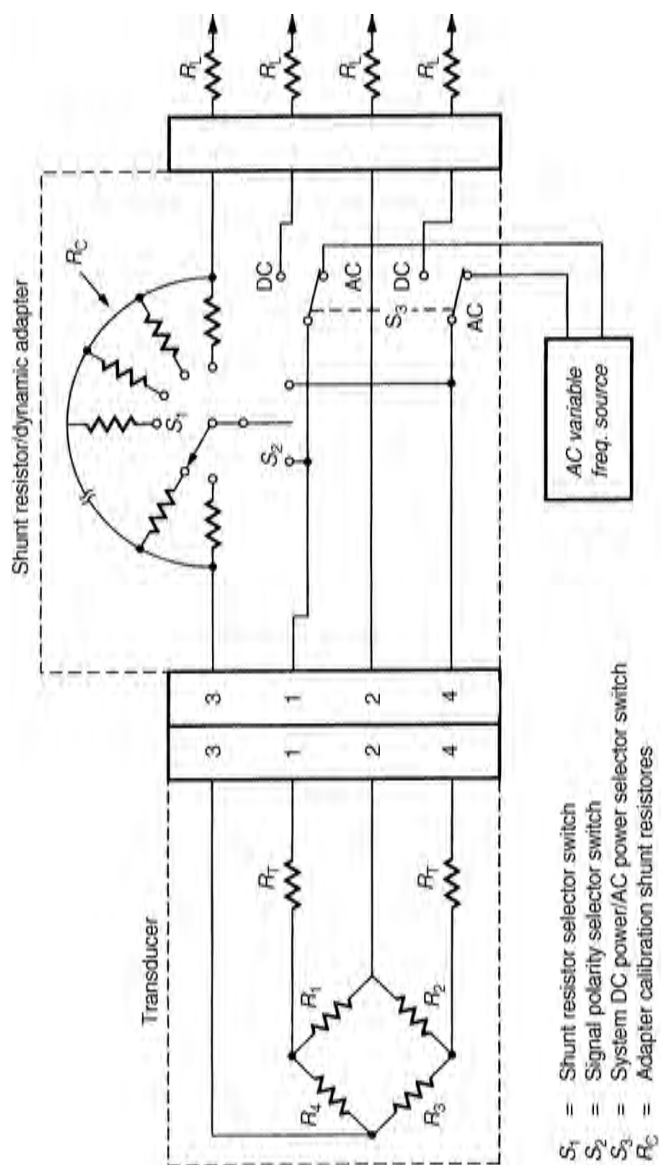


FIGURE 48.24 Shunt resistor adapter.

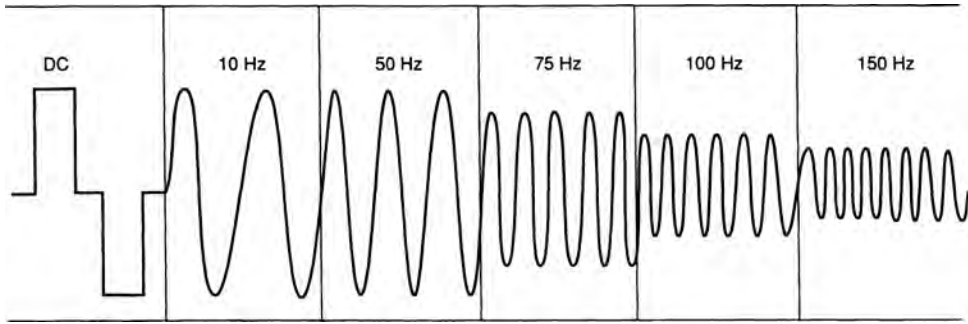


FIGURE 48.25 Oscilloscope frequency response display.

Several disadvantages are encountered when using the shunt resistor bridge as a calibration tool. Since some transducers are hard to model, it is difficult to ensure that the bridge is a representative model of the transducer under all conditions. Furthermore, a different bridge model is required for each major transducer design.

As a final note, remember that the resistance of semiconductor bridge transducers is strongly a function of temperature. When using shunt or series calibration techniques on semiconductor bridges, ambient temperature changes should be taken into account.

48.7 RESISTANCE BRIDGE TRANSDUCER MEASUREMENT SYSTEM CONSIDERATIONS

48.7.1 Bridge Excitation

When amplifiers and power supplies were formally designed around vacuum tubes, component drift was a problem in bridge transducer measurement systems. Alternating-current power supplies in bridge circuits eliminated many of these problems by operating at frequencies above AC. Most bridge transducer power supplies today are AC. When comparing AC supplies with AC, the following advantages are associated with AC:

1. simpler circuitry;
2. wider resultant instrumentation system frequency response;
3. no cable capacitive or inductive effects due to the excitation;
4. simpler shunt calibration and bridge balance circuitry.

Independent of type of supply, the power level selected has to take account of all variables, which affect the measurement. These include gage resistance, gage grid area, thermal conductivity of flexure to which gage is mounted, flexure mass, ambient test temperature, whether used on a static or dynamic test, accuracy requirements, and long- or short-term measurement. These variables account for the fact that a strain gage is a resistance which has to dissipate heat when current passes through it. Most of the heat is conducted away from the gage grid to the transducer flexure. The result of inadequate heat conduction is gage drift.

For transient measurements, a steady transducer zero reference is not as important as for static measurements. Bridge power can be significantly elevated to increase measurement system signal-to-noise ratio.

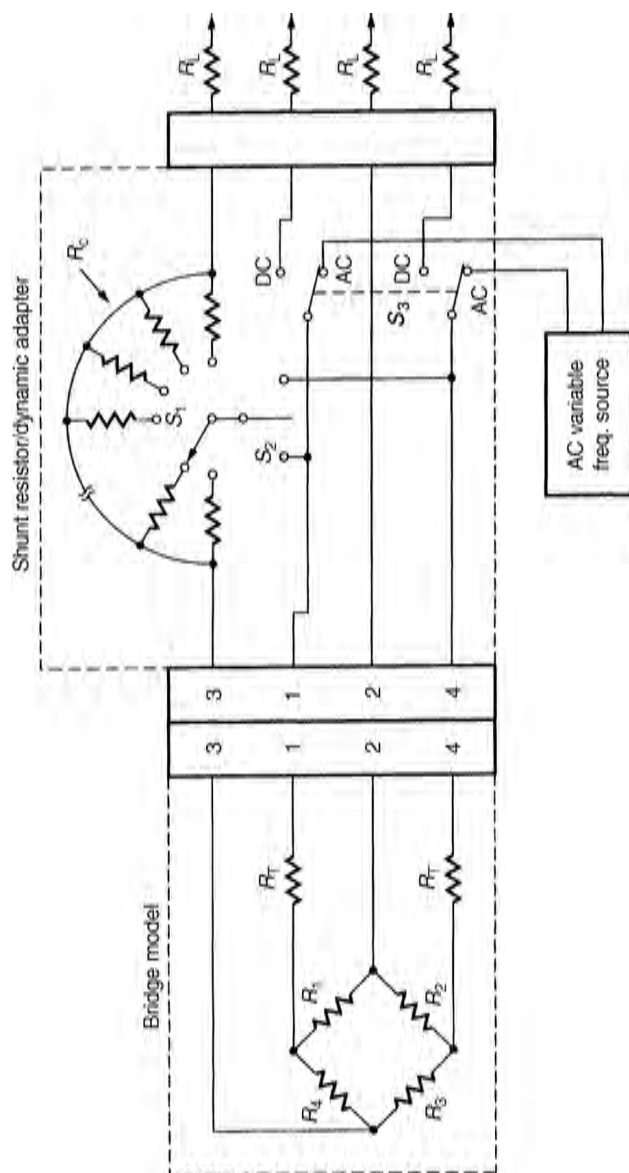


FIGURE 48.26 Shunt resistor bridge.

The following specifications define key performance parameters of AC output instrumentation power supplies. Input supply can be either AC or AC.

1. *Warmup Time*: The time necessary for the power supply to deliver nominal output voltage at full-rated load. It is usually specified over the range of operating temperatures.
2. *Line Regulation*: The change in steady-state AC output voltage resulting from an input voltage change over the specified range.
3. *Load Regulation*: The change in steady-state AC output voltage resulting from a full-range load change.
4. *Efficiency*: The ratio of the output power to the input power.
5. *Load Transient Recovery*: The time required for the output AC voltage to recover and stay within a specified band following a step change in load.
6. *Periodic and Random Deviation*: The AC ripple and the noise of the AC output voltage over a specified bandwidth with all other parameters held constant.
7. *Stability (Drift)*: The deviation in the AC output voltage from AC to an upper limit that coincides with the lower limit as specified above in 6.
8. *Temperature Coefficient*: The change in output voltage per degree change in ambient temperature.

Test Methods for Vehicle Telemetry Systems (2003) defines test procedures for these specifications.

48.7.2 Signal Amplification

In addition to providing a precision power source to bridge transducers, the resultant millivolt signals from these transducers often require amplification. This amplification is usually performed by a differential AC amplifier. A differential AC amplifier is an electronic circuit whose input lines are conductively isolated from the output lines, power, and chassis ground and whose output voltage is proportional to the differential input signal voltage. Ideally, both input lines have equal impedance and transfer characteristics with respect to the amplifier ground structure. The amplifier has a frequency response from 0 Hz to a value determined by the bandwidth of the amplifier.

Selecting amplifiers can be difficult because specification terminology is not universally standardized. Amplifier specifications are either referred to input (RTI) or referred to output (RTO). Discussing these specifications can lead to an understanding of the amplifiers themselves.

1. *Input Impedance*: The minimum impedance the amplifier will present when operated within its specification. It is the impedance seen between the two ungrounded input lines of the amplifier.
2. *Source Current*: The bias current flowing through the circuit comprised of the amplifier input terminals closed through the source resistance. The amplifier input transistors act as constant-current generators in series with the input terminals. This current can result in both offset voltage and common-mode voltage.

3. *Common Mode Rejection*: The measure of the conversion of common-mode voltage to normal differential signal. The common-mode input voltage is the voltage common with both inputs to the amplifier. A common-mode rejection of 60 dB implies that a 10-V signal applied simultaneously to both inputs produces an error signal RTI of 10 mV.
4. *Linearity*: The maximum deviation from the least-squares straight line established through the output voltage versus differential input voltage characteristic. In evaluating linearity, it is usually sufficient to test at the highest and lowest gains, since linearity will be worst at these settings.
5. *Gain Range*: The slope of the least-squares straight line established through the output voltage versus the differential input voltage characteristic of the amplifier. The gain range is the maximum and minimum values of gain available from the amplifier without causing any degradation in performance beyond the limits of the specification.
6. *Gain Stability with Temperature*: The change in amplifier gain as a function of ambient temperature for any gain in the specified gain range.
7. *Zero Stability with Temperature*: The change in output voltage with temperature. It must be specified as RTI or RTO, and this test is typically performed with the amplifier input leads terminated with the maximum source impedance and no signal applied. A warm-up period is usually specified for both this test and gain stability with temperature.
8. *Frequency Response*: The minimum frequency range over which the amplifier gain is within ± 3 dB of the AC level for all specified gains for any output signal amplitude within the linear output voltage range. In writing specifications, it is not uncommon for a user to also specify the desired phase characteristics over the frequency range of interest and the number of filter poles.
9. *Slew Rate*: The maximum rate at which the amplifier can change output voltage from the minimum to the maximum limit of linear output voltage range. It is expressed in volts per microsecond with a large-amplitude step voltage applied to the input of the amplifier and the amplifier driving a specified capacitive load. The usual source of slew rate difficulty is current limiting, and this specification (a nonlinear process) should not be confused with rise time (a linear process).
10. *Settling Time*: The time following the application of a step voltage input for the amplifier output voltage to settle to within a specified percentage of its final value.
11. *Overload Recovery*: The time required for the amplifier to recover from a specified differential input signal overload. It is specified as the number of microseconds from the end of the input overload to the time that the amplifier AC output voltage recovers to within the linear output voltage range. Amplifier gain must be specified.
12. *Noise*: Noise is divided into two components: RTI and RTO. RTI noise is that component of noise that varies directly with gain. It is measured with the amplifier input leads terminated in the maximum source impedance and no signal applied. The RTO noise is that component of noise which remains fixed with gain.
13. *Harmonic Distortion*: The maximum harmonic content for any amplifier frequency or output amplitude within the specified limits.

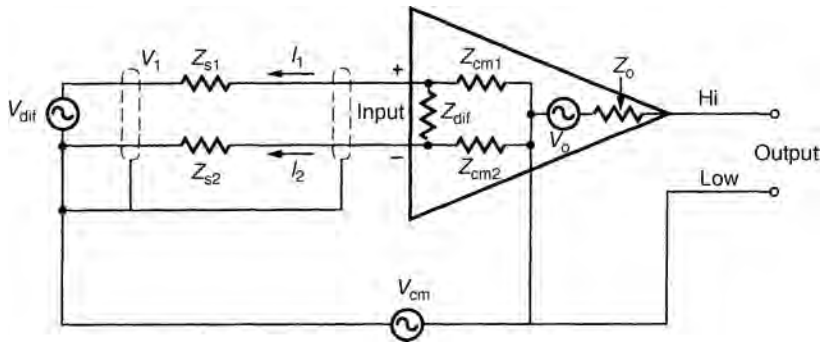


FIGURE 48.27 Basic AC amplifier circuit.

14. *Output Impedance:* The maximum impedance the amplifier will present when it is operated anywhere within its specification. This specification is important in resistive loading ratings or in determining the amount of capacitance which can be connected across the output without causing instability.

Test Methods for Vehicle Telemetry Systems (2003) describes test procedures for these specifications and discusses them further. Figure 48.27 presents the basic AC amplifier circuit. Jaquay (1972) provides additional discussion directed toward understanding AC instrumentation amplifiers.

48.7.3 Slip Rings

In many measurement applications, it is necessary to acquire data from rotating machinery. Turbines, rate tables, and centrifuges are examples of such machinery. If it is necessary to measure strain, pressure, torque, force, and so on, on the rotating machine component, signals from bridge transducers must be coupled from this component to a stationary instrumentation system. Instrumentation slip rings accomplish this function.

In their simplest form, slip rings consist of a metal ring on the rotating machine component against which a brush attached to the stationary machine portion is spring loaded to make ohmic contact. Precious metals are generally used for mating surfaces to minimize contact resistance.

Slip rings came into existence in the 1940s with initial application in the aircraft industry. In the 1950s, mercury slip rings came into existence. These latter rings, which first found application at Rolls Royce in England, use mercury as the signal transfer medium. The mercury is entrapped between the rotor and the stator of the ring assembly. Today, slip rings are capable of operating from very low RPM to tens of thousands of RPM.

Noise induced in slip rings is of the ohmic contact type, that is, it is roughly proportional to current. A high brush pressure reduces noise at the expense of increased brush wear. Brush wear is a function of the brush pressure, material, finish (usually microinch), and flatness. One technique for lowering contact noise is to mount several brushes in parallel on the same ring.

Because ohmic changes in the slip rings can be of the same order of magnitude as resistance changes in the bridge transducer, full bridges are almost always used on the

rotating part to avoid inserting slip rings within the bridge itself. Slip rings in the output circuits of bridge transducers using voltage monitoring do not create significant problems because any small resistance changes in the rings are in series with the large input impedance of the voltage-measuring device and are effectively ignored. Slip rings in the input circuits of bridge transducers operating from a constant-voltage source can create problems if they cause fluctuating voltage drops in series with the transducer. For this reason, constant-current sources are preferred when using slip rings.

Other techniques for extracting data from rotating machinery have evolved over the years. These include rotary transformers, light modulation, and radio frequency (RF) telemetry. Of these schemes, RF telemetry has displayed the most promise with commercially available low-power transmitters capable of operating up to 30,000g.

48.7.4 Noise Considerations

Many other sources besides slip rings can induce unwanted spurious signals in these transducers. Since the unamplified output from bridge transducers is typically ones or tens of millivolts and never more than a few hundred, they are easily influenced by noise sources. The following discussion defines noise, documents how to verify its existence (or hopefully nonexistence), and provides some hints as to how to suppress noise in bridge transducer measuring systems. Stein (1972) provides a basis for this discussion.

The output of measuring system components represents combinations of responses to environments. These environments can be divided into two categories: desired and all others (undesired). For example, consider a bridge pressure transducer in a hostile explosive environment. Its desired environment is pressure. Other undesired environments it encounters are temperature, acceleration, ionized gas, and so on. Ideally, the transducer would respond to pressure alone. In practice, an additional response is elicited from the transducer due to the other environments; usually, but not always, the response is small compared to the pressure response.

Two response types exist for a bridge transducer: self-generating and nonself-generating. Nonself-generating responses are due to changes in the material properties or geometries within a transducer. Power has to be applied to the transducer to elicit a nonself-generating response. For example, pressure applied to the diaphragm of a pressure transducer with bridge electrical power supplied modifies the impedance of the strain gage circuit and results in a millivolt output (nonself-generating). Self-generating responses are those attributable to various measurands applied to bridge transducers without electrical power supplied. Examples of these responses include thermoelectric-, photoelectric-, pyroelectric-, and magnetoelectric-induced voltages within the bridge circuit. Thus, there exist four environment–response combinations in bridge transducers. The nonself-generating response to the desired environment is defined as signal. The nonself-generating response to the undesired environment, as well as the self-generating response to both the desired and the undesired environment, is noise. Figure 48.28 illustrates the paths associated with these four combinations with path 4 being signal and paths 1, 2, and 3 being noise.

The quantifications of paths 1–4 can be accomplished by switching. If at some time during the test bridge power is switched off, Figure 48.28 indicates that only paths 1 and 3 will exist. Since these paths are both noise, the bridge transducer response ideally should approach zero. Similarly, if the desired environment can be switched off for some time period, only paths 1 and 2 remain. If path 1 was verified as being noise free when bridge power was removed, path 2 also becomes quantified. If paths 1, 2, and 3 are all

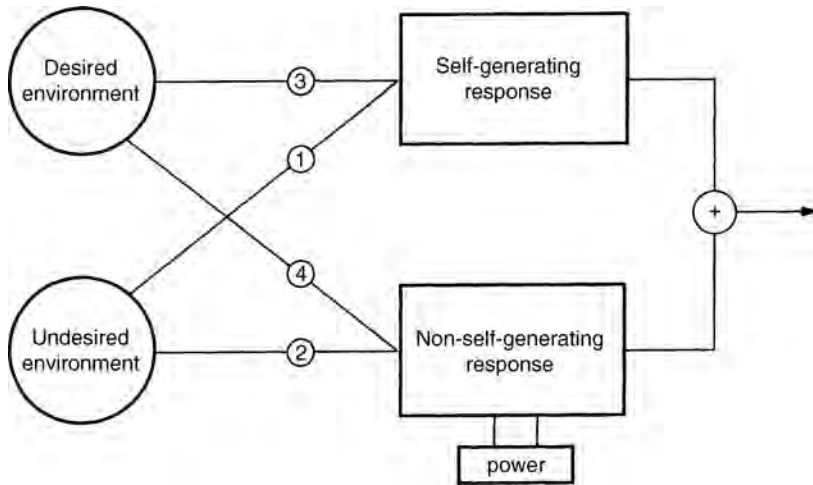


FIGURE 48.28 Bridge transducer model for noise hunting and documentation. (Adapted with permission from “Information as a ‘Noise Suppression’ Method,” by Peter Stein, Stein Engineering Services, Inc., Phoenix, AZ, LR/MSE Publ. 66, 1975.)

shown to provide negligible signal level, transducer output becomes attributable to path 4, which has been defined as signal. In summary:

Remove Bridge Power: Document paths 1 and 3.

Reapply Power and Remove Desired Environment: Document paths 1 and 2.

If the documented signal paths are of sufficiently small magnitude to be considered inconsequential during test, the nonself-generating response to the desired environment (signal) is recorded.

Some question may arise as to how to implement these procedures, particularly in transient measurement situations. For example, assume a bridge accelerometer is to be used to measure a transient acceleration event. Three accelerometers can be fielded in close physical proximity. The first can be mounted without power applied to document paths 1 and 3. The second can have power applied but be mounted in a piece of foam to isolate it from the acceleration environment, resulting in documentation of paths 1 and 2. The third can be powered and properly mounted to measure the acceleration environment. If the first two accelerometers produce no output, then the output from the third is the noise-free signal.

If noise is present in measuring systems containing bridge transducers, noise control efforts are dictated by the specific noise type. Electric and magnetic fields can be shielded, noise components at specific frequencies can be filtered, thermal transient effects can be absorbed or delayed, steady-state temperature effects can be compensated, and so on. The prerequisite to any noise control is documentation of its presence.

As noted earlier, most modern resistance bridge transducers use AC power supplies as opposed to DC power supplies. However, AC power supplies still have an important role to play with resistance bridge transducers. The AC power supplies accomplish noise suppression by separating the self-generating responses from the nonself-generating

responses in a transducer by moving the frequency content of a signal into some new range of the frequency spectrum. This procedure is known as amplitude modulation. Stein (1975) forms the basis for the following discussion.

Referring to Figure 48.28, amplitude modulation can eliminate paths 1 and 3 (noise) from the net output signal. Thus, self-generating electromotive force (emf), such as thermoelectric and electromagnetic ones, can be separated from emf attributable to resistance changes in a bridge transducer.

A design procedure is presented for an AC-powered bridge where the signal input to the self-generating response extends to frequency ω_1 , and the input to the nonself-generating response extends to ω_2 . Power supplied is at frequency ω_3 . The design method developed requires a knowledge of these frequencies. A procedure to determine these frequencies involves the following:

1. performing a frequency analysis of the signal recorded with no power supplied to determine ω_1 ;
2. applying power under normal operating conditions and comparing signal frequency content to the results of 1 to determine ω_2 .

An example follows where noise is present as a self-generating response and the bridge is powered first by a AC supply and then by an AC supply.

In this example, the nonself-generating response (signal) occurs at frequency ω_p and the self-generating response (noise) occurs at two frequencies bracketing ω_p ($\omega_a = \omega_p/2$ and $3\omega_a$). Figure 48.29 illustrates the wave shapes assumed for the self-generating and nonself-generating responses with AC power supplied ($\omega_3 = 0$). Figure 48.29a represents both self-generating response input and output, Figure 48.29b represents nonself-generating response input and output, and Figure 48.29c represents the total transducer output (summation of Figure 48.29a, b). It is seen that the two responses are hopelessly intermingled and that the signal cannot be separated from the noise. Figure 48.30 shows the frequency content of the wave shapes, with Figure 48.30a corresponding to Figure 48.29a, Figure 48.30b corresponding to Figure 48.29b, and Figure 48.30c corresponding to Figure 48.29c. The AC-powered bridge will be presented as a solution to measurement problems such as this. Frequency ω_3 is typically selected as 10 times ω_p .

Figure 48.31 describes this situation for the AC bridge. Figure 48.31a describes bridge power. Figure 48.31b describes the output from the nonself-generating response, which now contains frequencies at $\omega_c - \omega_p$ and $\omega_c + \omega_p$. Here, ω_c is defined to be the carrier frequency, ω_3 . Figure 48.31c describes the net transducer output, which is a summation of Figures 48.29a and 48.31b. Figure 48.32 describes the frequency content associated with Figure 48.31, respectively. The frequency content in Figure 48.32c associated with the time history of Figure 48.31c shows conclusively that the nonself-generating information has been moved from its original frequency, ω_p , to occupy a new frequency range, $\omega_c - \omega_p$ to $\omega_c + \omega_p$, while the self-generating response is left at ω_a and $3\omega_a$. The nonself-generating response is then in that part of the frequency spectrum where no appreciable noise exists and can be separated by band-pass filtering.

After bandpassing, a problem still remains in phase sensing. Figure 48.33 describes this problem. Figure 48.33 illustrates an amplitude-modulated signal after bandpassing to remove the effects of any self-generating response which may be present. This amplitude-modulated signal is ambiguous in that it could correspond to any of the lower four signal

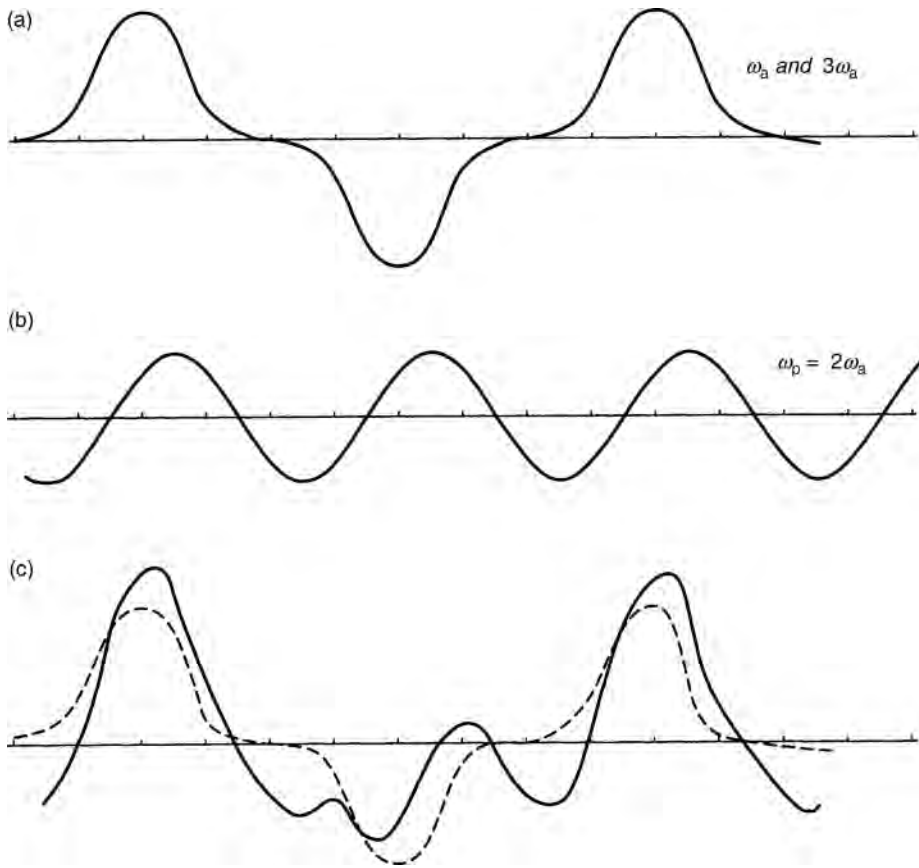


FIGURE 48.29 Signal wave shapes, AC bridge. (Adapted with permission from “Information as a ‘Noise Suppression’ Method,” by Peter Stein, Stein Engineering Services, Inc., Phoenix, AZ, LR/MSE Publ. 66, 1975.)

inputs to the nonself-generating response. The problem of phase sensing associated with a modulated signal is that of determining which portion of the modulated wave shape is positive and which is negative.

If a modulated signal emerges from a measuring system which is initially balanced (zero output for zero input), phase sensing must be done in a phase-sensitive manner. The general principle for all phase sensors is as follows:

If the system output is of the same sign at the same time as the time-varying supply power, the measurand must have been positive. If the signs are opposite, the measurand must have been negative.

Phase sensing is accomplished by a phase-sensitive demodulator. A reference signal is fed from the bridge supply power to the phase-sensitive demodulator. This signal is compared with the amplitude-modulated signal for phase determination. A half-wave rectifier with transformer coupled reference and amplitude-modulated signals forms one basis for a demodulator.

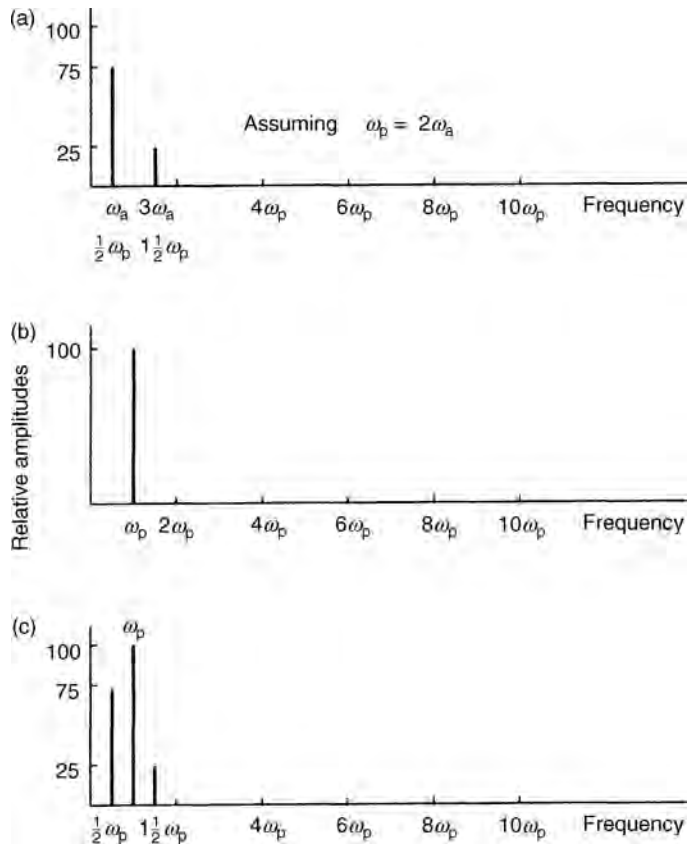


FIGURE 48.30 Signal frequency contents, AC bridge. (Adapted with permission from “Information as a ‘Noise Suppression’ Method,” by Peter Stein, Stein Engineering Services, Inc., Phoenix, AZ, LR/MSE Publ. 66, 1975.)

After phase sensing, a low-pass filter is required to separate the nonself-generating analog signal proportional to the measurand from the other frequencies which appear as sidebands around harmonics of power supply frequency. Final selection of a power supply frequency is a trade-off between the maximum frequency content in the measurand and the low-pass filter roll-off characteristics. While a 10:1 ratio is typical, the power supply frequency may vary between 3 and 20 times the maximum nonself-generating signal frequency.

Note that introduction of an AC power supply requires a bridge-balancing network incorporating complex impedance in the balance controls.

48.8 AC IMPEDANCE BRIDGE TRANSDUCERS

Having discussed bridge transducers that use resistive sensing elements and AC power supplies (amplitude modulation), a logical question is whether bridge

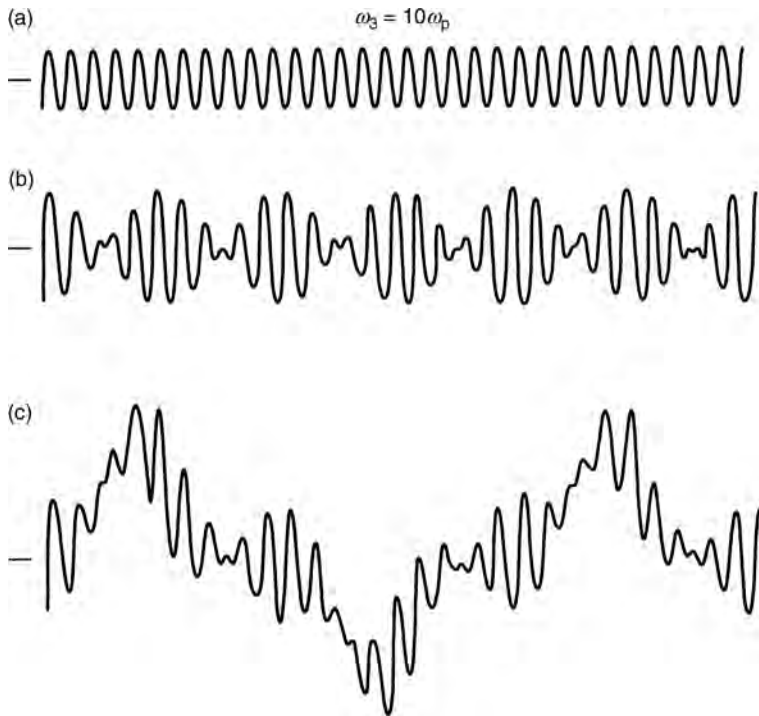


FIGURE 48.31 Signal wave shapes, AC bridge. (Adapted with permission from “Information as a ‘Noise Suppression’ Method,” by Peter Stein, Stein Engineering Services, Inc., Phoenix, AZ, LR/MSE Publ. 66, 1975.)

transducer sensing elements can be capacitive or inductive. In practice, many bridge transducers do employ capacitive or inductive elements. While resistance bridge-type transducers typically possess resonant frequencies in the tens or hundreds of kilohertz range, AC impedance bridge transducers typically possess resonant frequencies of less than 10 kHz. The larger physical size of AC impedance bridge transducers makes them environmentally more fragile but improves their performance by increasing their sensitivity to low-level measurands.

48.8.1 Inductive Bridges

Figure 48.34 shows an example of how variable-reluctance sensing elements can be incorporated into an AC bridge transducer. A differential pressure transducer containing a magnetically permeable stainless steel diaphragm as the mechanical flexure is portrayed. This diaphragm is clamped between two blocks and deflects when a pressure difference is created across it through the two ports shown. An E-core and coil assembly is embedded in each block. A small gap exists in front of each E-core. When the diaphragm is undeflected, a condition of equal inductance exists in each coil. When the diaphragm does deflect, an increase of gap in the magnetic flux path of one core occurs, with a resultant

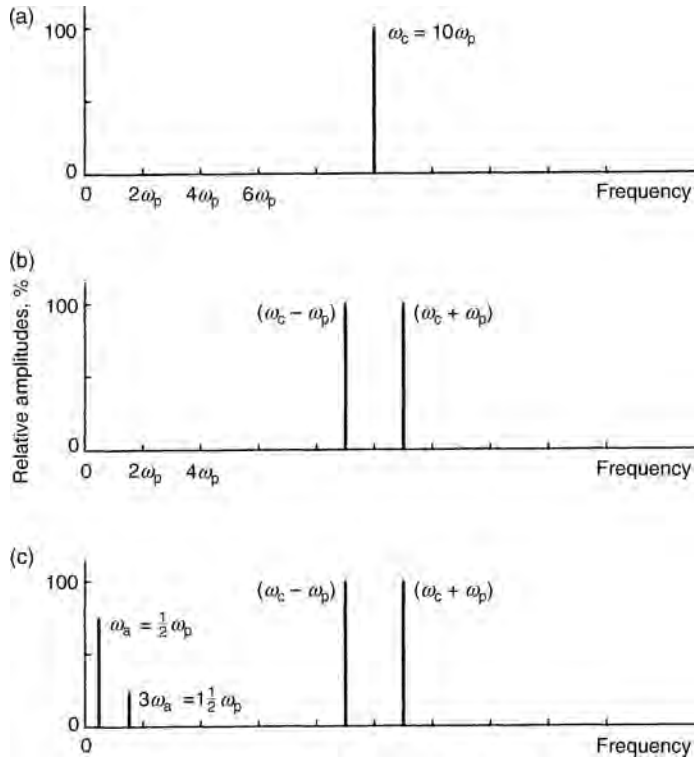


FIGURE 48.32 Signal frequency contents, AC bridge. (Adapted with permission from “Information as a ‘Noise Suppression’ Method,” by Peter Stein, Stein Engineering Services, Inc., Phoenix, AZ, LR/MSE Publ. 66, 1975.)

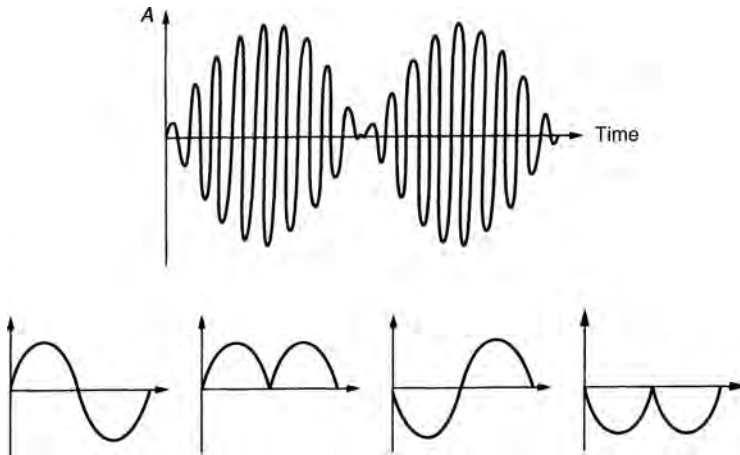


FIGURE 48.33 Problems of phase sensing. (Adapted with permission from “Information as a ‘Noise Suppression’ Method,” by Peter Stein, Stein Engineering Services, Inc., Phoenix, AZ, LR/MSE Publ. 66, 1975.)

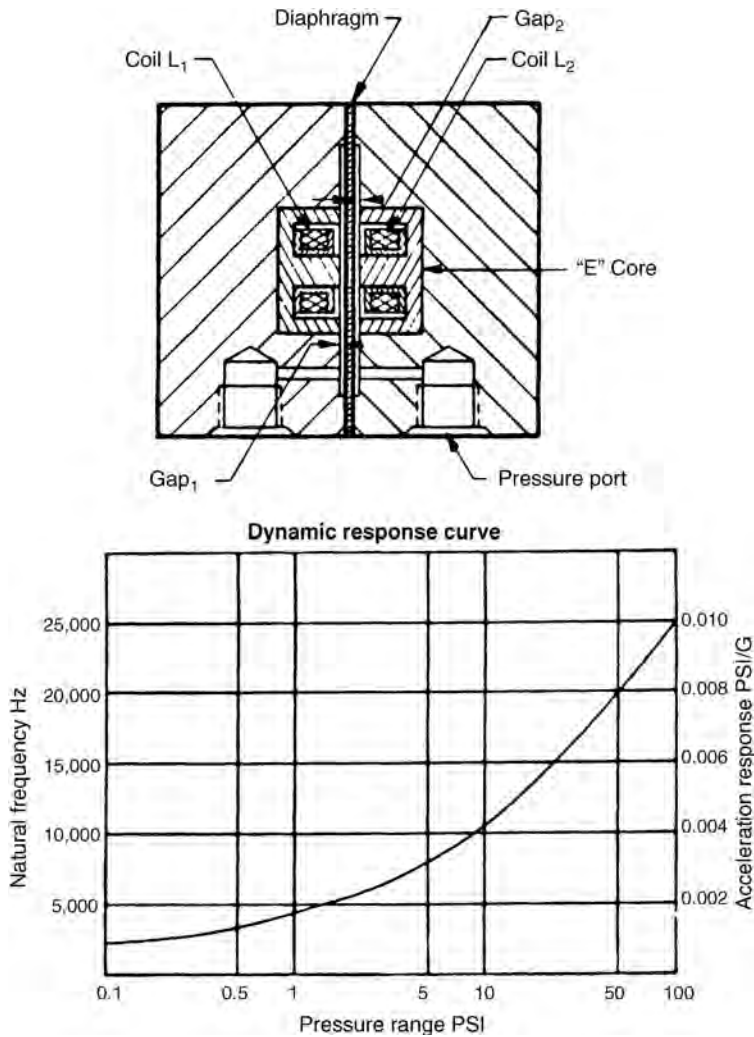


FIGURE 48.34 Variable reluctance sensing in a bridge transducer. (Courtesy of Validyne Engineering Corporation.)

decrease in the gap in the magnetic flux path of the other. Magnetic reluctance varies with gap, determining the inductance value. The diaphragm motion then changes the inductance of the two coils, one increasing and one decreasing. These two coils can be placed in adjacent arms of an AC-powered bridge. Resistive elements can be used to complete the bridge. Once the bridge is balanced, an amplitude-modulated signal results when a differential pressure is applied across the ports of the transducer. When the resultant signal is properly demodulated, the applied pressure can be quantified.

Eddy current inductive displacement-measuring systems are another example of the use of impedance as opposed to resistive bridges. Placing a coil with an AC flowing in it a nominal distance from a metal target induces a current flow on the surface and within the

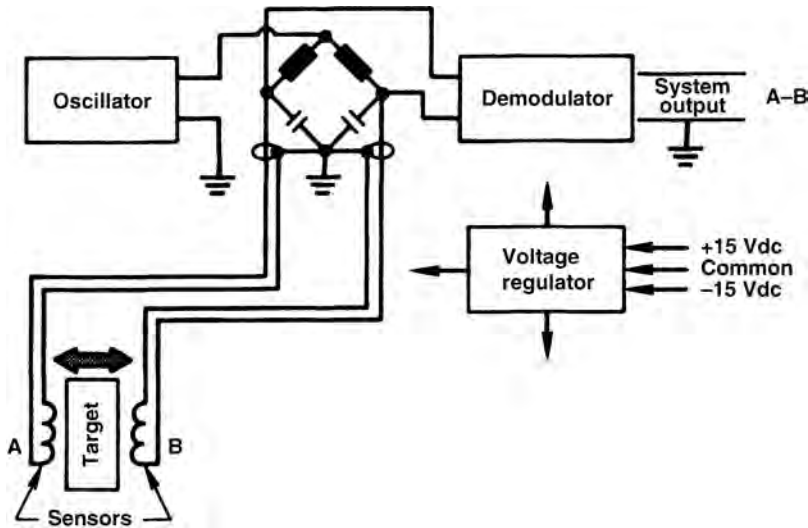


FIGURE 48.35 Eddy current inductive displacement measuring system. (Courtesy of Kaman Instrumentation Corporation, Measurement Systems Group, Colorado Springs, CO.)

target. The induced current produces a secondary magnetic field that opposes and reduces the intensity of the first field. Changes in the impedance of the exciting coil provide information about the target. The target can be the diaphragm of a pressure transducer, the seismic mass of an accelerometer, the flexure of a load cell, and so on. The coil is one leg of a balanced bridge network. Unbalanced bridge conditions are sensed and converted into a signal directly proportional to the distance between coil and target. Figure 48.35 schematically illustrates this conversion.

The electrical parameters of resistivity and permeability in the target material influence performance of eddy current transducers. For a specific material, displacement sensitivity is influenced by coil geometry and operating frequency.

Target thickness is generally not a limiting factor. At one “skin depth,” the eddy current density is only 36% of the maximum encountered on the target surface; at two “skin depths,” it is 13%. Figure 48.36 defines skin depth as a function of target resistivity and permeability.

Target shape and alignment also should be considered in application. A flat, circular target equal to the coil diameter appears as an infinite plane. Smaller target diameters produce smaller voltage unbalances in the impedance bridge. Since the transducer senses the average distance to the target, the nonparallelism effect is small up to 15° .

A differential transformer also is briefly mentioned here, although it does not operate in an impedance bridge. A linear variable differential transformer (LVDT) consists of three symmetrically spaced coils wound onto an insulated bobbin. A magnetic core moving through the bobbin provides a path for magnetic flux linkage between coils. The center coil is the primary and has an AC voltage applied. The two secondary coils are wired in a series-opposing circuit. When the core is centered between two secondary coils, the voltages in the two coils cancel. As the core is displaced, the phase-referenced and demodulated output signal provides a linear voltage output with displacement.

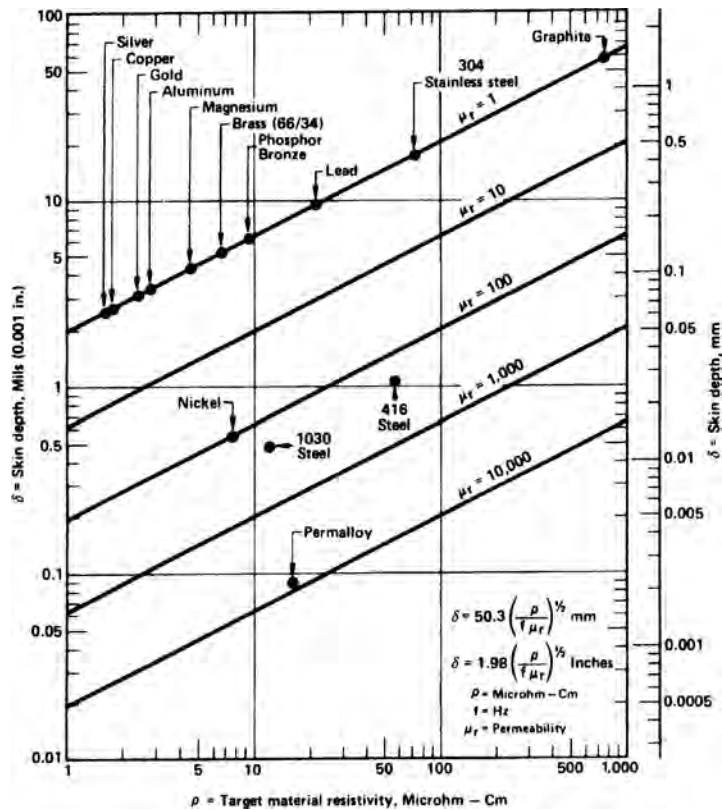


FIGURE 48.36 Skin depth versus target resistivity and permeability at 1 MHz. (Courtesy of Kaman Instrumentation Corporation, Measurement Systems Group, Colorado Springs, CO.)

48.8.2 Capacitive Bridges

Capacitance sensors can be integrated into bridge transducers. The capacitance between two metal plates separated by an air gap is $C = kKA/h$, where C is capacitance, K is the dielectric constant for the material between the plates, A is the plate overlapping area, h is the gap thickness between the two plates, and k is a proportionality constant. The range of a capacitance sensor can be shown to remain linear with changes in area but become nonlinear when the change in gap displacement becomes a significant portion of the original gap.

Again, it should be pointed out that advantages of small size and enhanced dynamic response are to be found with resistance bridge transducers. Increased sensitivity will be displayed by impedance bridge transducers. Both impedance bridges and AC-powered resistance bridges offer noise suppression through separating nonself-generating responses from self-generating responses.

REFERENCES

- Dove RC, Adams PH. *Experimental Stress Analysis and Motion Measurements*. Columbus (OH): Charles E. Merrill Books; 1964.

- End-to-End Test Methods for Telemetry Systems. Document 118-79. In: *Test Methods for Telemetry Systems and Subsystems*. Vol. 1, White Sands Missile Range, NM: Secretariat, Range Commanders Council; 1979.
- Jaquay JW. Understanding DC instrumentation amplifiers. In: *Instruments and Controls*. September; 1972.
- Murray WM, Stein PK. *Strain Gage Techniques*. Engineering Extension UCLA and Society for Experimental Stress Analysis; 1953.
- Perry CC, Lissner HR. *The Strain Gage Primer*. 2nd ed. New York: McGraw-Hill; 1962.
- Stein PK. *Measurement Engineering*. Phoenix (AZ): Stein Engineering Services; 1964.
- Stein PK. *A Unified Approach to Handling of Noise in Measuring Systems*. Nevilly-sur-Seine, France: AGARD LS-50, NATO; September 1972.
- Stein PK. *Information Conversion as a Noise Suppression Method*. Phoenix (AZ): Lf/MSE Publication 66, Stein Engineering Services; 1975.
- TECHNICAL DATA TD4354-1. *When and How—Semiconductor Strain Gages*. BLH Electronics, Waltham, MA, June; 1975.
- Temperature Compensation of Bridge Type Transducers. Statham Instrument Notes Number 5, Gould, Oxnard, CA, August; 1951.
- Test Methods for Vehicle Telemetry Systems. Document 118-03. In: *Test Methods for Telemetry Systems and Subsystems*. Vol. 5. White Sands Missile Range, NM: Secretariat, Range Commanders Council; 2003.
- Walter PL. Deriving the transfer function of spatial averaging transducers. *ISA Transactions* 1980;19(3).
- Window AL, Hollister GS. *Strain Gauge Technology*. London and New York: Applied Science Publishers; 1982.

FURTHER READINGS

- Dally, JW, Riley WF, McConnell K. *Instrumentation for Engineering Measurements*. New York: Wiley; 1988.
- Doebelin, EO. *Measurement Systems: Application and Design*. McGraw-Hill Science-/Engineering/Math; 5th Pkg ed. ISBN: 007292201X June 4; 2003.

49

SIGNAL PROCESSING

JOHN TURNBULL

- 49.1 Frequency-domain analysis of linear systems
- 49.2 Basic analog filters
 - 49.2.1 Butterworth
 - 49.2.2 Tchebyshev
 - 49.2.3 Inverse Tchebyshev
 - 49.2.4 Elliptical
 - 49.2.5 Arbitrary frequency response curve fitting by method of least squares
 - 49.2.6 Circuit prototypes for pole and zero placement for realization of filters designed from rational functions
- 49.3 Basic digital filter
 - 49.3.1 z -transforms
 - 49.3.2 Design of FIR filters
 - 49.3.3 Design of IIR filters
 - 49.3.4 Design of various filters from low-pass prototypes
 - 49.3.5 Frequency-domain filtering
- 49.4 Stability and phase analysis
 - 49.4.1 Stability analysis
 - 49.4.2 Phase analysis
 - 49.4.3 Comparison of FIR and IIR filters
- 49.5 Extracting signal from noise
- References

49.1 FREQUENCY-DOMAIN ANALYSIS OF LINEAR SYSTEMS

Signals are any carriers of information. Our objective in signal processing involves the encoding of information for the purpose of transmission of information or decoding the information at the receiving end of the transmission. Unfortunately, the signal is often

corrupted by noise during our transmission, and hence it is our objective to extract the information from the noise. The standard method most commonly used for this involves filters that exploit some separation of the signal and noise in the frequency domain. To this end, it is useful to use frequency-domain tools such as the Fourier transform and the Laplace transform in designing and analyzing various filters. The Fourier transform of a function of a time is

$$\mathcal{F}\{f(t)\} = F(\omega) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} f(t) e^{j\omega t} dt \quad j^2 = -1. \quad (49.1)$$

For continuous systems, the transfer characteristics of a filter system is a function that gives information of the gain versus frequency. The Laplace transform for a given time-domain function is

$$\mathcal{L}\{f(t)\} = F(s) = \int_0^{\infty} f(t) e^{-st} dt. \quad (49.2)$$

The steady-state Laplace transform (i.e., neglecting transients) for the derivative and integral of a given function is

$$\mathcal{L}\left\{\frac{df(t)}{dt}\right\} = sF(s) \quad \mathcal{L}\left\{\int f(t) dt\right\} = \frac{F(s)}{s}. \quad (49.3)$$

By convention, functions in the time domain use t as the independent variable and functions in the Laplace domain use s as the independent variable. For this reason, the Laplace domain is commonly called the *S-domain*. The Fourier transform and the Laplace transform are similar but different in two respects: (1) the Fourier transform integrates the signal over all time while the Laplace transform integrates for positive times only and (2) the exponent of the kernel in the Laplace transform is complex with both real and imaginary components while the exponent in the Fourier transform has imaginary component and no real component. Using Euler's identity, $e^{j\omega} = \cos(\omega) + j\sin(\omega)$, we see that

$$\begin{aligned} \mathcal{F}\{f(t)\} &= F(\omega) = \int_{-\infty}^{\infty} f(t) e^{j\omega t} dt \\ &= \int_{-\infty}^{\infty} f(t) \cos(\omega t) dt + j \int_{-\infty}^{\infty} f(t) \sin(\omega t) dt \\ &= \langle \cos(\omega t) \rangle + j \langle \sin(\omega t) \rangle. \end{aligned} \quad (49.4)$$

We approximate the Fourier transform from discrete samples $f(k) \leftrightarrow F(k)$, where $0 \leq k \leq N$ and $F(k) = \alpha_k + j\beta_k$:

$$\alpha_k = \sum_{i=0}^{N-1} f(i) \cos\left(2\pi \frac{i \times k}{N}\right) \quad \beta_k = \sum_{i=0}^{N-1} f(i) \sin\left(2\pi \frac{i \times k}{N}\right). \quad (49.5)$$

However, if the number of points we transform is a composite number and not a prime number, we can restructure our calculations to eliminate some of the calculations. Furthermore, of the factors that are themselves composite, we can further factor and eliminate more calculations. For this reason, the most efficient vector sizes are those that are highly composite. As an example, consider the simple case of transforming four points. We can express Equation (49.5) for this case in matrix form as

$$\begin{bmatrix} c_0 \\ c_1 \\ c_2 \\ c_3 \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & \rho & \rho^2 & \rho^3 \\ 1 & \rho^2 & \rho^4 & \rho^6 \\ 1 & \rho^3 & \rho^6 & \rho^9 \end{bmatrix} \begin{bmatrix} f_0 \\ f_1 \\ f_2 \\ f_3 \end{bmatrix},$$

where ρ is the principal root of 1. The n th principal root of 1 is $\cos(2\pi/n) + j \sin(2\pi/n)$. We can then factor the matrix into two sparse matrices:

$$\begin{bmatrix} c_0 \\ c_1 \\ c_2 \\ c_3 \end{bmatrix} = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & \rho^2 & 0 & 0 \\ 0 & 0 & 1 & \rho \\ 0 & 0 & 1 & \rho^3 \end{bmatrix} \begin{bmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 1 & 0 & \rho^2 & 0 \\ 0 & 1 & 0 & \rho^2 \end{bmatrix} \begin{bmatrix} f_0 \\ f_1 \\ f_2 \\ f_3 \end{bmatrix}.$$

Although it is possible to implement this efficient algorithm—known as the fast Fourier transform (FFT)—for any vector size that is a composite number, it is most commonly implemented for vector sizes in which all factors are 2. The effort in evaluating Equation (49.5) increases with the square of the number of points to transform, or $O(n^2)$. In contrast, the effort for the FFT of order 2^p increases proportionately to $O(n \log(n))$. Finally, we can use the FFT to estimate the power spectral density (PSD) of a given discrete signal by computing the square of the magnitude of the FFT. The following algorithm describes the method for computing the fast Fourier transform (Press et al., 1988).

FAST FOURIER TRANSFORM 1 *Given a vector $\{x_1, \dots, x_n\}$ of complex numbers, where n is some integer and there exists some integer p such that $n = 2^p$. This algorithm outputs the Fourier transform overwriting the input vector \mathbf{x} .*

```

k ← 1
For i = 1 To n
    If k > i
        swap (xi, xk)
    End If
    m = n/2
    While m ≥ 1 And k > m
        k ← k - m
        m ← m/2
    End While
    k ← k + m
Next i
Mmax ← 1
    
```



```

While  $n > M_{max}$ 
     $i_c \leftarrow 2 \cdot M_{max}$ 
     $\theta \leftarrow \pi / M_{max}$ 
     $w_p \leftarrow -2 \sin^2(\theta/2) - j \cdot \sin(\theta)$ 
     $w \leftarrow 1$ 
    For  $m = 1$  To  $M_{max}$ 
        For  $i = m$  To  $n$  By  $i_c$ 
             $k \leftarrow i + M_{max}$ 
             $x_{temp} \leftarrow x_i - w \cdot x_k$ 
             $x_i \leftarrow x_i + w \cdot x_k$ 
             $x_k \leftarrow x_{temp}$ 
        Next  $i$ 
         $w \leftarrow w \cdot (w_p + 1)$ 
    Next  $m$ 
     $M_{max} = i_c$ 
End While

```

49.2 BASIC ANALOG FILTERS

Linear filters apply frequency-specific gains to a signal. This is often done to enhance desired portions of the spectrum while attenuating or eliminating other portions. Four common filters are low pass, high pass, bandpass, and band reject. The objective of an ideal low-pass filter is to eliminate a range of undesired high frequencies from a signal and leave the remaining portion undistorted. To this end, an ideal low-pass filter will have a gain of 1 for all frequencies less than some desired cutoff frequency f_c and a gain of 0 for all frequencies greater than f_c , as seen in Figure 49.1. There are various rational functions that approximate this ideal. But because of the discontinuity in the ideal low-pass response, all realizations of this ideal will be an approximation. The various approximation functions generally trade off between three characteristics: passband ripple,

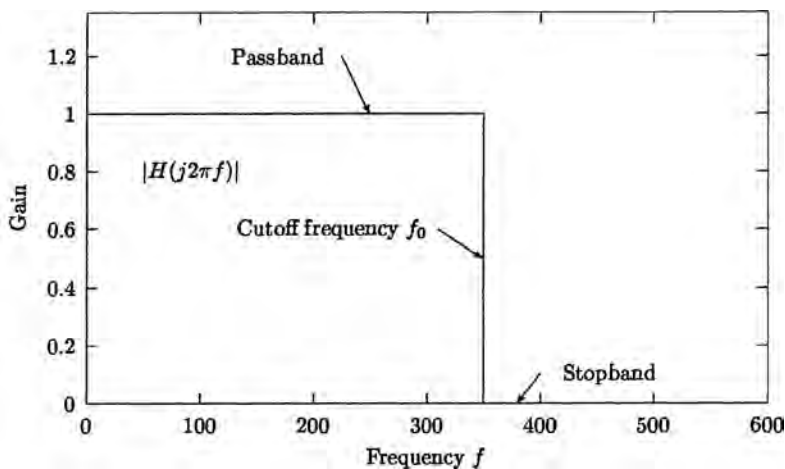


FIGURE 49.1 Frequency response for an ideal low-pass filter.

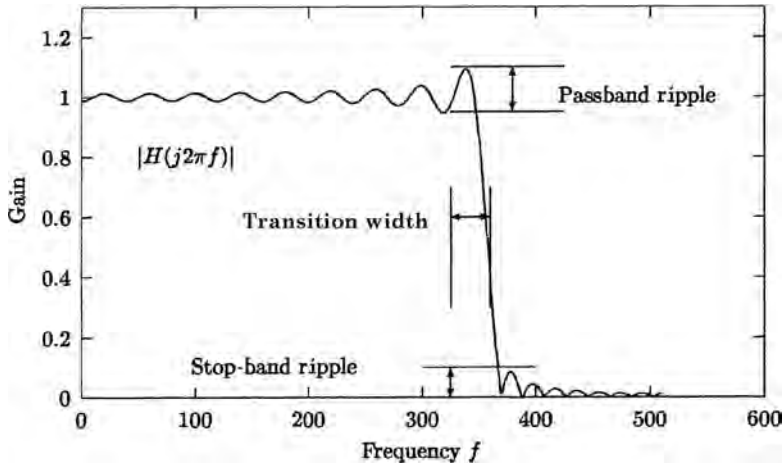


FIGURE 49.2 Typical frequency response for an approximate low-pass filter.

stop-band ripple, and the transition width, shown in Figure 49.2. Four common rational function approximations for low-pass filters are the Butterworth, the Tchebyshev Types I and II, and the elliptical filter. By convention, we use $H(s)$ as the transfer function from which we determine the frequency response, where

$$H(s) = \frac{V_{\text{output}}(s)}{V_{\text{input}}(s)} \quad (49.6)$$

is the output-to-input gain of a standard two-port system.

49.2.1 Butterworth

The Butterworth filter has a smooth passband region (frequencies less than f_c hertz) and a smooth stop band (frequencies greater than f_c) and a comparatively wide transition region as shown in Figures 49.3 and 49.4. Let s_c be the cutoff frequency; then a low-pass rational function approximation is as follows:

$$|H(s)|^2 = \frac{1}{1 + (s/s_c)^{2N}}, \quad (49.7)$$

where $2\pi f_c = s_c$. In factored form

$$H(s) = \frac{1}{(s - p_0)(s - p_1) \cdots (s - p_{N-1})}, \quad (49.8)$$

where $p_i = \alpha_i + j\beta_i$ and

$$\begin{aligned} \alpha_i &= 2\pi f_c \cos\left(\frac{\pi}{2N}(N + 2i + 1)\right) & i = 0, \dots, N - 1 \\ \beta_i &= 2\pi f_c \sin\left(\frac{\pi}{2N}(N + 2i + 1)\right) & i = 0, \dots, N - 1 \end{aligned} \quad (49.9)$$

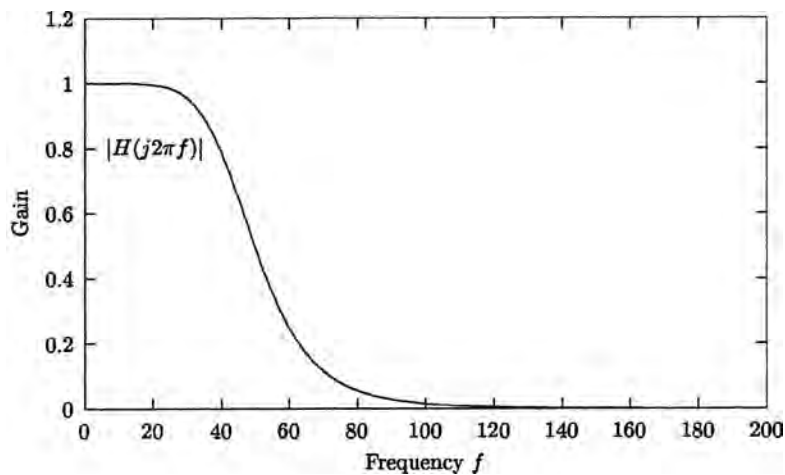


FIGURE 49.3 Frequency response to a third-order Butterworth filter.

49.2.2 Tchebyshev

Unlike the Butterworth rational function, the Tchebyshev (Type I) rational function permits some ripple to occur in the passband (frequencies less than f_c in the low-pass filter), in exchange for a sharper transition region compared to a Butterworth filter of equal order N , as seen in Figure 49.5:

$$|H(s)|^2 = \frac{1}{1 + \epsilon^2 T_N^2(s/s_c)}, \tag{49.10}$$

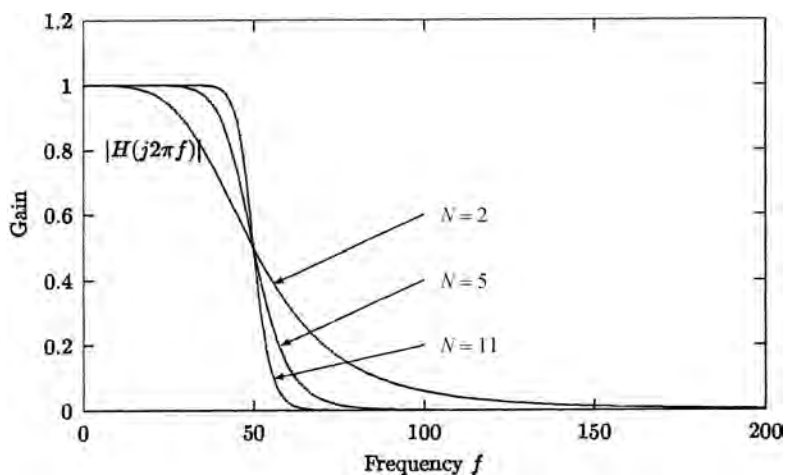


FIGURE 49.4 Frequency response to Butterworth filters of order 2, 5, and 11.

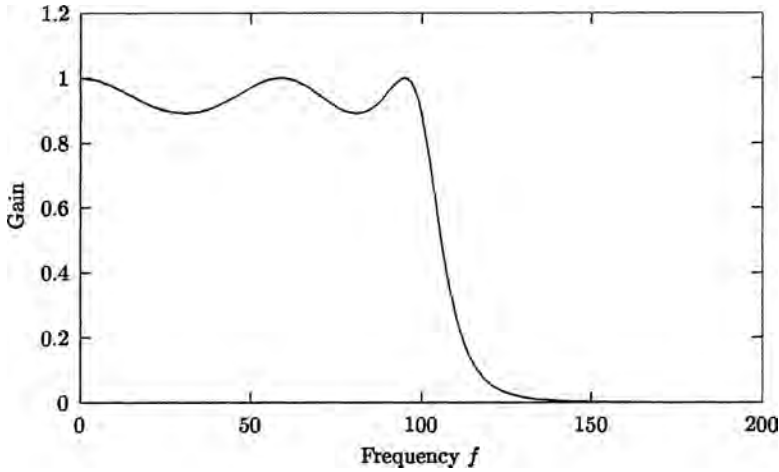


FIGURE 49.5 Frequency response to a fifth-order Tchebyshev filter.

where T_N is the Tchebyshev polynomial defined as

$$T_N(s) = \cos[n \cos^{-1}(s)]. \quad (49.11)$$

The pole placements $p_i = \alpha_i + j\beta_i$ for the Tchebyshev Type I filter are

$$\begin{aligned} \alpha_i &= 2\pi f_c \sinh(v_0) \cos\left(\frac{\pi}{2N}(N+2i+1)\right) \quad i = 0, \dots, N-1 \\ \beta_i &= 2\pi f_c \cosh(v_0) \sin\left(\frac{\pi}{2N}(N+2i+1)\right) \quad i = 0, \dots, N-1 \end{aligned}, \quad (49.12)$$

where

$$v_0 = \frac{1}{N} \sinh^{-1}\left(\frac{1}{\epsilon}\right), \quad (49.13)$$

$$\epsilon = \sqrt{\frac{1}{(1-r)^2} - 1} \quad 0 < r < 1, \quad (49.14)$$

where r is this amplitude of the ripple in proportion to the gain of the passband.

49.2.3 Inverse Tchebyshev

The inverse Tchebyshev (or Tchebyshev Type II) filter has a smooth passband, ripple in the stop band (frequencies less than f_c for the low-pass filter), and a sharper transition region compared to the Butterworth function of equal order N , as seen in Figure 49.6:

$$|H(s)|^2 = \frac{\epsilon^2 T_N^2(s_c/s)}{1 + \epsilon^2 T_N^2(s_c/s)}. \quad (49.15)$$

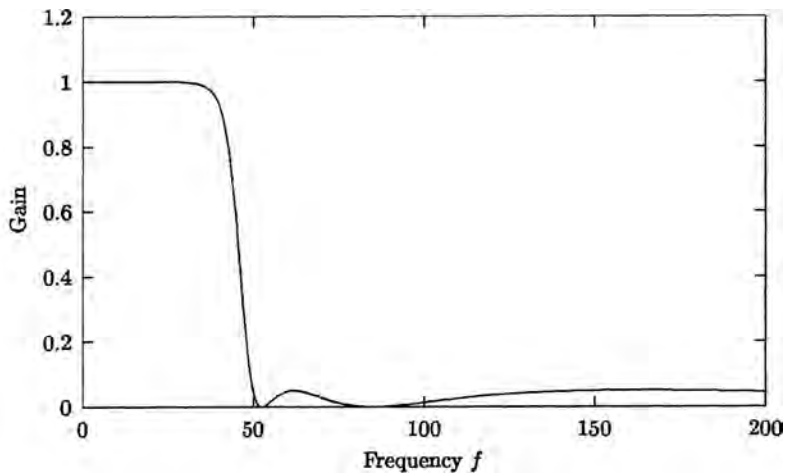


FIGURE 49.6 Frequency response to a fifth-order inverse Tchebyshev filter.

The zero placements $\zeta_i = \alpha_i + j\beta_i$ for the Tchebyshev Type II filter are

$$\zeta_i = \frac{1}{\sin(i\pi/2N)} \quad 0 \leq i \leq N-1. \quad (49.16)$$

The pole placements $p_i = \alpha_i + j\beta_i$ for the Tchebyshev Type II filter are simply the reciprocals of the pole placements computed for the Tchebyshev Type I filter.

49.2.4 Elliptical

The elliptical filter has ripple in both the passband and the stop band but, in exchange, has the narrowest transition region for equal filter order N , as seen in Figure 49.7. The

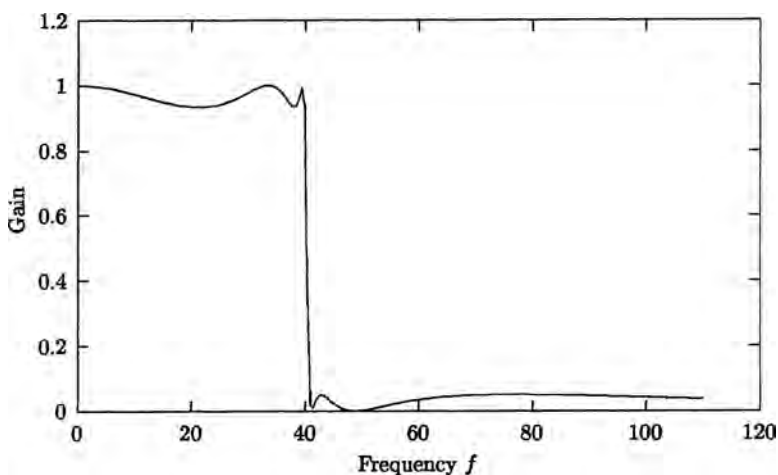


FIGURE 49.7 Frequency response to a fifth-order elliptical filter.

derivation and implementation for the determination of the poles and zeros involve the Jacobian elliptic function:

$$f(t, k) = \int_0^t \frac{dx}{\sqrt{1 - k^2 \sin^2(x)}}. \quad (49.17)$$

The method is beyond the scope of this chapter. The interested reader is referred to Parks and Burrus (1987); Vlček and Unbehauen (1989).

49.2.5 Arbitrary Frequency Response Curve Fitting by Method of Least Squares

It is possible to design a filter to approximate a desired frequency response $F(\omega)$ by the method of least squares. Consider a transfer function in factored form as

$$H(s) = G \frac{\prod_{i=1}^M s - \zeta_i}{\prod_{k=1}^N s - p_k}. \quad (49.18)$$

Minimize

$$\chi^2(G; \boldsymbol{\zeta}; \mathbf{p}) = \int_D [|H(j\omega)| - F(\omega)]^2 d\omega, \quad (49.19)$$

subject to

$$\Re(p_k) < 0 \quad \text{for all } k = 1, \dots, N. \quad (49.20)$$

Unfortunately, this results in a nonlinear system of equations. Furthermore, the topography of this objective function is generally complicated with many local minima, making standard gradient-descent methods unfeasible. It is usually best to use finite-impulse-response (FIR) filtering or filtering in the frequency domain for these types of problems. If an infinite-impulse-response (IIR) filter is desired, the reader is referred to Prony's method, which linearizes this system and finds an approximate optimal solution (Marple, 1987).

49.2.6 Circuit Prototypes for Pole and Zero Placement for Realization of Filters Designed from Rational Functions

The voltage–current relation for a resistor (R), inductor (L), and capacitor (C) is

$$v_r = iR \quad v_l = L \frac{di}{dt} \quad v_c = \frac{1}{C} \int_{-\infty}^t i \, dr \rightarrow . \quad (49.21)$$

These relationships are represented in the S -domain as

$$V_r(s) = RI(s) \quad V_l(s) = sLI(s) \quad V_c = \frac{I(s)}{sC}. \quad (49.22)$$

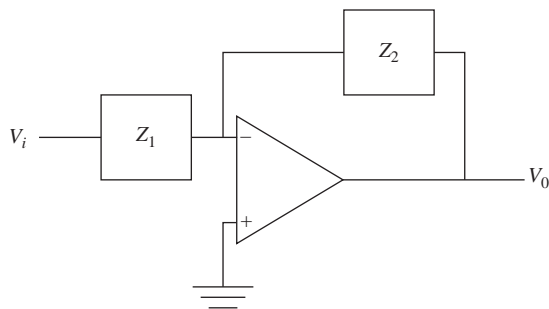


FIGURE 49.8 Prototype circuit element for construction of analog filter.

Thus, the general transfer function for any linear circuit involving standard passive, active, and reactive devices is a rational function, that is, a ratio of polynomials:

$$\begin{aligned} H(s) &= \frac{\sum_{i=0}^{M-1} a_i s^i}{\sum_{k=0}^{N-1} b_k s^k} \\ &= \frac{V_0}{V_i} = -\frac{Z_2}{Z_1} \end{aligned} \tag{49.23}$$

Thus one can construct any arbitrary transfer function through a serial placement of this building block circuit prototype shown in Figure 49.8. Table 49.1 gives circuit elements for Z_1 and Z_2 to construct the basic prototype circuits.

49.3 BASIC DIGITAL FILTER

Basic linear digital filters are of two types: those that have a finite response to an impulse, or FIR, and those that have an infinite response to an impulse (IIR). The general form of a linear digital filter is

$$y_k = b_1 y_{k-1} + b_2 y_{k-2} + \cdots + b_{k-m} y_{k-m} + a_0 x_k + a_1 x_{k-1} + \cdots + a_n x_{k-n}, \tag{49.24}$$

where $k - i$ represents the i th delay. That is, the k th output from a linear digital filter is some linear combination of previous inputs and outputs. The filter will have a finite response to an impulse if $b_1 = b_2 = \cdots = b_m = 0$; otherwise, the filter is of type IIR.

TABLE 49.1 Circuit Elements for the Construction of Basic Transfer Function Prototypes

Single pole	$Z_1 \leftarrow$ resistor $Z_2 \leftarrow RC$ in parallel
Single zero	$Z_1 \leftarrow RC$ in parallel $Z_2 \leftarrow$ resistor
Complex-conjugate pole pair	$Z_1 \leftarrow LRC$ in series $Z_2 \leftarrow$ capacitor
Complex-conjugate zero pair	$Z_1 \leftarrow$ capacitor $Z_2 \leftarrow LRC$ series

49.3.1 z -Transforms

The z -transform is used to analyze the frequency response and stability of a system of difference equations in much the same way that the Laplace transform is used to analyze the frequency response and stability of a system of differential equations. The z -transform of Equation (49.24) is

$$H(z) = \frac{a_0 + a_1 z^{-1} + \cdots + a_M z^{-M}}{1 - b_1 z^{-1} - \cdots - b_N z^{-N}} = \frac{\sum_0^M a_j z^{-j}}{1 - \sum_{j=1}^N b_j z^{-j}}. \quad (49.25)$$

We determine the frequency response by $|H(e^{j\omega})|$, where $0 \leq \omega \leq 2\pi$ and corresponds to the scaled frequencies of our sampled system from 0 Hz up to the sampled frequency. The function $e^{j\omega}$ is a periodic signal. For $\pi < \omega \leq 2\pi$, $e^{j\omega} = e^{j(\pi - \omega)}$, and for $2\pi < \omega \leq 4\pi$, $e^{j\omega} = e^{j(\omega - 2\pi)}$. Thus, frequencies greater than π , corresponding to the Nyquist frequency, or one-half of the sampling frequency assume an identical characteristic to an analogous frequency less than π . This phenomenon is called *aliasing* and is illustrated in Figure 49.9.

49.3.2 Design of FIR Filters

It is possible to use the Fourier transform to determine the coefficients to an FIR filter. However, the Fourier coefficients are generally complex numbers, and when working with real signals, it is desirable to have a real coefficients in our filter. To do this, we apply Euler's identity and observe from Equation (49.4) that the coefficients will be real if the inner products with the sine function are all zero. We can artificially construct our desired frequency response so that this will be so. To do this, suppose $F(\omega)$ is the desired frequency response where $0 \leq \omega \leq \pi$.

Step 1. Augment the domain of the function over $0 \leq \omega \leq 2\pi$.

Step 2. Augment the function values from $\pi \leq \omega \leq 2\pi$ as $F(\omega) = F(2\pi - \omega)$.

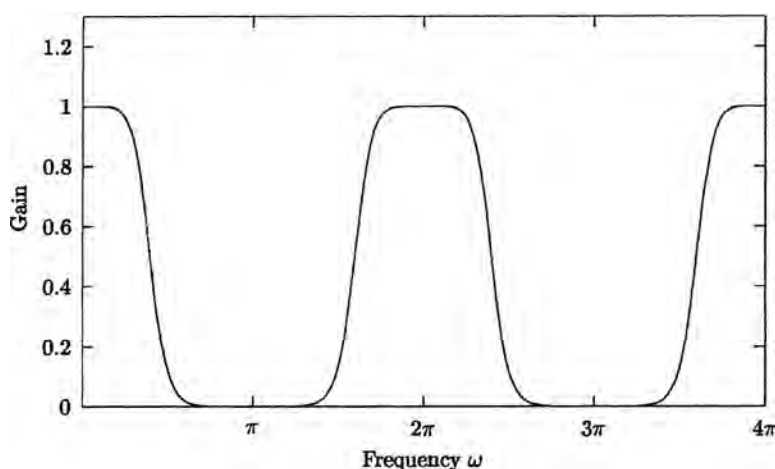


FIGURE 49.9 Demonstration of aliasing from digital filter response.

Step 3. Compute the discrete cosine Transform [α_i coefficients from Equation (49.5)] over the range $0 \leq \omega \leq 2\pi$.

For a discrete system with a desired FIR filter of length N :

Step 1. Augment the domain of the function over $1 \leq i \leq 2N$.

Step 2. Augment the function values from $N + 1 \leq i \leq 2N$ as $F(i) = F(2N - i)$.

Step 3. Construct the FIR filter with $2N$ points from the discrete cosine transform of the desired frequency response.

Step 4. Keep the first N coefficients and truncate the remaining coefficients.

The discrete cosine transform of an ideal low-pass filter is a sinc function, defined as

$$\text{sinc}(x) \begin{cases} \frac{\sin(\pi x)}{\pi x} & \text{if } x \neq 0 \\ 1 & \text{if } x = 0 \end{cases}. \quad (49.26)$$

Therefore, the coefficients to a low-pass FIR filter with cutoff frequency f_c and length $2N - 1$ are determined using Equation (49.27), where ω_c is the desired cutoff frequency divided by the Nyquist frequency:

$$h(i) = \omega_c \text{sinc} \left[\omega_c \left(i - \left\lfloor \frac{N}{2} \right\rfloor \right) \right] \quad 1 \leq i \leq N, \quad (49.27)$$

where $\lfloor x \rfloor$ is the greatest integer less than or equal to x and ω_c is the desired cutoff frequency divided by the Nyquist frequency.

49.3.2.1 Windowing This FIR filter will have ripple in the passband and in the stop band. It is possible to suppress these ripples and smooth the frequency response, but the trade-off will be an increased transition width. The method for suppressing these ripples is with the application of a windowing function. There is a large class of windowing functions that allow the designer to determine how he or she wishes to trade off the transition width and how much ripple is to be tolerated. The design of an FIR filter with windowing involves the use of Equation (49.26) for the determination of the FIR coefficients followed by the component-by-component product of the coefficients with the windowing values, that is, $h'(i) = h(i) \text{win}(i)$. Below is a list of common windows (Oppenheim and Schaffer, 1975):

Rectangular:

$$\text{win}(i) = 1 \quad 0 \leq i \leq N - 1. \quad (49.28)$$

Bartlett:

$$\text{win}(i) = \begin{cases} \frac{2i}{N-1} & 0 \leq i \leq \frac{N-1}{2} \\ 2 - \frac{2i}{N-1} & \frac{N-1}{2} \leq i \leq N-1 \end{cases}. \quad (49.29)$$

TABLE 49.2 Comparison of Characteristics for Commonly Used Windowing Functions

Window Name	Minimum Stop-Band Attenuation (dB)
Rectangular	-21
Bartlett	-25
Hanning	-44
Hamming	-53
Blackman	-74

Hanning:

$$\text{win}(i) = \frac{1}{2} \left[1 - \cos\left(\frac{2\pi i}{N-1}\right) \right] \quad 0 \leq i \leq N-1. \quad (49.30)$$

Hamming:

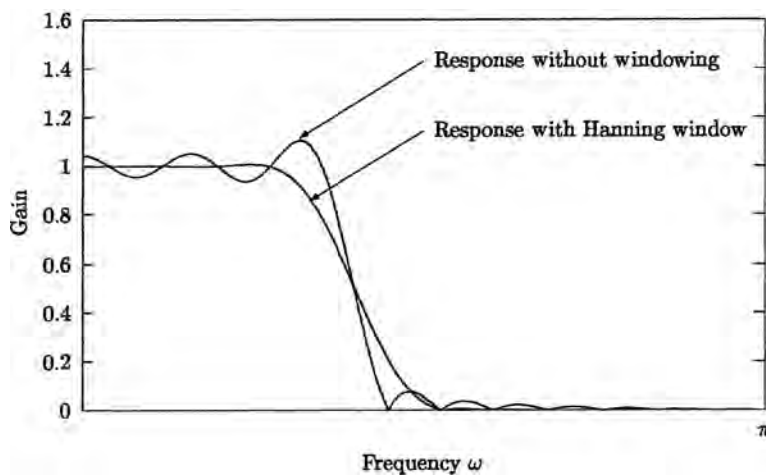
$$\text{win}(i) = 0.54 - 0.46 \cos\left(\frac{2\pi i}{N-1}\right) \quad 0 \leq i \leq N-1. \quad (49.31)$$

Blackman:

$$\text{win}(i) = 0.42 - 0.5 \cos\left(\frac{2\pi i}{N-1}\right) + 0.08 \cos\left(\frac{4\pi i}{N-1}\right) \quad 0 \leq i \leq N-1. \quad (49.32)$$

Table 49.2 gives a list of several common windowing functions together with their characteristics (Oppenheim and Schaffer, 1975).

Figure 49.10 demonstrates the effect of a Hanning window.

**FIGURE 49.10** Comparison of FIR filter with Hanning window and with windowing.

49.3.2.2 FIR High-Pass and Bandpass Design The design of a high-pass filter is simply $1 - H(z)$. In the time domain, this is

$$h(i) = \omega_c \operatorname{sinc} \left[\omega_c \left(i - \left\lfloor \frac{N}{2} \right\rfloor \right) \right] - \omega_c \delta \left(\left\lfloor \frac{N}{2} \right\rfloor \right) \quad 1 \leq i \leq N, \quad (49.33)$$

where

$$\delta(i) = \begin{cases} 1 & i = 0 \\ 0 & i \neq 0 \end{cases}. \quad (49.34)$$

We construct a bandpass by filtering the data with a high pass-filter, then filtering the output with a low-pass filter. Or we can combine the two filters together into a single filter by convolving the coefficients.

49.3.3 Design of IIR Filters

The common strategy in designing IIR filters is as follows:

Step 1. Design a rational function in the S -domain in factored form that best approximates the desired frequency response characteristics (using Butterworth, Tchebyshev, elliptical, etc.).

Step 2. Transform the poles and zeros into the z -domain.

Step 3. Reconstruct the rational function.

Step 4. Realize the difference equation by inverse z -transform of the rational polynomials.

There are different S - to z -transforms. The two most common are the impulse-invariant and the bilinear transformation. The impulse-invariant transformation is

$$z = e^{sT}, \quad (49.35)$$

where T is the sampling period. This method is usually not used because it can cause aliasing. The bilinear transformation avoids this but distorts the mapping in other ways for which we need to compensate. The bilinear transformation is

$$z = \frac{2/T + s}{2/T - s}. \quad (49.36)$$

The S - to z -transformation is not exact and always involves various trade-offs. Because of this, the actual placement of the cutoff in a designed digital filter is misplaced and the error increases for cutoff frequencies closer to the Nyquist frequency. To compensate for this effect, we apply Equation (49.37) in the design of our filter:

$$f'_c = \frac{1}{\pi T} \tan(\pi f_c T). \quad (49.37)$$

This process is called “prewarping.”

49.3.3.1 Example IIR Design For a digital system with a sampling rate of 100 Hz, a third-order low-pass Tchebyshev filter with a cutoff frequency at 15 Hz and with 20% ripple in the passband is designed as follows:

Step 1. Compute the Nyquist frequency $f_N = 100 \text{ Hz}/2 = 50 \text{ Hz}$.

Step 2. Compute f' using Equation (49.37) to prewarp the cutoff frequency:

$$f' = 16.22 \text{ Hz}.$$

Step 3. Compute v_0 using Equations (49.13) and (49.14):

$$\epsilon = 0.75 \quad v_0 = 0.3662.$$

Step 4. Determine the poles and zeros of the analog system using Equation (49.12):

$$\begin{aligned} p_1 &= -19.0789 + j94.2364 \\ p_2 &= -38.1578 \\ p_3 &= -19.0789 - j94.2364. \end{aligned}$$

Step 5. Map the poles from the S -domain into the z -domain using the bilinear transformation Equation (49.36) and $T = (1/\text{sampling rate}) = 0.01 \text{ s}$:

$$\begin{aligned} pz_1 &= 0.5407 + j0.6627 \\ pz_2 &= 0.6796 \\ pz_3 &= 0.5407 - j0.6627. \end{aligned}$$

Step 6. Map the zeros from the S -domain into the z -domain also using Equation (49.36). Although we may be tempted to conclude that there are no zeros in the S -domain, since our numerator is constant, we note that $H(s) \rightarrow 0$ as $s \rightarrow \infty$. Thus mapping the bilinear transformation for this case yields

$$\lim_{s \rightarrow \infty} \frac{2/T + s}{2/T - s} = -1.$$

And since this is a third-order system,

$$z_1 = -1 \quad z_2 = -1 \quad z_3 = -1$$

Step 7. Expand the numerator and denominator polynomials:

$$\frac{(z - \zeta_1)(z - \zeta_2)(z - \zeta_3)}{(z - pz_1)(z - pz_2)(z - pz_3)} = \frac{1 + 3z^{-1} + 3z^{-2} + z^{-3}}{-1 + 1.7611z^{-1} - 1.466z^{-2} + 0.4972z^{-3}}.$$

Step 8. Normalize the filter so that the gain in the passband will be 1. In this case, we know that the gain should be 1 at $\omega = 0$. Hence, we evaluate $|H(e^{j\omega})|$ for $\omega = 0$: $|H(e^0)| = 38.4$. The normalized transfer function is then

$$H(z) = \frac{0.026 + 0.078z^{-1} + 0.078z^{-2} + 0.026z^{-3}}{-1 + 1.7611z^{-1} - 1.466z^{-2} + 0.4972z^{-3}}.$$

Step 9. Realize the difference equation from inverse z-transformation of the derived transfer function:

$$y_n = 1.7611y_{n-1} - 1.466y_{n-2} + 0.4972y_{n-3} + 0.026x_n + 0.078x_{n-1} + 0.078x_{n-2} + 0.026x_{n-3}.$$

49.3.4 Design of Various Filters from Low-Pass Prototypes

The procedure for designing a high-pass, bandpass, or band-reject IIR filter is as follows: First, design, by pole-zero placement, a low-pass filter (Butterworth, Tchebyshev, etc.) with an arbitrary cutoff frequency (though for practical considerations, it is best to choose a value midway between 0 Hz and Nyquist), transform the poles and zeros according to the following formulas, reconstitute a new transfer function from the transformed poles and zeros, then realize the digital filter by taking the inverse z-transform of the new transfer function. The formulas with an example are given below, where ω_L and ω_H are the low- and high-frequency cutoffs, respectively, normalized between 0 and π , that is, $\omega_L = 2\pi f_L/\text{sample rate}$, where f_L is the cutoff frequency in hertz; ϕ_L is the normalized cutoff frequency of the low-pass prototype (Oppenheim and Schaffer, 1975).

49.3.4.1 High Pass

$$z' = -\frac{1 + Az}{Z + A}, \quad (49.38)$$

where Z is the pole or zero to be transformed and z' is the transformed pole or zero and

$$A = -\frac{\cos[(\omega_H + \phi_L)/2]}{\cos[(\omega_H - \phi_L)/2]}. \quad (49.39)$$

49.3.4.2 Bandpass The bandpass filter has two transitions: a rising edge and a falling edge. For this reason, we need twice as many coefficients for the same approximate transition width as the prototype filter. Thus, the order of these polynomials will be twice the order of polynomials in the prototype. Each pole from the prototype will transform into a pair of poles (z'_1 and z'_2). Likewise, each zero will transform into a pair of zeros:

$$z'_1 = \frac{(A + AZ) + \sqrt{(A + AZ)^2 - 4(Z + B)(BZ + 1)}}{2(Z + B)} \quad (49.40)$$

$$z'_2 = \frac{(A + AZ) - \sqrt{(A + AZ)^2 - 4(Z + B)(BZ + 1)}}{2(Z + B)} \quad (49.41)$$

$$A = \frac{2CD}{D + 1} \quad (49.42)$$

$$B = \frac{D - 1}{D + 1} \quad (49.43)$$

$$C = \frac{\cos[(\omega_H + \omega_L)/2]}{\cos[(\omega_H - \omega_L)/2]} \quad (49.44)$$

$$D = \cot\left(\frac{\omega_H - \omega_L}{2}\right) \tan \frac{\phi_L}{2}. \quad (49.45)$$

49.3.4.3 Band Reject

$$z'_1 = \frac{(AZ - A) + \sqrt{(AZ - A)^2 - 4(Z - B)(BZ - 1)}}{2(Z - B)} \quad (49.46)$$

$$z'_2 = \frac{(AZ - A) - \sqrt{(AZ - A)^2 - 4(Z - B)(BZ - 1)}}{2(Z - B)} \quad (49.47)$$

$$C = \frac{\cos[(\omega_H + \omega_L)/2]}{\cos[(\omega_H - \omega_L)/2]} \quad (49.48)$$

$$D = \tan\left(\frac{\omega_H - \omega_L}{2}\right) \tan \frac{\phi_L}{2}. \quad (49.49)$$

As an example, design a Tchebyshev (Type I) bandpass filter for a system sampled at 100 Hz with a low cutoff frequency of 20 Hz and a high cutoff frequency of 35 Hz.

Step 1. We can use the poles and zeros designed in Section 49.3.3 as the low-pass prototype.

Step 2. Convert the low and high cutoff frequencies to values normalized between 0 and π , where π corresponds to the Nyquist frequency:

$$\omega_L = 2\pi \times \frac{20}{100} = 1.2566 \quad \omega_H = 2\pi \times \frac{35}{100} = 2.1991. \quad (49.50)$$

Step 3. Map the poles from the prototype using Equations (49.40) and (49.41):

$$p_1 \rightarrow \begin{cases} -0.6095 - j0.9105 \\ 0.2940 + j1.0722 \end{cases} \quad (49.51)$$

$$p_2 \rightarrow \begin{cases} -0.2302 - j1.2527 \\ -0.2302 + j1.2527 \end{cases} \quad (49.52)$$

$$p_3 \rightarrow \begin{cases} -0.6095 + j0.9105 \\ 0.2940 - j1.0722 \end{cases}. \quad (49.53)$$

Then map the zeros from the prototype using Equations (49.40) and (49.41):

$$\zeta_1 \rightarrow \begin{cases} -1 \\ 1 \end{cases} \quad (49.54)$$

$$\zeta_2 \rightarrow \begin{cases} -1 \\ 1 \end{cases} \quad (49.55)$$

$$\zeta_3 \rightarrow \begin{cases} -1 \\ 1 \end{cases}. \quad (49.56)$$

Step 4. Expand the numerator and denominator polynomials:

$$\frac{(z - \zeta_1) \cdots (z - \zeta_6)}{(z - pz_1) \cdots (z - pz_6)} = \frac{1 - 3z^{-2} + 3z^{-4} - z^{-6}}{(1 + 1.091z^{-1} + 3.632z^{-2} + 2.616z^{-3} + 4.642z^{-4} + 1.982z^{-5} + 2.407z^{-6})}. \quad (49.57)$$

Step 5. Normalize the transfer function so that it will have unity gain in the passband. For this, we estimate

$$M = \max_{0 \leq \omega \leq \pi} |H(e^{j\omega})|. \quad (49.58)$$

Then compute

$$H_{\text{normalized}}(z) = \frac{1}{M} H(z). \quad (49.59)$$

For this example, $M = 14.45$.

Step 6. Realize the difference equation from inverse z -transformation of the derived transfer function:

$$y_n = -1.091y_{n-1} - 3.632y_{n-2} - 2.616y_{n-3} - 4.642y_{n-4} - 1.982y_{n-5} - 2.407y_{n-6} + 0.0692x_n - 0.2076x_{n-2} + 0.2076x_{n-4} - 0.0692x_{n-6}. \quad (49.60)$$

49.3.5 Frequency-Domain Filtering

It is possible to filter the data in the frequency domain. The method involves the use of the Fourier transform. We Fourier transform the data, multiply by the desired frequency response, then inverse Fourier transform the data. This is similar to the FIR filters discussed earlier. Deriving the FIR coefficients by performing a discrete cosine transform (DCT) of the desired frequency response and then convolving the coefficients with the data is equivalent to filtering the data in the frequency domain. One difference, however, is that the frequency domain filtering is generally done on blocks of data and not on streaming data, as is done in the time domain, which can be of concern when processing highly nonstationary data with abrupt transients. The inverse Fourier transform is

$$\mathcal{F}^{-1}\{F(\omega)\} = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} f(\omega) e^{-j\omega t} d\omega. \quad (49.61)$$

49.4 STABILITY AND PHASE ANALYSIS

49.4.1 Stability Analysis

Consider a transfer function

$$H(s) = \frac{1}{s - p}, \quad (49.62)$$

where the pole p is a complex number $\alpha + j\beta$. The inverse Laplace transform of this function is $e^{\alpha t}[\cos(\beta t) + j \sin(\beta t)]$. This function is bounded as $t \rightarrow \infty$ if and only if $\alpha < 0$. From this, we can determine the stability of a function by inspecting the real components of all poles of a given transfer function. The procedure for a rational function (a ratio of polynomials) would be to factor the polynomials in the denominator and inspect to ensure that the real components to all of the poles are less than zero. Suppose

$$\begin{aligned} H(s) &= \frac{a_0 + a_1s + \cdots + a_ms^m}{b_0 + b_1s + \cdots + b_ns^2} \\ &= \frac{(s - \zeta_0)(s - \zeta_1) \cdots (s - \zeta_m)}{(s - p_0)(s - p_1) \cdots (s - p_n)} \end{aligned} \quad (49.63)$$

In a similar way, by inspection of the S -to- z transformation $z = e^s$ we see that the entire left half of the plane in the S -domain maps inside the unit circle in the z -domain. For this reason, we analyze the stability of systems in the z -domain by inspecting the poles of the transfer function. The system is stable if the norm of all poles is less than 1.

49.4.2 Phase Analysis

While processing the data in real time, our filters must act on the signal history. For this reason, there will always be some delay in the output of our process. Worse, certain filters will delay some frequency components by more or less than other frequency components. This results in a phase distortion of the filter. For a certain class of FIR filters, it is possible to design filters that shift each frequency component by a time delay in proportion to the frequency. In this way, all frequency components are shifted by an equal time delay. Although it is possible to design certain nonreal-time, noncausal IIR filters that are phase shift distortionless, in general, IIR filters will produce some phase shift distortion. We can determine the actual phase shift for each frequency component by computing

$$\arg H(j\omega) = \angle H_{\text{real}} + jH_{\text{imag}} = \tan^{-1} \frac{H_{\text{imag}}}{H_{\text{real}}}.$$

The arctan will produce the principal value of the phase shift, not necessarily the cumulative phase shift, since the arctan function produces principal value $-\pi \leq \tan^{-1}(\phi) \leq \pi$. It is possible to recover the accumulated phase shift by factoring the rational function into its binomial parts, then expressing this in exponential form as a summation. One can then determine the principal angle on each part and the accumulated phase shift by summing the parts.

49.4.3 Comparison of FIR and IIR Filters

There are various factors when deciding on a particular filter for a given application. Table 49.3 summarizes these.

TABLE 49.3 Comparison of FIR and IIR Characteristics

	FIR	IIR
Run time efficiency	Less efficient; requires high-order filter	Higher efficiency; usually possible to achieve a desired design specification in fewer computations
Stability	Always stable	Stable if all poles are inside the unit circle
Phase shift distortion	Can be designed to be phase shift distortionless	Generally distorts phase
Ease of design	Simpler design process, usually involving Fourier transforms or solving linear systems	Design is more complex, involving special functions or solving nonlinear systems

49.5 EXTRACTING SIGNAL FROM NOISE

The PSD of white noise is uniformly distributed over all frequencies. Therefore, it is possible to detect the PSD signature of a signal corrupted by white noise by inspecting spectral components that rise above some baseline. From this, we can design a matching filter to optimally extract the signal from the noise. Figures 49.11 and 49.12 illustrate this procedure.

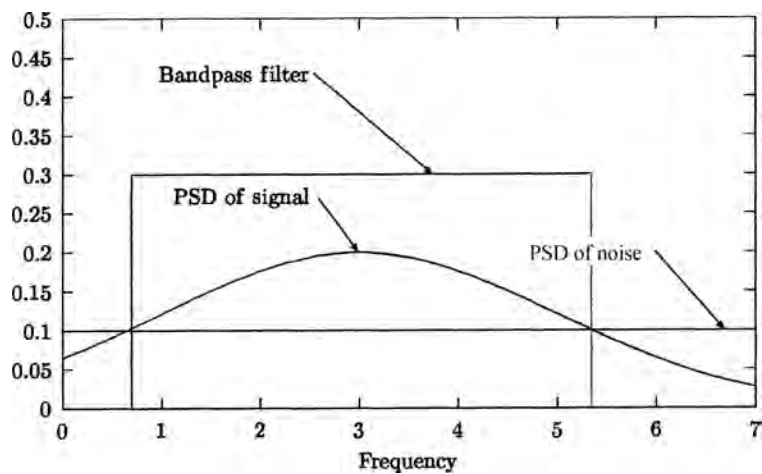


FIGURE 49.11 Use of bandpass filter for discriminating signal from noise.

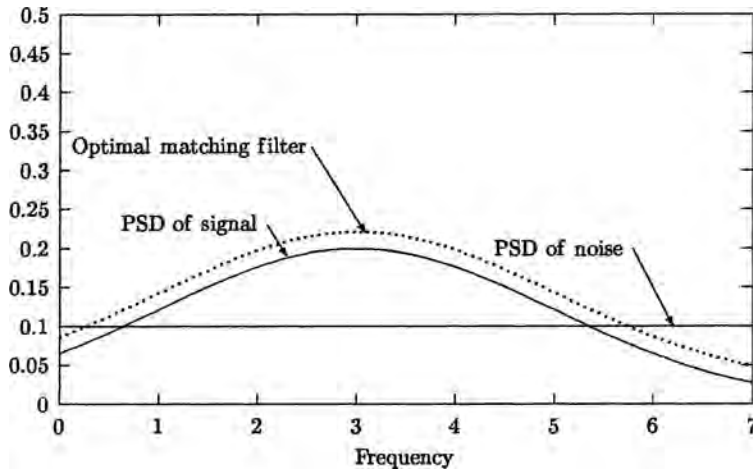


FIGURE 49.12 Improved matching filter for better discrimination of signal from noise.

REFERENCES

- Marple S. *Digital Spectral Analysis with Applications*. Englewood Cliffs (NJ): Prentice-Hall; 1987.
- Oppenheim A, Schaffer R. *Digital Signal Processing*. Englewood Cliffs (NJ): Prentice-Hall; 1975.
- Parks T, Burrus C. *Digital Filter Design*. New York: Wiley; 1987.
- Press W, Flannery B, Teukolsky A, Vetterline W. *Numerical Recipes in C*. Cambridge, UK: Cambridge University Press; 1988.
- Vlček M, Unbehauen R. Analytical solutions for design of IIR equiripple filters. *IEEE Transactions on Acoustics, Speech, and Signal Processing* 1989;37(10): 1518–1531.

50

DATA ACQUISITION AND DISPLAY SYSTEMS

PHILIP C. MILLIMAN

- 50.1 Introduction
- 50.2 Data acquisition
- 50.3 Process data acquisition
 - 50.3.1 Sampling interval
 - 50.3.2 Accuracy and precision of data
 - 50.3.3 Time-based versus event-driven collection
- 50.4 Data conditioning
 - 50.4.1 Simple linear fit
 - 50.4.2 Nonlinear relationships
 - 50.4.3 Filtering
 - 50.4.4 Compression techniques
 - 50.4.5 More on sampling and compression
- 50.5 Data storage
 - 50.5.1 In-memory storage
 - 50.5.2 File storage
 - 50.5.3 Database storage
 - 50.5.4 Using third-party data acquisition systems
- 50.6 Data display and reporting
 - 50.6.1 Current-value inspection
 - 50.6.2 Display of individual data points
 - 50.6.3 Display of historical data
- 50.7 Data analysis
 - 50.7.1 Distributed systems
 - 50.7.2 System error analysis
- 50.8 Data communications
 - 50.8.1 Serial communications
 - 50.8.2 Parallel communications
 - 50.8.3 Networks
 - 50.8.4 OSI standard
 - 50.8.5 OPC standard
 - 50.8.6 Benefits of standard communications

- 50.9 Other data acquisition and display topics
 - 50.9.1 Data chain
 - 50.9.2 Web programs and interfaces
 - 50.9.3 Configuration versus implementation
 - 50.9.4 Store and forward
 - 50.9.5 Additional communications topics
- 50.10 Summary
- References

50.1 INTRODUCTION

The industry has changed significantly since this chapter was first written in the months before 1990. The personal computer has become part and parcel of everyday life. Control systems have become increasingly based on standard systems and interfaces; sensors themselves are often based on just smaller versions of the same operating system as large manufacturing systems. This has tended to change the focus from the technology of data acquisition to the software and systems to support data acquisition.

The trend has been away from requiring the engineer to understand the science of how sensors work and the lowest levels of data acquisition and more toward the engineer understanding the collection, coordination, storage, access, and manipulation of data. With that in mind, this chapter has been updated to focus more on the latter and less on the former. Other chapters in this book cover details of the electronics, transducers, sampling, and calibration.

To control any process or understand what occurs during the life cycle of a process, the system (a human or machine) must have information about what is occurring. In the simplest of control loops, the measured variable must be converted into usable units, comparison in some form to a target occurs, and a response is determined. At the plant level, improvement of plant operation relies upon understanding the relationships between processes within the plant (not only current, but historical), which in turn requires collecting data throughout the plant, characterizing the relationship of the data with other data, storing the data in such a way as to be retrievable in a useful, timely way, and manipulating the data for presentation and hopefully providing an aid to understanding the relationships between processes. In today's competitive environment, focusing on local control and ignoring the interaction between processes, both internal to the plant and external, can be disastrous. If one is not focused on improvement, one can bet the competitor is. Larger corporations, especially, can bring analytical tools to bear to improve local processes, plantwide processes, and their relationships to external influences, such as the supply chain. On the other hand, today's computing tools bring very powerful data acquisition and analysis capability within the reach of the average technician with a little bit of motivation.

Data acquisition and display systems have changed dramatically. Twenty years ago, terms referring to specialized systems such as SCADA (supervisory control and data acquisition) and data loggers were common terms. Now, with the proliferation and broadening role of computer systems and their intrusion into every aspect of manufacturing,

many of the features that used to be in specialized instruments and systems are now part of the everyday tools available to anyone with a computer. This chapter attempts to cover aspects of data acquisition and manipulation that may help the engineer better understand issues and give a foundation for using and even constructing tools. The organization is as follows:

- The initial sections cover the nature of data and the acquisition and conversion of data to usable units and includes some discussion of useful display techniques. The discussion attempts to identify issues of which the engineer should be aware and give guidelines on how to manage data.
- The latter sections cover the coordination, storage, access, and manipulation of data. A discussion of pros and cons of different strategies should help the reader understand the trade-offs in system selection and construction. It is difficult to do this without describing specific technologies and brands, but the author has endeavored to level the discussion in such a way that changes in technology will not change the value of the discussion. Time will tell if the approach is effective.

50.2 DATA ACQUISITION

Data acquisition includes the following: (1) acquiring raw data from the process being measured and (2) converting data into usable units. Included in this section are also some topics of data display closely related to the nature of the data being acquired. Other aspects of data display will be covered in later sections.

In process industries, much of the data are analog in nature, such as pressure, temperature, and flow rate. The values acquired are sampled representations of process data that have a scale and a range, with various issues around effective range and whether values over a range are linear or more complex. When acquired in a data acquisition system there are a number of issues that must be addressed related to how data is sampled; how it is represented in the computer as a digital value but still able to be manipulated as an analog number; and how a continuously changing value can be stored without exceeding the capacity of storage or computation of the acquisition system. Discrete manufacturing still has a number of analog data sources, but a larger proportion involves discrete data, such as motor stops, starts, and pulses. These have their own issues of acquisition and storage and are often related to attributes of the process. The next section deals primarily with process data, additionally covering some discrete data and issues around data collection, representation, and storage.

As businesses begin to broaden the scope of optimization and understand their global processes, the context of the data in terms of product, plant conditions, market conditions, and other environmental aspects has increasingly added discrete data to the set of data to be obtained. The interaction of the process with factors such as which crew is managing the process, which customer's needs are highest priority, legal controls such as environmental limits impacting allowable process rates, operator decisions, which product is being manufactured, grade achieved, and a large number of other factors become important when a company is competing with other companies that have already achieved excellent local control of processes. These data involve less understanding how to deal with continuous data and more with the coordination of data within and between processes. These can be termed manufacturing attributes to emphasize their importance in

providing an environment around process data. Later sections deal with manufacturing attribute data and issues around their collection and coordination with process data.

50.3 PROCESS DATA ACQUISITION

Most modern data acquisition is via digital systems that may have a lower level analog collection mechanism but is now so removed from the engineer that the engineer is only concerned with the digital portion of the system. The ability to use digital microprocessors as building blocks for data collection, the prevalence of computer tools, and the creation of widely available operating systems that operate on small footprints have virtually eliminated the need for analog instrumentation. While the data may be analog in nature, the technology has been developed to such a degree that the engineer decreasingly needs to pay attention to the analog aspects of the data.

A digital-to-analog (D/A) and analog-to-digital (A/D) converter performs the actual processing required to bring analog information from or to the process. While the resulting signal may be digital, it is a representation of a continuous number that has characteristics that, if not understood, can result in erroneous conclusions from data, including missing data, misinterpreting trends, or improperly weighting certain values.

The engineer should be aware of several features of analog data to ensure that the data are used properly. An understanding of sampling interval, scaling, and linearization will facilitate the use of data once collected.

50.3.1 Sampling Interval

One of the important steps with any data collection process includes the proper choice of sampling interval. As an example of the impact of selection of sampling interval, or frequency, a sine wave with a period of 1 s (Figure 50.1a) is measured with several sampling intervals. Both 0.5-s (Figure 50.1b) and 0.1-s (Figure 50.1c) intervals provide different impressions of what is actually happening. The 1-s sampling rate being in phase with the sine wave yields the impression that we are measuring a nonvarying level. The 0.5-s sampling rate yields several different results depending on what phase shift is encountered. This is known as aliasing (see Johnson, 1984, pp. 122–125). If a 0.1-s period is used, we finally begin to obtain a realistic idea of what the waveform truly looks like.

The sampling interval has a different impact when collecting manufacturing attribute data. With manufacturing attribute data, every change in value has importance. They provide a context to the process that assists with the tieback to business interactions. The values are often coded. The sampling must be close enough to the time of occurrence to allow determination of state in relation to other events. The sampling interval must be fast enough to capture any change in state. Sampling at slower than the change rate of the data will mean lost events—potentially critical in a situation where the count of items processed is important. Sampling at a rate slightly faster than the maximum change rate of the manufacturing attribute data assures that no change will be missed. Another important consideration is to know at what time an event occurred. For instance, if a value changes only infrequently but other related manufacturing attribute values are changing at a faster rate, then the scan rate has to be fast enough to match the fastest change rate of all the related manufacturing attribute data. This is sometimes called the master scan rate, meaning that the frequency of scanning must be fast enough to capture faster events and

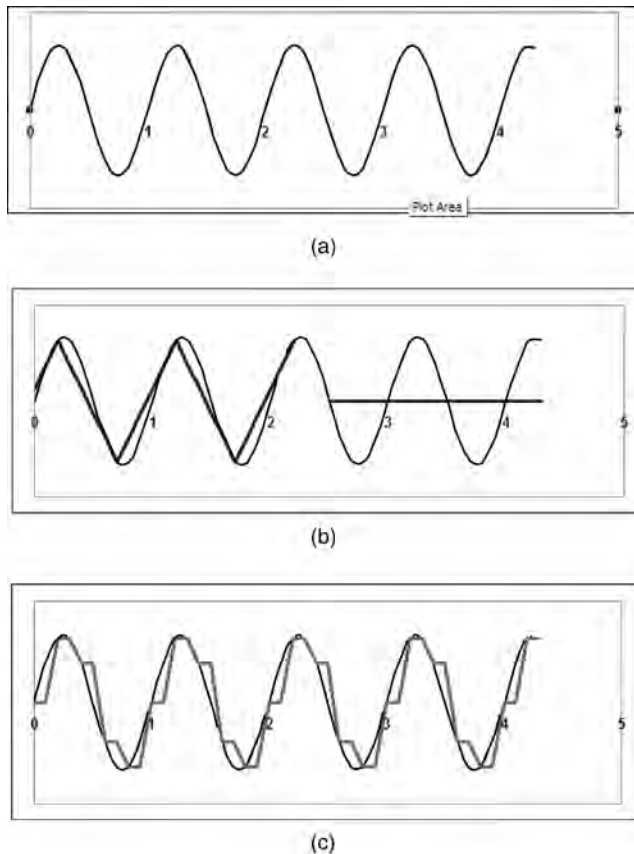


FIGURE 50.1 Sampling (a) sine wave form; (b) aliasing of data; (c) sampled every 10 s.

determine the state of other variables relative to those events. Similarly, if there are related analog data, the scan rate may have to be fast enough to even characterize the curve of the analog data (remember Figure 50.1).

The capacity of the target system must also be taken into account. Storing large volumes of data is becoming more feasible as systems increase in speed and power, but the retrieval and organization of those data may become a time-consuming, overly complex task with too much data or poorly organized data. Consequently, even though storage itself is less of an issue, other factors impact how much data are retained and how organized for later retrieval. Later sections examine approaches for organizing and retrieving data.

50.3.2 Accuracy and Precision of Data

Accuracy and precision are dependent on the sampling interval as well as the resolution of the system (Murrill, 1981; Liptak, 2003, pp. 78–80; and Chapter 1 in this volume). When dealing with the A/D conversion process, the step size or number of bits used is critical when determining the system precision and accuracy (Johnson, 1984, pp. 78–81). Figure 50.2 illustrates the difference between accuracy and precision of data. Table 50.1 illustrates the effect the number of bits has on the precision.

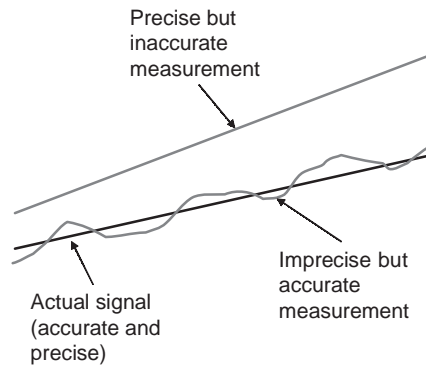


FIGURE 50.2 Difference between accuracy and precision.

This also interacts with range, which will be discussed later, since having an accurate number over a small percent of the desired range would not allow the ability to fully characterize the process. For example, highly accurate readings with 1% moisture resolution over a range from 10 to 20% moisture content would be inadequate if one were attempting to measure moisture over a 5–40% range. When selecting transducers, it is necessary that they have both the accuracy and the range needed for the process being observed. When selecting converters, one should be aware of the settling time (governs how often readings can be obtained), resolution of the converter (affects range and detail of measurements), and accuracy of sensor.

Chapter 1 includes some characteristics of transducers, including calibration, and the sampling of data.

One should be aware that an event that has been stabilized in a data collection system may be offset in time, resulting in a potential discrepancy between events or values when values are compared from different sources or from multiplexed data. It should be verified that the transducer is collecting the data fast enough to allow one to have relevant times of collection in the data acquisition system. Also, one should assure that the potential relationship between events from different sources and their intended use is understood when considering the speed and accuracy of transducers.

50.3.3 Time-Based Versus Event-Driven Collection

There are two major approaches when collecting data with a general-purpose data acquisition system. In one approach, data are collected on a regular frequency based on

TABLE 50.1 Relationship Between Number of Bits and Precision

Number of Bits	Steps	Resolution on 5-V Measurement	Percent Resolution
8	256	0.01950	0.3900
10	1,024	0.00488	0.0980
12	4,096	0.00122	0.0240
16	65,536	0.00008	0.0015

time, such as once per second. This is easy to institute and it is relatively easy to analyze the data and their relationships after the fact. This approach tends to require more data storage and can make it difficult to identify events or the interactions with manufacturing attribute data. The other approach is event-based acquisition. An event is identified, such as when a package is dropped onto a platform, the time of that event is recorded, and the values of related variables are collected for that time. The sampling rate of the transducers to acquire the other variables may be important, as their values may become irrelevant if too long a time interval has passed after the event has occurred when the related variables are sampled. Batch processes, such as mixing a tankful of chemicals, often have some data collected only at the start and end of the process. Other data may be recorded at fixed time intervals during the batch process. Depending on the needs, the data during the actual reaction may be of great or of little use. The time between sampled events may be several minutes, hours, or even days in length, but the time of the event may be critical, as well as detailed data at the time of the event, resulting in a common tactic of using high-speed scanning to detect the occurrence of an infrequent event. Approaches for combining and analyzing data will be covered in a later section.

50.4 DATA CONDITIONING

Often the data obtained from a process are not in the form or units desired. This section describes several methods of transforming data to produce proper units, reduce storage quantity, and reduce noise.

There are many reasons why process measurements might need to be transformed in order to be useful. Usually, the signals obtained will be values whose units (e.g., voltage, current) are other than the desired units (e.g., temperature, pressure). For example, the measurement from a pressure transducer may be in the range 4–20 mA. To use this as a pressure measurement in pounds per square inch (PSI), one would need to convert it using some equation. As environmental conditions change, the performance characteristics of many sensors change. A parametric model (equation) can be used to convert between types of units or to correct for changes in the parameters of the model. The parameters for this equation may be derived through a process known as calibration (Chapter 1 covers much of the process of calibration and sampling). This involves determining the parameters of some equation by placing the sensor in known environmental conditions (such as freezing or boiling water) and recording the voltage or other measurable quantity it produces. Some (normally simple) calculations will then produce the parameters desired. (See the following discussion of simple linear fit for the procedure for a simple, two-parameter equation.) The complexity of the model increases when the measured value is not directly proportional to the desired units (nonlinear). Additionally, as sensors get dirty or age, the parameters might need adjusting. There are a variety of control techniques to assist with compensating for changes in the environment around the sensor, including adaptive control.

50.4.1 Simple Linear Fit

The simplest formula for converting a measured value to the desired units is a simple linear equation. The form of the equation is $y = ax + b$, where x is the measured value, y

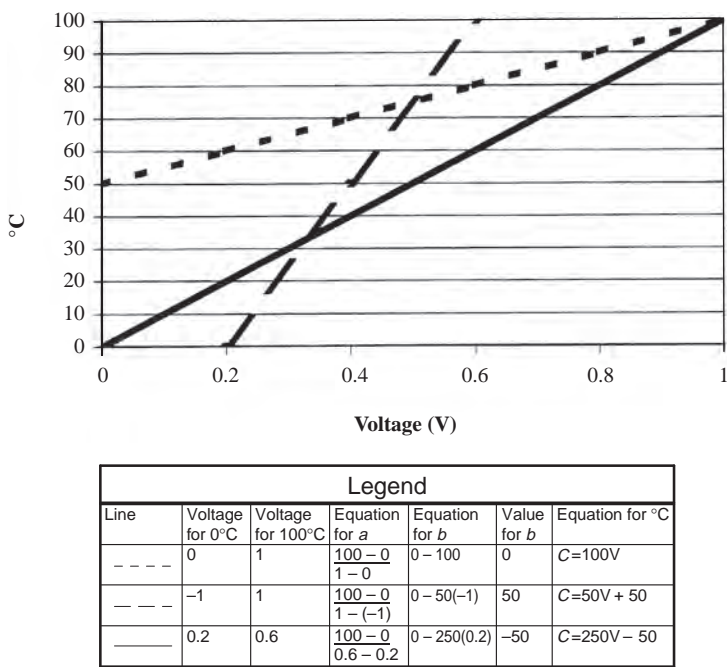


FIGURE 50.3 Relationship between measured values and engineering units.

represents the value in the units desired, and *a* and *b* are parameters to adjust the slope and offset, respectively. The procedure for finding *a* and *b* is as follows:

1. create a known state for the sensor in the low range. An example would be to put a temperature sensor in ice water;
2. determine the value obtained from the sensor;
3. create a known state for the sensor in the high range. An example would be to immerse the sensor in boiling water;
4. determine the value obtained from the sensor;
5. calculate the values of *a* and *b* from these values using the equations.

$$a = \frac{\text{actual high value} - \text{actual low value}}{\text{measured high value} - \text{measured low value}} \tag{50.1}$$

$$b = \text{actual low value} - (a \times \text{measured low value}). \tag{50.2}$$

Figure 50.3 demonstrates example relationships between measured values and engineering units.

50.4.2 Nonlinear Relationships

Often, there is not a simple linear relationship between the engineering units and the measured units (Figure 50.4). Instead, for a constantly rising pressure or temperature, the

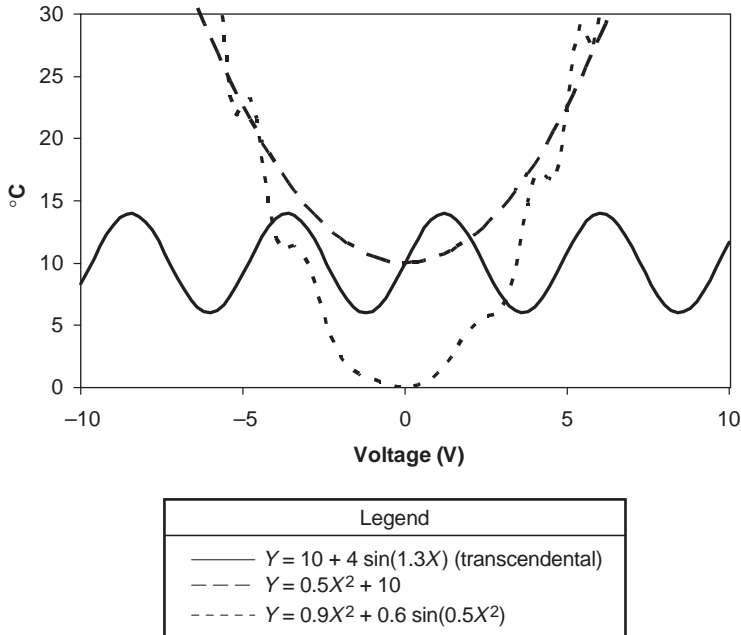


FIGURE 50.4 More complex data relationships.

measured value would form some curve. If possible, we use a portion of the sensor's range where it is linear, and we can use Equations (50.1) and (50.2). When this is not possible, we have to characterize the sensor by a different equation, which could be a polynomial, a transcendental, or a combination of a series of functions.

One can imagine several sensors which are linear in different ranges to be used in conjunction to create a larger range of operational data. This variety of formulas should make one point clear: without an understanding of the basic model of the sensor, one cannot know what type of conversion to use. Many sensors have known differences in output depending on the range of sensed data. Be aware of the effect environmental conditions have on the sensor readings. If the characteristics of a sensor are unknown, then the sensor must be measured under a variety of conditions to determine the basic relationship between the measured values and the engineering units. Some knowledge of the theory of the sensor's mechanism will help to give an idea of which model to use. The development and evaluation of a model is beyond the scope of this chapter, but other chapters in this volume provide assistance.

50.4.3 Filtering

Even after data are converted into the appropriate units, the data may have characteristics that inhibit understanding the important relationships for which one is looking. For instance, the data may have occasional fluctuations caused by factors other than the process or the process may have short-term perturbations, which are not really an indication of the major process factors.

Filtering is a technique that allows one to retain the essence of the data while minimizing the effects of fluctuations. The data may then appear to be “smoothed.” In fact, the terms “filtering” and “smoothing” are often interchanged. Filtering can occur when the data are still in an analog state (Johnson, 1984, p. 54) or can occur after the data are converted into digital data (digital filtering). Measurement variability comes from a variety of sources. The process itself may undergo fluctuations that result in variation in measurement but that are only temporary and should be ignored. For instance, if the level of an open tank of water were to be measured but waves cause fluctuations in the height of a float, then the exact value at any given time would not be an accurate reflection of the level of the tank. The sensor itself may have fluctuations due to variability in its method for acquiring data. For instance, the proximity of a 60-Hz line may induce a 60-Hz sinusoidal variation in the signal (measured value). Examples of filtering approaches are as follows and are also given in other chapters in this volume (Wright and Edgar, 1983, p. 538):

- (a) *Repeated Sample Average*: take a number N of samples at once and average them:

$$\frac{1}{N} \sum \text{Value}(i)$$

- (b) *Finite-Length Average (Moving)*: take the average of the last N measurements, averaging them to obtain a current calculated value.

- (c) *Digital Filters*:

$$y = (1 - \alpha)y_{i-1} + \alpha x_{i-1}.$$

The simple average is useful when repeated samples are taken at approximately the same point of time. The more samples, the more random noise is removed. Chapter 1 addresses some of the issues with sampling and the concept of population distribution. The formula for an average is shown in (a) above.

However, if the noise appeared for all the samples (as when all the samples are taken at just the time that a wave ripples through a tank), then this average would still have the noise value. A moving average can be taken over time [see (b) above] with the same formula, but each value would be from the same sensor, only displaced in time. A disadvantage of this approach is that one has to keep a list of previous values at least as long as the time span one wishes to average.

A simpler approach is the first-order digital filter, where a portion of the current sample is combined with a portion of previous samples. This composite value, since it contains more than one measured value, will tend to discard transitory information and retain information that has existed over more than one scan. The formula for a first-order digital filter is described in (c) above, where α is a factor selected by the user. The more one wants the data filtered, the smaller the choice of α ; the less one wants filtered, the larger the choice of α . Alpha must be between 0 and 1 inclusive.

The moving average or digital filter can tend to make the appearance of important events to be later than the event occurred in the real world. This can be mitigated somewhat for moving averages by including data centered on the point of time of interest in the moving-average calculation. These filters can be cascaded, that is, the output of a filter can be used as the input to another filter.

A danger with any filter is that valuable information might be lost. This is related to the concept of compression, which is covered in the next section. When data are not continuous, with peaks or exceptions being important elements to record, simple filters such as moving-average or digital filters are not adequate. Some laboratory instruments such as a gas chromatograph may have profiles that correspond to certain types of data (a peak may correspond to the existence of an element). The data acquisition system can be trained to look for these profiles through a pre-existing set of instructions or the human operator could indicate which profiles correspond to an element and the data acquisition system would build a set of rules. An example is to record the average of a sample of a set of data but also record the minimum and maximum. In situations where moisture or other physical attributes are measured, this is a common practice. Voice recognition systems often operate on a similar set of procedures. The operator speaks some words on request into the computer and it builds an internal profile of how the operator speaks to use later on new words. One area where pattern matching is used is in error-correcting serial data transmission. When serial data are being transmitted, a common practice to reduce errors is to insert known patterns into the data stream before and after the data. What if there is noise on the line? One can then look for a start-of-message pattern and an end-of-message pattern. Any data coming over the line that are not bracketed by these characters would be ignored or flagged as extraneous transmission.

50.4.4 Compression Techniques

For high-speed or long-duration data collection sessions there may be massive amounts of data collected. It is a difficult decision to determine how much detail to retain. The trade-offs are not just in space but also in the time required to store and later retrieve the data. Sampling techniques, also covered in other chapters, provide a way of retaining much of the important features of the data while eliminating the less important noise or redundant data. As an example, 1000 points of data collected each second for 1 year in a database could easily approach 0.5 terabyte when index files and other overhead are taken into account. Even if one has a large disk farm, the time required to get to a specific data element can be prohibitive. Often, systems are indexed by time of collection of the data point. This speeds up retrieval when a specific time frame is desired but is notoriously slow when relationships between data are explored or when events are searched for based on value and not on time. Approaches to reducing data volume include the following:

- Reduce the volume of data stored by various compression tools, such as discovering repeating data, storing one copy, and then indicating how many times the data are repeated. There are many techniques for compressing data, covered elsewhere. Zip files are a common instance of compression to make the data consume less storage and to take less time in transmission from one computer to another. Compression tends to increase the storage and retrieval time slightly. Increasingly, file systems associated with common operating systems include compression as a standard option or feature of mass storage. These systems are quite good at compressing repeated data but are less effective when data vary but have a mathematical relationship, such as a straight line between two points.
- Normalize the data. When the developer knows relationships between data, redundancy can be avoided by normalizing the data—following some basic principles to organize the data in such a way that redundancy is avoided (Date, 1990). C. J. Date

(1990) describes the levels of normalization of data in a relational database. For instance, if a person has several addresses, then one could store the person's name once, store each address, and store the links from the person to the address. While very similar to compression, it relies on the developer identifying and taking advantage of the relationships between data to eliminate redundancy and reduce space. This creates significant effort in planning for acquisition and storage of data. It pays off in reduced storage and significantly improved retrieval and analysis times.

- Eliminate nonessential data. If one is not interested in the shape of a sinusoidal signal, for instance, but only interested in how many cycles occurred during a given time frame, then sampling techniques can be used to characterize the data without having to store significant data.

The engineer or researcher has to make assumptions about how the data will be used and factor those into the acquisition and storage system. A project attempting to discover the relationships between waveforms would require high-frequency sampling and probably time-based storage. A project attempting to record the number of times a boiler went over a certain temperature level might have a high-speed scanning capability but only store those values that were above the temperature limit. An inventory tracking system may have triggers that cause scanning only when some event occurs.

Often, a batch or pallet of product may contain a large number of items. The items can be sampled for some process attribute. The customer may want to know summary statistics about the pallet, but the storage of all the data may not be feasible. In this case, statistical results can normally be derived from summary data. Average, standard deviation, total, correlation, maximum, and minimum are easily calculated from summary, accumulated data¹:

Averages: Keep a running sum of data and count of readings:

$$\text{Average} = \text{Sum of values} / \text{count of values}$$

Totals: Keep a running sum of data.

Standard Deviation:

$$\sqrt{\frac{n \sum_{i=1}^n x_i^2 - (\sum_{i=1}^n x_i)^2}{n(n-1)}}$$

Correlation:

$$r = \frac{n \sum x_i y_i - (\sum x_i)(\sum y_i)}{\sqrt{[n \sum x_i^2 - (\sum x_i)^2][n \sum y_i^2 - (\sum y_i)^2]}} \quad (50.3)$$

Range: Save largest and smallest values.

Median: Find the middle value of a distribution, which requires keeping all values.

¹ Standard deviation and correlation from Beyer (1976, pp. 473, 477).

The median (the true center of the data) requires the raw data to be calculated. A compromise for depicting the distribution of data without having to store the full details is to store a distribution of the data. For instance, the range of possible important data can be broken into a series of totals, reflecting the count of items that fit into the particular total. A histogram representing the distribution of data can be created from the totals without requiring the full set of original data. In addition, the median can be approximated using this technique. The distribution can also be used to supply data for statistics based on distribution of data, such as the Taguchi loss function (Crossley, 2000, pp. 397–400).

50.4.5 More on Sampling and Compression

Rather than just sample the data, why not save all the changed values of the data, discarding values which are the same or within some limits of the previous reading? This really applies best to continuous processes. Quite significant space reduction can be maintained in processes that are slowly changing and have only occasional large upsets. Variations of this technique can provide additional improvements. For instance, rather than just checking to see if the current value is the same as or within some limits from the previous reading, see if it is on the same line or curve as the previous value. This can result in a great reduction of storage requirements at the loss of a slight amount of accuracy in reconstruction. The more flexible the compression technique, the more work must be done to reconstruct the data later for examination. For instance, if the user of the data acquisition system wants to retrieve a data point within data that has been reduced to a line segment, the user or the system must determine which line segment is wanted using the time stamp for the beginning and ending of the line segment interval and then recalculate the point from the equation. This is referred to as a boxcar algorithm. Values that are close to the line segment can be treated as on the line segment if one can afford to lose some accuracy (http://www.aspentech.com/publication_files/White_Paper_for_IP_21.pdf). The formula for a boxcar has to take into account the length of the interval (maximum), the height of the box (how much noise is allowed), and how peak or exception values are treated. For instance, Table 50.2 presents a set of data with several types of compression applied for data sampled at a constant interval.

In Table 50.2, the simple repeating-value compression will not lose data but will result in little or no compression if the data are changing value frequently, including having any noise. The boxcar compression technique results in much higher compression for slowly changing data with only the loss of fine detail data (depending on the height of the window). For data that are nonlinear or changing frequently, the boxcar compression method results in little compression. Process information systems often use the boxcar compression method. If data are slow moving with occasional bursts of activity, the boxcar and repeating-value methods can result in dramatic reductions in space required. If data changes tend to be linear, then the boxcar algorithm tends to be superior to the repeated-value approach. For an extreme example, see Table 50.3.

The raw data would have resulted in 631 data points being stored. The boxcar method would result in five data points being stored, less than 1% of the storage required. In the example above, the repeated-value method would have resulted in almost exactly the same compression as the boxcar. However, if there had been a 0.1% slope in data throughout the period, the boxcar would remain the same but the repeated-value method would have resulted in no compression. A deadband (the height of the boxcar) around the repeated value (meaning two values within some small deviation from each other would

TABLE 50.2 Examples of Compression Techniques

Time Stamp and Raw Value		Simple Repeating Value			Boxcar Compression		
Time	Value	Time	Value	Count	Start Time	Start Value	Slope
12:00	1	12:00	1.0	3	12:00	1	0
12:01	1	12:03	2.0	1	12:03	2	1
12:02	1	12:04	3.0	1	12:08	7	−2
12:03	2	12:05	4.0	1	12:11	1	4
12:04	3	12:06	5.0	1	12:12	5	0
12:05	4	12:07	6.0	1	12:16	5	−4
12:06	5	12:08	7.0	1			
12:07	6	12:09	5.0	1			
12:08	7	12:10	3.0	1			
12:09	5	12:11	1.0	1			
12:10	3	12:12	5.0	5			
12:11	1	12:17	1.0	1			
12:12	5						
12:13	5						
12:14	5						
12:15	5						
12:16	5						
12:17	1						

be counted as the same value) would result in very high compression in the above example. A long ramp-up of the value during that time frame would have further differentiated the two compression methods.

There are many techniques for compression of data. As aforementioned, many rely on assumptions about the underlying nature of the data, such as being continuous data.

TABLE 50.3 Data Compression: Raw Data Versus Boxcar

Raw Data ^a		Boxcar Method		
Start Time	Start Value	Start Time	Start Value	Slope
12:00	1	12:00	1	0
12:01	1	15:00	1	1
...		15:01	2	3
14:59	1	15:02	5	−4
15:00	1	15:03	1	0
15:01	2			
15:02	5			
15:03	1			
15:04	1			
...				
18:59	1			
19:00	1			

^aGaps represent no change in data.

Where data are directly related to events, more traditional compression techniques which look for repeating patterns in the data may be used. These are typically performed by operating systems and database systems and can therefore be taken advantage of with little or no work on the part of the engineer.

50.5 DATA STORAGE

In whatever ways data are sampled, collected, filtered, smoothed, and/or compressed, at some point the data must be stored on some media if long-term data are to be analyzed (covered later in this chapter). There are several approaches to data storage that will be discussed in brief here.

50.5.1 In-Memory Storage

There are normally limitations on how much data can be stored, particularly when low-frequency events have high-frequency data surrounding them that are of interest. For example, if scientists are monitoring Mt. St. Helens for seismic data, it would be prohibitive to capture millisecond data for years while waiting for an eruption. It would be of interest to capture data at high density just before, during, and after each eruption but not in the quiet times in the intervening years. Collecting the millisecond data on many sensors would overflow the storage capability of most systems. There are techniques to store subsets of the data that allow high-density data from constrained time intervals to be stored.

An approach for collecting and later reporting high-density data around an event of interest is to collect the data continuously using the triggered snapshot method. High-speed data are temporarily retained for a fixed time interval or memory capacity, with the start and end time of the data moving forward with time. Older data are discarded as the time range moves past it. This moving window is useful for creating trends and summary data for that interval. The user can be shown dynamic displays that update over time and reflect characteristics of the moving window of the process. Periodically, a set of the data can be extracted to mass storage, especially triggered by some event of interest. An event is recognized by some means (automatic or user generated) that the engineer has preconfigured to cause the transfer of the current instance of the moving window to permanent storage.

The relationship between the trigger and the moving window can be configured several ways, as depicted in Figure 50.5. The handling of the data involves moving values through a data array, adding more recent values at the end and pushing the rest toward the beginning—a queue.

The triggered snapshot is particularly useful when knowledge about the sequence of events just before the event of interest can help discover problems. As an example in manufacturing, in sawmills there are often very high-speed sequences of events, such as where a board may come out of one conveyor and is transferred to another conveyor and some event such as the board leaving the conveyor occurs. High-speed video can be always in progress, and the detection of the board leaving the system can be used to trigger the transfer to permanent storage of the video. Events of concern can be safety issues and the triggered snapshot method can be used to help eliminate potential life-threatening situations. The triggered snapshot method is particularly useful for discovering the causes of unusual events.

Data can be stored around trigger in three ways:

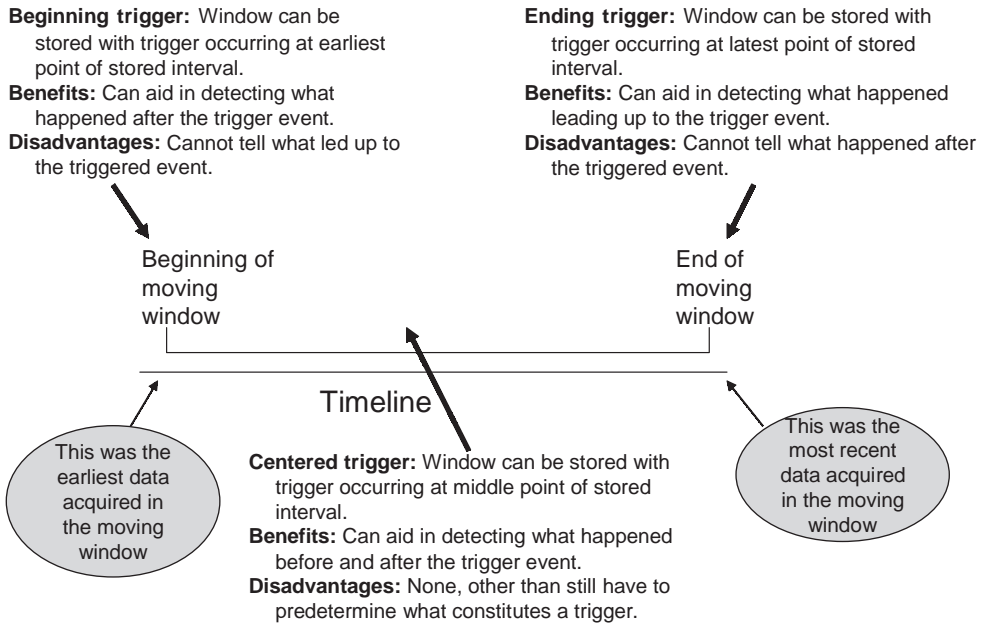


FIGURE 50.5 Relationship between trigger and moving window.

This has the advantage of allowing monitoring and analysis of high-speed events and still capturing some data to enable determining some data relationships. This is most useful if the engineer has some idea of what events may yield valuable relationships. It is much less useful when events, triggers, or relationships are unknown or unexpected. Sampling data at slower intervals may serve to allow accidental capture and identification of useful relationships, but the work required to find those relationships is much higher and of questionable probability of success.

As an example, in the sawmill many variables are changing state at high speed. For diagnostic purposes, it is valuable to see a high-speed snapshot of states of photo eyes compared to saw drops, gate openings, and grade decisions. However, the volume of data is normally too great for storage and later analysis. There are some events that are of more importance than others, such as when a gate is opening early or a saw is failing to drop consistently. These can often be recognized and the data captured in the window of time before and after the event can be stored, allowing later analysis of what led to the event and what happened shortly afterward. Some characteristic data can be summarized for each time window, stored, and used later for analysis, such as the number of photo eye changes, number of saw drops, and number of gate openings. More complex relationships can be tallied to aid in diagnostics, such as number of gate openings for grade 2. The more complex the relationship, the more difficult the programming task to ensure capturing the incidence to storage. A typical pattern is to collect process variables that may be of interest, often from a programmable logic controller. As a given problem begins to be identified, additional logic can be added to examine relationships between process inputs and sequences of events, creating a new variable

TABLE 50.4 Example of Storing Multiple Variables in Files

Filename	Process Data 2004-03-12, 14, 23, 05
Data in file	Timestamp, Temperature A, Temperature B, Temperature C, Rate 1, Rate 2, Rate 3
	2004-03-20, 12:34, 15, 14, 15, 12, 13, 13
	2004-03-20 12:35, 15, 14, 14, 12, 13, 13
	2004-03-20 12:36, 16, 14, 13, 12, 13, 13
	2004-03-20 12:37, 15, 14, 12, 12, 13, 13
	2004-03-20 12:38, 15, 14, 11, 12, 13, 13
	2004-03-20 12:39, 14, 14, 11, 12, 13, 13
	2004-03-20 12:40, 15, 14, 10, 12, 13, 13
	2004-03-20 12:41, 15, 14, 10, 12, 13, 13
	2004-03-20 12:42, 16, 14, 10, 12, 13, 13
	2004-03-20 12:43, 15, 14, 09, 12, 13, 13

that reflects some attribute of that relationship. The data collection system can store the results of that variable, such as the sum of the number of times it occurred in some larger time interval, allowing it to be low enough in volume to be mass stored for later analysis.

50.5.2 File Storage

An easy way to store data is in a file, often a comma-delimited file. This is easy to program and can easily be imported into analysis tools such as spreadsheets. It is not well suited to the compression techniques mentioned earlier because of the complexity of storing and interpreting data. However, for storing records of multiple variables collected at a time interval this can be a very useful technique. An example is shown in Table 50.4.

In the above example, the filename was chosen so that it would be unique. The date and time were concatenated together, eliminating every invalid character with an underscore. This helps to prevent files from being overwritten accidentally and facilitates store-and-forward techniques described below. Files can be sorted by date or name. Often, they are stored in a directory structure so the number of files in any one directory does not get too great. This speeds up file search in a given directory but can make programs more complex that search for files across directories. An example directory structure is the following:

```
C:
DataDirectory
  2003
    11-files created during November 2003 are stored here
    12
  2004
    01
    02
    03
```

Archiving files to backup media is easy in this file organization because one only needs to reference the particular directory for the time frame desired. Files can fill up mass

media and so either a manual process to check for file limits, backup old files, and delete them must be instituted or a program to provide the same functions would need to be created.

A major deficiency with a file-based system is that when the time range of a search is larger than one file, the analysis can become very difficult. One may be searching for events, specific time frames that go across file boundaries, or relationships between variables that may not be effectively evaluated within the time frame of one file. The analysis task usually consists of importing a number of files into some analysis tool and then using the analysis tool to look for relationships. This means that the importation process has to include the organizing of data and identifying relationships between events, a difficult task at best. A common tactic is to import data via a script or macro for a given time range, so the user only has to specify a beginning and ending time.

50.5.3 Database Storage

Database technology has been improving for many years, resulting in database management systems being increasingly the storage tool of choice for data acquisition systems. Database management systems provide organization tools, compression of data, access aids in the form of indexes, and easy access for analysis tools. A special benefit of database management systems is that they allow the combination of discrete data and time-based data collected on different time intervals. Relational databases are now the dominant database management system type. Data are organized in tables. Each table is composed of a set of rows, each row having a fixed set of columns. Indexes are provided to speed access to data. In data acquisition systems, the designer often adds a time stamp column to each row to facilitate retrieval and analysis of data. An example of a simple database is given in Table 50.5.

The TimeData table contains data that are sampled every second, whereas the BatchEvent table contains a record for each batch that has occurred. The SQL language is a common language used to examine and extract data in the tables. An example of its power is that a query can be constructed to use the BeginBatchTime and EndBatchTime to extract data from the TimeData table and combine it with related batch events in the BatchEvent table. A sample query to combine event- and time-based data is as follows:

TABLE 50.5 Example Database Structure

Time Data Table		Batch Event Table	
Timestamp	Datetime	Timestamp	Datetime
TemperatureA	Float	BatchNumber	Integer
TemperatureB	Float	MixPercent	Float
TemperatureC	Float	InputMaterialAQty	Integer
Rate1	Integer	InputMaterialBQty	Integer
Rate2	Integer	OutputProductType	Varchar
Rate3	Integer	OutputProductQty	Integer
		BeginBatchTime	Datetime
		EndBatchTime	Datetime

```

Select BatchEvent.Timestamp, Batchevent.BatchNumber,
Batchevent.MixPercent,
Min(TimeData.TemperatureA),Max(TimeData.TemperatureA)
From BatchEvent, TimeData
Where BatchEvent.OutputProductType( ' BENCH' and TimeData.Timestamp
between BatchEvent.BeginBatchTime and BatchEvent.EndBatchTime
Group by BatchEvent.BatchNumber
Order by BatchEvent.BatchNumber

```

The above query searches for batches that created a certain output product type (BENCH) and reports the maximum and minimum temperatures from those batches. This can facilitate research, for example, on what conditions lead to the best yield of a particular product.

The ease of performing this operation is a particular advantage of relational databases. There are some disadvantages, including overhead due to the access methods, extra space requirements due to the creation of indexes, and costs and complexity associated with the database management system. Indexes may add as much or more than 100% to the size of a database. Old data must be managed and removed as with any other storage system. This typically is via an automated program, since the structure is not as simple as just looking for the file creation date of a file.

50.5.4 Using Third-Party Data Acquisition Systems

When data storage is fairly simple, it is not hard to store data in the aforementioned methods, but when one is using sophisticated methods of data compression, it is recommended that systems that have robust implementations of those methods be used rather than attempting to reinvent the wheel. They can be quite complex to implement reliably. Transfer of data to those systems can occur through a variety of methods, including creation of files that are captured by the other systems, insertion of data into standard interfaces such as OPC or message buses, or insertion into database tables which are monitored by the other systems (often ODBC links). Third-party systems often have software development kit (SDK) interfaces that allow the engineer with some programming skills to store data directly into the system.

Process historians are optimized for storing time-based data. A technique used to provide some relational capability to the data is the following:

- store-related data at exactly the same time stamp (time stored in the database for when the data elements were collected),
- treat data stored with the same time stamp as being part of the same record,
- select a set of these “records” based on a time range,
- search a variable for some attribute value (e.g., having some value or range of values),
- provide to the display system the values of some related variable in the same record having the same time stamp as the desired attribute variables.

This is functionally the same as performing a relational database query on a set of records in a table, with criteria based on values in some columns.

50.6 DATA DISPLAY AND REPORTING

There are a variety of ways to reference and display data acquired from a sensor and stored in suitable media. The current value can be inspected, values can be stored for inspection later, values can be trended, alarm conditions can be detected and reported, or some output back to the process can be performed.

50.6.1 Current-Value Inspection

Often, one wants to see the data as they are being collected. This can be of critical importance in experiments which are hard or costly to repeat, allowing the researcher to react to situations as they occur. As it is collected, each data item will be called the current value for that sensor. Current data are usually stored in high-speed storage (the computer main memory). As new values are obtained, they replace the value from the last reading. The collection rate can vary widely (Table 50.6).

For instance, detecting the profile at 10-mm intervals for a log moving at 100 m/min requires values to be obtained for each sensor 167 times per second. In continuous processes, data may only need to be acquired once per minute, as in monitoring the level of a large vat. It is useful to remember that human reaction time is in terms of tenths of a second, so displaying data at a faster rate would only be useful if it were easier to program the data. Do not waste time and energy attempting to record data at a high frequency if the only reason is for display to an operator, even if the operator must immediately react to an alert. Human-machine interfaces often show data changes at the time the new value arrives from a sensor. They have display elements that are tied to sensing points. Process historians (data acquisition systems architected for the long-term storage of process data) provide tools to extract data and present the data to the analyst. Their display systems normally provide update tools that automatically refresh the user's display at some display refresh rate (often in terms of seconds, such as 10 s). The data being collected by the historian may be changing faster, and the data may be stored at a faster rate, but the display normally is still refreshed at the standard refresh rate. It is useful to have the time displayed when the value was collected, as there is often a time delay between the collection and the display of data values. This is especially true where the data collection system may be disconnected temporarily from the display system. The data may come back in a rush when the link is reconnected and is useful for the user (and systems performing analysis) to provide a context for what time the data represent.

TABLE 50.6 Data Collection Rates: Examples

Type of Operation	Time per Event
Discrete manufacturing operations	
Assembly line manufacturing/assembly	0.01 to multiple seconds
Video image processing	0.001 s
Parts machining	0.002–0.02 s
Continuous processes	
Paper machine	1–60 s
Boiler	Several seconds to several minutes
Refinery	Seconds to minutes
Dissolving operations	1 s–20 min

50.6.2 Display of Individual Data Points

Display of the data is normally in text or some simple bar graph representation. Other techniques include button or light indicators where color may represent the state of some value or range of values of the current value. Coded values may take the current value and translate it into some form that provides more value to the user such as zero being translated to the string “FALSE” on the display.

Often, one can better understand the data being obtained by using an analog representation (Bailey, 1982, pp. 243–254; Tufte, 1983). This involves representing the measured quantity by some other continuously variable quantity such as position, intensity, or rotation. A common example is the traditional wristwatch. The hours and minutes are determined by the position of a line indicator on a circle. A common analog representation for data acquisition is the faceplate. This is a bar graph where the height of the bar corresponds to the value being measured. Often, lines or symbols may be overlaid on the bar to indicate high or low ranges. A frequent indicator is the meter. A needle rotates in a circle with the degrees of movement corresponding to the value obtained from a sensor. Many voltmeters use this technique (Figure 50.6). Increasingly sophisticated calculations can be established to translate a flow rate, for example, into a cost number, providing the user with immediate feedback on the costs being incurred by the current process rate.

A common technique for representing trends of current value is to create a simple array and plot it as a trend line on the display. As new values are gathered, the array values are shifted through the array, with old values shifted out at one end of the array while the new values are shifted in at the other. This is a simple technique that provides some of the benefits of data storage without requiring the complexity of actually storing data in mass storage and managing it.

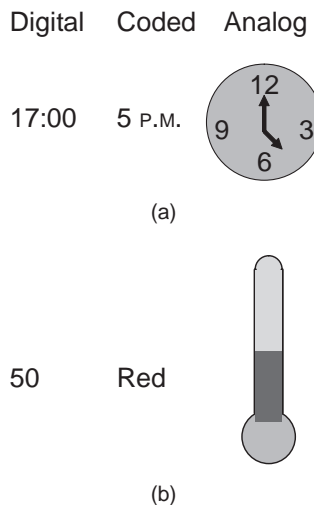


FIGURE 50.6 Comparison of digital, coded, and analog data representation: (a) time; (b) temperature, °F.

50.6.3 Display of Historical Data

There are two main issues with display of historical data:

- selection of the data,
- the representation the data will have.

50.6.3.1 Selection of Historical Data Selection of historical data involves several factors, including time frame and attributes of the data. Identifying a time frame is probably the most common activity in selecting historical data. How the data are updated can be an important consideration when comparing third-party historians.

The time frame is often referenced by a span (such as a number of hours) and a starting point which can be absolute time (e.g., 2004-12-14 16:22:03) or relative time (e.g., -4H for starting 4 h in the past). Another option is to provide an absolute start time and an absolute end time. It is common to have the time frame updating (moving forward with time) if the start point is relative (but check your particular vendor's software for their practice) and to be fixed if the start point is absolute. For example, if the span is 2 h and the start point is -2 h at 15:00, the starting point on a trend line would be 13:00 and the ending point would be 15:00. Ten minutes later the starting point on the trend line would be 13:10 and the ending point would be 15:10.

For relational data, there is often a desired time frame as described earlier (to restrict the size of the data to be searched) and some selection criteria based on characteristics of the data itself or of related data, including events. For example, one may wish to find those manufactured units for the past month that were for product *X* and see how many had quality defects. As described earlier under third-party data acquisition systems, there are techniques for selecting data from process history databases that approach (but do not equal) relational capability.

50.6.3.2 Representation of Historical Data The primary difference between current data and historical data is that there are multiple data points, normally with an order defined by the time they were acquired (for process history data) and/or by their relationship to other variables and events (for relational data). Once the time frame and relationships are selected, the data must have some method of representation on a display.

Historical data have a number of potential representation techniques. Multiple data values can be combined into a single number such as average, standard deviation, mode, maximum, minimum, range, variance, and so on. These can then be represented by techniques for individual data elements as described earlier. The equations in Section 50.4.4 are examples of summary statistics formulas.

The simplest form to represent historical data is the list. Create a column for each variable of interest. If they are all collected at the same time (the "records" described above), then each time data were collected can be used as the first value on the left in each row. The data for each variable of interest can be placed on the row corresponding to its time stamp, similar to that for files shown in Table 50.4.

Where data were collected with different time stamps, a new row can be created whenever a new value is obtained and values only placed in the column-row combinations where there is a corresponding time stamp between the row's time stamp and the time stamp of the variable in question. While useful, the problem with this is that there are now holes in the data. This list is very useful when viewing a trend line or graphical tool to validate numbers, to verify time stamps, and diagnose problems with collection.

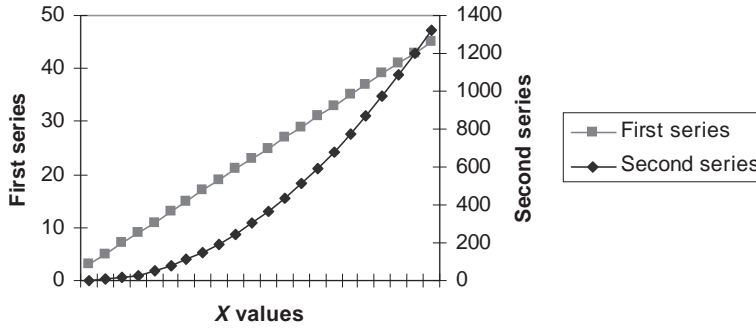


FIGURE 50.7 Multiple-axis chart.

Often, one desires to view the relationships between data and time or other variables. Trend plots typically are used to show the relationship between variables and time. The X axis is typically represents the time range, and the Y axis represents the value of the desired variable. The range of data on the Y axis can vary depending on how one wishes to view the data. The maximum and minimum of the data can be used to set the top and bottom of the range of the Y axis, respectively. This can have two undesired effects. It may make small movements in data appear to be very large when the range is small. In cases where there are spikes in the data where a value is disproportionately high or low, representing the Y axis based on the maximum and minimum could make it difficult to view normal variation in the data. One has to understand the potential use of the data to choose the Y -axis scale appropriately.

When more than one variable is shown on a trend chart, the selection of scale of the Y axis becomes more complex. If all the variables are representative of the same domain, such as all temperatures, then perhaps the same Y axis can be used for all of them. Often, however, the viewer is attempting to compare relative variations in data, sort of a poor man's correlation analysis. In this case, it may be useful to have multiple Y axes and select the range of each of them such that they represent the range of one or more of the variables being viewed (Figure 50.7).

Another approach to compare variation between two variables is to use one variable for the X value and the other variable for the Y value (an X - Y chart). This is useful when two variables are related by sample time or some other selection technique that results in a paired relationship between the two variables. The correlation function in Section 4.4 represents the mathematical correlation between two variables and can be used to determine the strength of that relationship. Chapter 1 discusses correlation and the calculation of the line through a distribution of data.

50.7 DATA ANALYSIS

50.7.1 Distributed Systems

Distributed systems are a powerful approach to data acquisition systems because they combine some of the best of both stand-alone and host-based systems. The data acquisition portion is located on a small processor that has communication capability to a host computer system. The small system collects the data, possibly reducing some to a more

compact form, and then sends the data to the host systems for analysis. The host system can analyze the data when it has the available time to do so. Only the data acquisition portion needs to be very responsive to the process. If the data acquisition task gets too big for the small system, the cost of expansion is limited to moving the data acquisition software to a new computer or splitting it up over several computers and changes to the host computer portion are not required. The major disadvantage of distributed systems is that they suffer from a more complex overall architecture even though the individual parts are simple. This leads to problems with understanding error sources and increases the potential errors because of more parts. Unless the communications are designed carefully, the messages sent between the small systems and the host system may be inflexible, causing increased effort when one wants to change the type of data being collected. Distributed systems may be expensive because of the number of individual components and the complexity required but often fit well with environments where one already has a host computer.

50.7.2 System Error Analysis

The errors that can occur at different stages in the data acquisition process must be analyzed, as they can add up to make the data meaningless. For instance, one may have very accurate sensors, but by the time the data reach the host computer they might have been converted into integer data or from real to integer and back to real again. This can lead to dangerous assumptions about the accuracy of the received numbers, because each conversion can cause rounding or other errors. It is the responsibility of the person setting up the acquisition system and the analyst to examine each source of potential error, discover its magnitude, and reduce it to the point where it will not have a significant impact on the conclusions to be derived from the data. Use of the filtering techniques described earlier under data collection can be of use to eliminate random error. It is not within the scope of this chapter to cover system error analysis, but Chapter 1 gives some foundation.

50.8 DATA COMMUNICATIONS

Data communications are involved in many aspects of data acquisition systems. The communications between the sensing and control elements and data acquisition devices, as well as the communications between the data acquisition system and other computer systems, can be carried out in many ways. This section will cover some aspects of communications, especially as they pertain to computer systems.

50.8.1 Serial Communications

A serial communication link means that data sent over a communications line is spread out over time on one physical data path. For instance, if a character is sent from a sensor to a computer, each bit making up the character (normally eight bits) will be sent one after the other (Table 50.7). This is often useful for low-cost, low-speed (usually less than 10,000 cps) rates of data transfer.

TABLE 50.7 Time Sequence of Bits Sent over Serial Communication Line

Character “A” is (in bit form)	
Bit number 6 5 4 3 2 1 0	
Bit value 1 0 0 0 0 0 1	
The communications are using the RS232C communications standard and sending the ASCII character A	
Bit Value	Time
Start bit	0 s after start
1 (bit 0 of A)	1/9600 s after start
0 (bit 1 of A)	2/9600 s after start
0 (bit 2 of A)	3/9600 s after start
0 (bit 3 of A)	4/9600 s after start
0 (bit 4 of A)	5/9600 s after start
0 (bit 5 of A)	6/9600 s after start
1 (bit 6 of A)	7/9600 s after start
Parity bit	8/9600 s after start
Stop bit	9/9600 s after start

50.8.2 Parallel Communications

A serial communication link may not require very many wires, but the time spent to transfer data can add up. A way to improve the speed of communications is to use parallel communication links. This is done by having a number of wires to carry data. For instance, sending the same “A” over a nine-wire bus would only require one transfer (Table 50.8).

TABLE 50.8 Time Sequence of Bits Sent over Parallel Communications Interface

Character “A” is (in bit form)	
Bit number 6 5 4 3 2 1 0	
Bit value 1 0 0 0 0 0 1	
The communications are using a hypothetical nine-line parallel communications bus sending the ASCII character A	
Bit Value	Time
Start bit	0 s after start
1 (bit 0 of A)	0/9600 s after start
0 (bit 1 of A)	0/9600 s after start
0 (bit 2 of A)	0/9600 s after start
0 (bit 3 of A)	0/9600 s after start
0 (bit 4 of A)	0/9600 s after start
0 (bit 5 of A)	0/9600 s after start
1 (bit 6 of A)	0/9600 s after start
Parity bit	0/9600 s after start
If the bus could handle the same rate of change of bits as the serial interface, then the next character could be sent 1/9600 s after the first character (the A)	

50.8.3 Networks

Ethernet with Transmission Control Protocol/Internet Protocol (TCP/IP) has become the dominant communications network protocol for data collection. There are proprietary process control and data acquisition networks that serve special purposes, but Ethernet has proven to be versatile for everything from office communications to data collection from smart sensors. Many computers can be connected to the same network segments. The use of switches and routers provides ways to isolate and limit communications to improve performance and security. Firewalls provide filters and protection for entire classes of messages and sources.

While Ethernet cannot guarantee delivery (being based on a collision detection and retransmit strategy), it has been shown to provide excellent response to moderate network activity. Communications speeds are regularly being improved to provide an even greater range of applicability.

50.8.4 OSI Standard

The International Standards Organization has developed a set of standards for discussing communications between cooperating systems called the Open Systems Interconnect (OSI) model (see Table 50.9). This defines communications protocols in terms of seven layers (American National Standards Institute, 1981). While not providing for specific interface protocols, the OSI model has had a significant impact on communications because it has provided a framework for compartmentalizing aspects of communications to allow the handoff of information from one device to another in a standard way. For instance, the transmission of data from one media type to another (such as copper wire to fiber to satellite to copper wire and then to wireless) is a result of standards enabling the seamless transfer of messages in a way that is transparent to the user.

50.8.5 OPC Standard

A recent standard of use in manufacturing is the OPC (OLE for process control) standard, which provides for a standard way of communicating with process equipment. It is sponsored by the OPC Foundation and originated as an extension for process control from the Microsoft OLE functionality (<http://www.opcfoundation.org>).

Functions provided by OPC include ability to browse the variable database of a device and monitor data on demand or when events occur. The capability of OPC has been

TABLE 50.9 Open Systems Interconnect Model

Layer	Principle	Example
7. Application	Application	Millwide reporting
6. Presentation	Display, format, edit	Convert ASCII into EBCDIC
5. Session	Establish communications	Log onto remote computer
4. Transport	Virtual circuits	Make sure all message parts got there in order
3. Network	Route to other networks	Talk to Internet
2. Data link	Correct errors	Send character downline
	Synchronize communications	Send Ack-Nak
1. Physical	Electrical interface	Wire and voltages

expanded to work with Web communication methods such as XML and cover complex data such as record structures. The power of OPC is that data from an instrument can be available via a standard network interface so that any data acquisition program that uses the OPC interface can gain access to any OPC device. The need to know the protocol of each device or adhere to the wiring of specialized communications or use custom database access methods is eliminated through the use of a standard protocol. Multiple programs can be simultaneously monitoring the same piece of data, performing different functions at different time intervals, as events occur. The last point is particularly significant, because much of the work of data acquisition systems is spent in polling for changes in data or otherwise attempting to determine when an event has taken place. A program can subscribe to an OPC item and it will be notified when the item changes value, reducing the complexity of monitoring data dramatically.

As an example of the power of this approach, consider the following example (Figure 50.8). A device can collect the identification from a unit of material such as unit number, color, and manufacturing date and make it available via OPC. As the unit is processed in a manufacturing center, another device collects defect counts and makes the unit number available via OPC. A human-machine interface program can monitor both sources with the same interface protocol and software and display it live for an operator to see. Simultaneously, another program can collect the defect counts and summarize them into totals. Yet another program can monitor the totals and wait for the unit number to change, triggering a transaction to a database or an email if there was a problem. The power of the OPC interface is that it provides real-time access to the data from each of the sources and multiple programs can monitor the same OPC sources to perform work. Diagnostics can monitor the same data to evaluate system processing, downtime, or quality issues. Other programs can sample and store data to log files or diagnostic files for further analysis. All can be operating in parallel with no need to understand the internals of the other programs, via a standard interface and standard protocols that support asynchronous delivery of data.

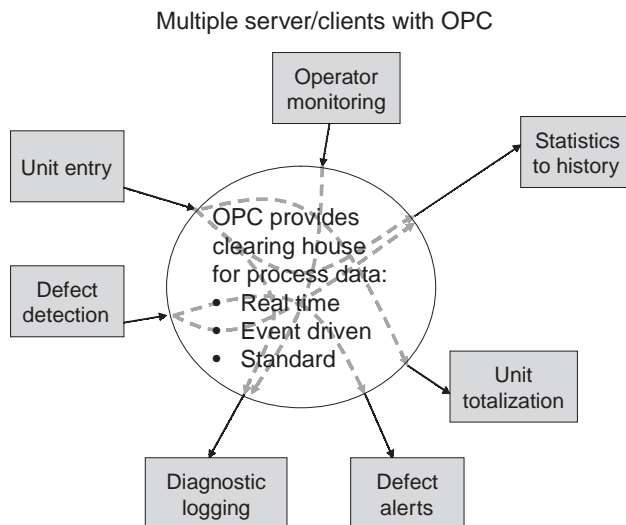


FIGURE 50.8 Example use of OPC communications.

50.8.6 Benefits of Standard Communications

When implementing data acquisition and display systems, the ability to communicate in a standard fashion can play a large role in the cost of the system. This is realized in a variety of ways:

1. Different sensors can be connected to a system without having to buy a whole new system.
2. Data can be sent to other systems as needed for further processing.
3. As technology or need changes, portions of a system can be mixed and matched.
4. Increased competition from vendors tends to bring prices down.
5. A standard will have many people using products based on the standard, resulting in more vendors and greater availability of parts with a greater variety of options.

50.9 OTHER DATA ACQUISITION AND DISPLAY TOPICS

50.9.1 Data Chain

As materials and parts move through a manufacturing operation, the data collected are separated by time and type of data. Combining that data together presents a number of challenges. One can consider the first piece of data collected about some object to be the beginning of a chain of data and each successive data acquisition point in the manufacturing process to be a link in that chain until the end of the chain, where the last piece of data about a manufactured item is collected. Often, the steps of the manufacturing process may proceed from raw material to some intermediate work-in-process unit, to some other step that may be time based, then to some other step that may be finished-unit based. Many varieties of the above exist. Each link can represent one of the following:

- Data collected from a start time to an end time.
- A set of attributes about a particular manufacturing unit with an associated time of processing.

Often referred to as the genealogy, the steps can be linked together through:

- Some assumptions based on time stamp relationship of one link back to the previous one.
- Recording of units that entered or left the time-based portion of the process and the beginning and ending of the entry time.
- Some assumptions about the mixing of elements of the manufactured item.

Using the techniques described above for combination of time-based and event data, a set of data for the whole life cycle of a manufactured item can be created. Where mixing occurs, the data will be less accurate but may provide clues as to the factors that went into the final product, such as proportions of additives. As an example, Table 50.10 shows various queries that can be combined to provide one picture of all data sources for a manufactured unit.

TABLE 50.10 Data Chain Sample Queries

State in Process	Important Data (Typical)	Example Data	Simplified Query to Combine Data with Previous Step
Raw material inventory	Identifiers: raw material batch ID Data: supplier, quality, time in inventory	Lot 1: 5% rejects; lot 2: 7% rejects	
Intermediate goods processing	Identifiers: raw material lots consumed; raw material batch ID; start and end time of entry to process; time delay through process Data: piece count, rejects, downgrades, process characteristics such as rate, temperature, modification to materials	3:00–4:00 raw materials from lot 1; 1000 pieces; 10 rejects; 200° 4:00–5:00 raw materials from lot 2; 1045 pieces; 14 rejects; 205°; 10 residence time in process	Use raw material batch ID to match to raw materials characteristics; material run at 4:00–5:00 had 7% rejects when delivered to plant
Intermediate package creation	Identifiers: raw package ID, start/end time of creation Data: piece count, package dimensions	Package PR1 created: 3:30–4:10; 100 pieces; package PR2 created: 4:10–5:10; 150 pieces	Use time of manufacture, time lag through process to identify characteristics from intermediate process and raw materials lots; package PR1 was processed at 200°, was created from lot 1 and was from a lot that had 5% rejects detected when delivered to the plant
Intermediate inventory	Identifier: intermediate package ID Data: location in inventory	Package PR1 location warehouse 1	Use ID of intermediate package to match to package at intermediate package creation; package PR1 had 100 pieces
Finished goods processing	Identifier: finishing batch ID, intermediate package consumed, start/end time of consumption; time delay through process Data: piece count, rejects, downgrades, process characteristics such as rate, temperature, modification to materials	Intermediate packages consumed at breakdown: 5:30–6:00 PR1; 95 pieces; 400° 6:00–6:30 PR2; 147 pieces; 390° 6:30–6:40 PR3; 25 pieces; 395° 6:40–7:00 PR4; 40 pieces; 410° Typical residence time: 15	Use ID of intermediate package to match to intermediate inventory; in this example, package PR1 came from warehouse 1, lost 5 pieces in consumption
Finished goods package creation	Identifiers: finished package ID; start/end time of creation Data: piece count, grade, package dimensions, customer order	Finished package PF1 created from 6:00 to 6:30; 40 pieces; prime grade; order PO5670	Use time of manufacture, time lag through process to identify characteristics from breakdown, and which raw packages sourced this finished package. There may be significant mixing. In this example, PR1 and PR2 would be sources for PF1. PF1 was possibly created at 400°, stored in warehouse 1, lost 5 pieces when loading into finished goods process, was probably processed at 200°, and was probably from lot 1

Starting at the finished-good item, the batch ID of the previous step acts as a link into the range of time data in the previous step. If the list of items broken down is retained from the prior step, then those can be used to link back to the previous time frame.

Depending on the amount of mixing, the results will be more or less indicative of what actually happened. The smaller the batch sizes, the easier the tracking back to source data will be. Time lags between process steps can dramatically impact the ability to assume when the raw materials for a particular item were processed.

50.9.2 Web Programs and Interfaces

Web interfaces have improved to the point that user interfaces can be written in Web browser screens. This eliminates the effort and organization required to deploy code across a company. The application is written for a Web interface. When the user uses a Web browser to access the page, functionality is downloaded to the user's page or is executed in such a fashion as to obtain the results of a query and transmit a data page to the user. The developer does not have to get involved in the process of manually installing software on the user's machine. This reduces the demand on the desktop computer and allows the developer to make a change and have it proliferated to all users when they next reference the Web.

50.9.3 Configuration Versus Implementation

As a general rule of thumb, third-party data acquisition and storage systems provide configurable tools for acquisition and display. For simple applications which do not require great flexibility in program functions, such as generating alarms, unusual graphics types, control of the process, or integration into larger systems, it is appropriate to use these, often simple, question-answer or menu-type systems. When the system must be very flexible or customized, it may be more appropriate to write a custom program. When considering this approach, be cautious, for the cost of implementing, a program is often much higher than expected. For instance, if one wanted to perform simple data acquisition and storage from a sensor, the cost to write a program would probably be higher than buying a small off-the-shelf system and entering the parameters for data collection. Writing a program involves analysis, design, development, debugging the program, and testing of results. The cost of documenting a program is often a large unplanned cost. If the results are intended to be used to make economic or process-related decisions, then the program must be tested carefully. Additionally, maintenance of the program can be quite expensive. Someone must be trained in the technologies used to build the program, the logic of the program, and the installation of the program. Another factor to consider is that costs of improvement of third-party software are borne by many customers and driven by many customers. The net result to the user of this software is that it is normally constantly improving, constantly tested, and maintained by a group of developers whose primary job is software development. One reason to build and maintain software internally is that a company can keep special knowledge within the company and thus maintain competitive advantage.

50.9.4 Store and Forward

When data acquisition and data storage are on two separate machines, it is important to provide methods to retain data in case the link between systems is broken. Message buses

provide automated methods of maintaining a link between data acquisition systems. The developer inserts data into the message bus. If the link between the two systems is broken, the message bus queues up the data messages on the collection machine. When the data storage machine connection is reestablished, the message bus passes on the data to the data storage machine.

When a message bus is not available or feasible, a simplified mechanism can be created where a file representing each sample of data is created. If the data collection and data storage system are linked, then the data storage system monitors the directory of the collector for a new data file. If the data storage system detects one or more files on the data collection computer, then it will process them into storage. If the link is broken, then the files build up until the link is reestablished. A related technique is to store data in a database or similar mechanism on the data collection computer and scan it periodically for missing data from the storage computer. This is particularly useful when connection to the data collection computer is unreliable.

50.9.5 Additional Communications Topics

When considering transmission media, some points may provide value to the engineer. Fiber-optic cabling is less sensitive to noise than other transmission media. Wireless access points provide increased flexibility in positioning of sensors and greatly reduce wiring costs. Particularly, if one wishes to collect data from sites that may move, such as environmental sampling sites, the costs of wiring and rewiring can be quite significant. Using wireless transmission technology, it eliminates much of the wiring costs and facilitates moving the sensors from one location to another. Wireless transmission has a set of concerns that must be taken into account by the engineer, including security, since other units can monitor signals (still evolving) and ability to be jammed.

50.10 SUMMARY

The tremendous change in technology for data acquisition and display systems since this chapter was first written has driven us to take a different approach than with the first edition. The technologies for data acquisition and display have become more standardized. Engineers are increasingly reliant upon and versed in computing technologies. The combination of data from various sources into an integrated view of the process has facilitated process improvement and leads to competitive advantage.

This chapter has attempted to provide tools and techniques to aid in the acquisition, storage, and manipulation of process data, expanding from the previous edition into techniques to aid in the manipulation of data for integration and analysis.

REFERENCES

- American National Standards Institute. *Open Systems Interconnection—Basic Reference Model*, Draft Proposal 7498 97/16 N719. New York: American National Standards Institute; 1981.
- Bailey, RW. *Human Performance Engineering: A Guide for Systems Designers*, Englewood Cliffs (NJ): Prentice-Hall; 1982.

- Beyer, WH, editor. *CRC Standard Mathematical Tables*. 24th ed. Boca Raton (FL): CRC; 1976.
- Crossley, ML. *The Desk Reference of Statistical Quality Methods*. Milwaukee (WI): ASQ Quality Press; 2000.
- Date, CJ. *An Introduction to Database Systems*. 5th ed. Addison-Wesley; 1990.
- Johnson, CD. *Microprocessor-Based Process Control*. Englewood Cliffs (NJ): Prentice-Hall; 1984.
- Liptak, BG. System accuracy. In: Liptak BG, editor. *Instrument Engineer's Handbook*. Vol. 1, 4th ed. Boca Raton (FL): Process Measurement and Analysis, CRC; 2003.
- Murrill, PW. *Fundamentals of Process Control Theory*, Research Triangle Park (NC): Instrument Society of America; 1981.
- Tufte, ER. *The Visual Display of Quantitative Information*, Cheshire, England: Graphics; 1983.
- Wright, JD, Edgar TF. Digital Computer Control and Signal Processing Algorithms. In: *Real-Time Computing*. Mellichamp DA, editors. New York: Van Nostrand Reinhold; 1983.

Magazines that Carry Relevant Information

- Control Engineering International*: <http://www.controleng.com/>.
- Design Engineering*: <http://www.designengineering.co.uk/>.
- IEEE Control Systems Magazine*: <http://www.ieee.org/organizations/pubs/magazines/cs.htm>.
- Industrial Technology*: <http://www.industrialtechnology.co.uk/>.
- Instrumentation and Automation News*: <http://www.ianmag.com/>.
- Pollution Engineering Online*: <http://www.pollutionengineering.com/>.
- Scientific Computing and Instrumentation*: <http://www.scamag.com/>.
- Sensors Magazine*: www.sensorsmag.com.

51

MATHEMATICAL AND PHYSICAL UNITS, STANDARDS, AND TABLES*

JACK H. WESTBROOK

51.1 Symbols and abbreviations

Bibliography for letter symbols

Bibliography for graphic symbols

51.2 Mathematical tables

51.3 Statistical tables

51.4 Units and standards

51.4.1 Physical quantities and their relations

51.4.2 Dimensions and dimension systems

51.4.3 Dimension and unit systems

51.4.4 The international system of units

51.4.5 Application of SI prefixes

51.4.6 Other units

51.4.7 Length, mass, and time (English units and standards)

51.4.8 Standard of time

51.4.9 Force, energy, and power

51.4.10 Thermal units and standards temperature

51.4.11 Quantity of heat and some derived quantities

51.4.12 Chemical units and standards

51.4.13 Theoretical, or absolute, electrical units

51.4.14 Internationally adopted electrical units and standards

Bibliography for units and measurements

51.5 Tables of conversion factors

51.6 Standard sizes

51.6.1 Preferred numbers

51.6.2 Gages

51.6.3 Paper sizes

51.6.4 Sieve sizes

51.6.5 Standard structural sizes—steel

51.6.6 Standard structural shapes—aluminum

51.7 Standard screws

51.7.1 Nominal and minimum dressed sizes of American Standard Lumber

*This chapter is a revision and extension of Sections 1 and 3 of the third edition, which were written by Mott Souders and Ernst Weber, respectively. Section 51.4.4 is derived principally from ASTM's *Standard for Metric Practice*, ASTM E380-82, Philadelphia, 1982 (with permission). Section 51.6.1 is derived from *MIS Newsletter*, General Electric Co., 1980 (with permission).

Handbook of Measurement in Science and Engineering. Edited by Myer Kutz.
Copyright © 2013 John Wiley & Sons, Inc.

51.1 SYMBOLS AND ABBREVIATIONS

TABLE 51.1 Greek Alphabet

A	α	Alpha	H	η	Eta	N	ν	Nu	T	τ	Tau
B	β	Beta	Θ	ϑ θ	Theta	Ξ	ξ	Xi	Y	υ	Upsilon
Γ	γ	Gamma	I	ι	Iota	O	o	Omicron	Φ	ϕ	Phi
Δ	δ	Delta	K	κ	Kappa	Π	π	Pi	X	χ	Chi
E	ε	Epsilon	Λ	λ	Lambda	P	ρ	Rho	Ψ	ψ	Psi
Z	ζ	Zeta	M	μ	Mu	Σ	σ ζ	Sigma	Ω	ω	Omega

TABLE 51.2 Symbols for Mathematical Operations^a

<p><i>Addition and Subtraction</i></p> <p>$a + b$, a plus b $a - b$, a minus b $a \pm b$, a plus or minus b $a \mp b$, a minus or plus b</p> <p><i>Multiplication and Division</i></p> <p>$a \times b$, or $a \cdot b$, or ab, a times b $a \div b$, or $\frac{a}{b}$, or a/b, a divided by b</p> <p><i>Symbols of Aggregation</i></p> <p>() parentheses () parentheses { } braces –vinculum</p> <p><i>Equalities and Inequalities</i></p> <p>$a = b$, a equals b $a \approx b$, a approximately equals b $a \neq b$, a is not equal to b $a > b$, a is greater than b $a < b$, a is less than b $a \gg b$, a much larger than b $a \ll b$, a much smaller than b $a \geq b$, a equals or is greater than b $a \leq b$, a is less than or equals b $a \equiv b$, a is identical to b $a \rightarrow b$, or $a = b$, a approaches b as a limit</p> <p><i>Proportion</i></p> <p>$ab = cd$, or $a :: b :: c :: d$, a is to b as c is to d $a \propto b$, $a \sim b$, a varies directly as b %, percent</p>	<p><i>Powers and Roots</i></p> <p>a^2, a squared a^n, a raised to the nth power \sqrt{a}, square root of a $\sqrt[3]{a}$, cube root of a $\sqrt[n]{a}$, or $a^{1/n}$, nth root of a a^{-n}, $1/a^n$ $3.14 \times 10^4 = 31,400$ $3.14 \times 10^{-4} = 0.000314$</p> <p><i>Miscellaneous</i></p> <p>\bar{a}, mean value of a $a!$, $= 1 \cdot 2 \cdot 3 \cdot \dots \cdot a$, factorial a a = absolute value of a $P(n, r) = n(n - 1)(n - 2) \cdot \dots (n - r + 1)$ $C(n, r) = \frac{P(n, r)}{r!} = \binom{n}{r}$ = binomial coefficients i (or j) $= \sqrt{-1}$, imaginary unit $\pi = 3.1416$, ratio of the circumference to the diameter of a circle ∞, infinity</p> <p><i>Plane Geometry</i></p> <p>$<$, angle Δ, triangle \parallel, parallel \perp, perpendicular \odot, circle \square, parallelogram \therefore, therefore $^\circ$ ' ' ', degree, minute, second ' ' ', feet, inches</p>
---	---

TABLE 51.2 (Continued)

<p><i>Logarithms and Exponentials</i></p> <p>$\log a = \log_{10} a$, common logarithm of a or log of a to the base 10</p> <p>$\ln a = \log_e a$, natural logarithm of a or log of a to the base e ($e = 2.718$)</p> <p>$\log^{-1} a$, number whose log is a</p> <p>$\text{lb } x$ or $\log_2 x$ = binary logarithm of x exponential of x, $\exp x$, e^x</p>	<p>$u_{xy} = f_{xy}(x, y) = D_y(D_x u) = \frac{\partial^2 u}{\partial y \partial x}$, second partial derivative of $u = f(x, y)$ with respect to x and y</p> <p>Δy, increment of y</p> <p>dy, differential of y</p> <p>δy, variation of y</p> <p>$\sum_{i=a}^b$, summation over i from a to b</p> <p>$\lim_{x \rightarrow a}(y) = b, y \rightarrow b$ as $x \rightarrow a$</p> <p>\int, integral of</p> <p>\int_a^b, definite integral of</p>
<p><i>Trigonometry</i></p> <p>\sin, \cos, \tan $\text{cosec or csc, sec, cot or ctn}$ vers, covers</p> <p>$\sin^{-1}, \cos^{-1}, \text{etc.}$</p> <p>$\arcsin, \arccos$</p>	<p>$\left. \begin{array}{l} \text{trigonometric} \\ \text{functions} \end{array} \right\}$</p> <p>$\left. \begin{array}{l} \text{inverse of the functions} \end{array} \right\}$</p>
<p><i>Analytic Geometry</i></p> <p>$x, y, z; \xi, \eta, \zeta$, rectangular coordinates</p> <p>ρ, s, intrinsic coordinates</p> <p>ρ, radius of curvature</p> <p>s, length of arc</p> <p>r, θ, polar coordinates</p> <p>ψ, angle from radius vector to tangent</p> <p>r, θ, φ, spherical coordinates</p> <p>θ, colatitude</p> <p>φ, longitude</p> <p>r, θ, z, cylindrical coordinates</p> <p>e, eccentricity in conics</p> <p>ρ, semi latus rectum in conics</p> <p>$l = \cos \alpha, m = \cos \beta, n = \cos \gamma$, direction cosines</p>	<p><i>Vector Analysis</i></p> <p>i, j, k, unit vectors along the axes (right-handed system)</p> <p>$a \cdot b = (ab) = Sab$, scalar product of a and b</p> <p>$a \times b = [ab] = Vab$, vector product of a and b</p> <p>Vectors are indicated in print by boldfaced type.</p> <p>A , A, absolute value</p> <p>$\partial/\partial r, \nabla$, differential vector operator</p> <p>$\text{grad } \varphi, \nabla \varphi$, gradient</p> <p>$\text{div } A, \nabla \cdot A$, divergence</p> <p>$\text{curl } A, \text{rot } A, \nabla \times A$, curl</p> <p>$\Delta \varphi, \nabla^2 \varphi$, Laplacian</p> <p>$\square \varphi, \square^2 = \frac{1}{c^2} \frac{\partial^2}{\partial t^2} - \nabla^2$, D'Alembertian</p>
<p><i>Calculus</i></p> <p>$y = f(x)$, y is a function of x</p> <p>$y' = f'(x) = \frac{dy}{dx} = D_x y$, derivative of $y = f(x)$ with respect to x</p> <p>$y'' = f''(x) = \frac{d(y')}{dx} = D_x^2 y = \frac{d^2 y}{dx^2}$, second derivative of $y = f(x)$ with respect to x</p> <p>$u = f(x, y)$, u is a function of x and y</p> <p>$u_x = f_x(x, y) = D_x(u) = \frac{\partial u}{\partial x}$ —, partial derivative of $u = f(x, y)$ with respect to x</p>	<p><i>Logic and Boolean Algebra</i></p> <p>$a \in A$, a is contained in set A</p> <p>$A \cap B, A \cdot B$, logical multiplication. Intersection of set A and set B, A AND B</p> <p>$A \cup B, A + B$, logical addition. Union of set A and set B. A OR B</p> <p>$A \oplus B$, exclusive OR</p> <p>$A \supset B$, logical inclusion. Inclusion of set B in set A</p> <p>$A \ominus B$, complement of set B in set A</p> <p>\bar{A}, \overline{A}, logical complementation. NOT set A. Negation</p> <p>$\emptyset, 0$, logical impossibility. Empty (null) set. Zero state</p> <p>$I, 1$, logical certainty. Universal set. One state</p>

^aReferences: Mathematic signs and symbols for use in the physical sciences and technology, ANSI Y10.20 – 1975.

TABLE 51.3 Abbreviations^a for Scientific and Engineering Terms^b

Name of Term	Abbreviation	Name of Term	Abbreviation
absolute	abs	constant	const
acre	spell out	continental horsepower	cont hp
acre-foot	acre-ft	cord	cd
air horsepower	air hp	cosecant	csc
alternating-current (as adjective)	a-c	cosine	cos
ampere	amp or A	cosine of the amplitude, an elliptic function	cn
ampere-hour	amp-hr	cost, insurance, and freight	cif
amplitude, an elliptic function	am.	cotangent	cot
Angstrom unit	Å	coulomb	spell out or C
antilogarithm	antilog	counter electromotive force	cemf
atmosphere	atm	cubic	cu
atomic weight	at. wt.	cubic centimeter	cu cm, cm ³ (liquid, meaning milliliter, ml)
average	avg	cubic foot	cu ft
avoirdupois	avdp	cubic feet per minute	cfm
azimuth	az or α	cubic feet per second	cfs
barometer	bar.	cubic inch	cu in.
barrel	bbl	cubic meter	cu m or m ³
Baumé	Bé	cubic micron	cu μ or μ^3 or cu mu
board feet (feet board measure)	fbm	cubic millimeter	cu mm or mm ³
boiler pressure	spell out	cubic yard	cu yd
boiling point	bp	current density	spell out
brake horsepower	bhp	cycles per second	spell out or cps or Hz
brake horsepower-hour	bhp-hr	cylinder	cyl
Brinell hardness number	Bhn	day	spell out
British thermal unit ^c	Btu or B	decibel	db
bushel	bu	degree ^d	deg or °
calorie	cal	degree centigrade	C
candle	c	degree Fahrenheit	F
candle-hour	c-hr	degree Kelvin	K
candlepower	cp	degree Rankine	R
cent	c or g	delta amplitude, an elliptic function	dn
center to center	c to g	diameter	diam
centigram	cg	direct-current (as adjective)	d-c
centiliter	cl	dollar	\$
centimeter	cm	dozen	doz
centimeter-gram-second (system)	cgs	dram	dr
chemical	chem	efficiency	eff
chemically pure	cp	electric	elec
circular	cir	electromotive force	emf
circular mils	cir mils	elevation	el
coefficient	coef	equation	eq
cologarithm	colog	external	ext
concentrate	conc		
conductivity	cond		

TABLE 51.3 (Continued)

Name of Term	Abbreviation	Name of Term	Abbreviation
farad	spell out or F	inside diameter	ID
feet board measure (board feet)	fbm	intermediate-pressure (adjective)	i-p
feet per minute	fpm	internal	int
feet per second	fps	joule	J
fluid	fl	kilocalorie	kcal
foot	ft	kilocycles per second	kcps
foot-candle	ft-c	kilogram	kg
foot-Lambert	ft-L	kilogram-calorie	kg-cal
foot-pound	ft-lb	kilogram-meter	kg-m
foot-pound-second (system)	fps	kilograms per cubic meter	kg per cu m or kg/m ³
foot-second (see cubic feet per second)		kilograms per second	kgps
franc	fr	kiloliter	kl
free aboard ship	spell out	kilometer	km
free alongside ship	spell out	kilometers per second	kmps
free on board	fob	kilovolt	kv
freezing point	fp	kilovolt-ampere	kva
frequency	spell out	kilowatt	kw
fusion point	fnp	kilowatthour	kwhr
gallons per minute	gpm	lambert	L
gallons per second	gps	latitude	lat or ϕ
grain	spell out	least common multiple	lcm
gram	g	linear foot	lin ft
gram-calorie	g-cal	liquid	liq
greatest common divisor	gcd	lira	spell out
haversine	hav	liter	L
hectare	ha	logarithm (common)	log
henry	H	logarithm (natural)	log _e or ln
high-pressure (adjective)	h-p	longitude	long. or λ
hogshead	hhd	low-pressure (as adjective)	l-p
horsepower	hp	lumen	lm
horsepower-hour	hp-hr	lumen-hour	lm-hr
hour	hr	lumens per watt	lpw
hour (in astronomical tables)	h	mass	spell out
hundred	C	mathematics (ical)	math
hundredweight (112 lb)	cwt	maximum	max
hyperbolic cosine	cosh	mean effective pressure	mep
hyperbolic sine	sinh	mean horizontal candlepower	mhcp
hyperbolic tangent	tanh	megacycle	spell out
inch	in.	megohm	spell out
inch-pound	in. lb	melting point	mp
inches per second	ips	meter	m
indicated horsepower	ihp	meter-kilogram	m-kg
indicated horsepower- hour	ihp-hr	mho	spell out

(continued)

TABLE 51.3 (Continued)

Name of Term	Abbreviation	Name of Term	Abbreviation
microampere	μ a or mu a	pound-foot	lb-ft
microfarad	μ F	pound-inch	lb-in.
microinch	μ in.	pound sterling	£
micromicron	$\mu\mu$ or mu mu	pounds per brake horse- power-hour	lb per bhp-hr
micron	μ or mu	pounds per cubic foot	lb per cu ft
microvolt	μ v	pounds per square foot	psf
microwatt	μ w or mu w	pounds per square inch	psi
mile	spell out	pounds per square inch absolute	psia
miles per hour	mph	power factor	spell out or pf
miles per hour per second	mphps	quart	qt
milliampere	ma	radian	spell out
milligram	mg	reactive kilovolt-ampere	kvar
millihenry	mH	reactive volt-ampere	var
millilambert	mL	revolutions per minute	rpm
milliliter	ml	revolutions per second	rps
millimeter	mm	rod	spell out
millimicron	$m\mu$ or m mu	root mean square	rms
million	spell out	secant	sec
million gallons per day	mgd	second	sec
millivolt	mV	second (angular measure)	"
minimum	min	second-foot (see cubic feet per second)	
minute	min	second (time)	s
minute (angular measure)		(in astronomical tables)	
minute (time) (in astronomical tables)	m	shaft horsepower	shp
mole	spell out	shilling	s
molecular weight	mol. wt	sine	sin
month	spell out	sine of the amplitude, an elliptic function	sn
National Electrical Code	NEC	specific gravity	sp gr
ohm	spell out or Ω	specific heat	sp ht
ohm-centimeter	ohm-cm	spherical candle power	scp
ounce	oz	square	sq
ounce-foot	oz-ft	square centimeter	sq cm or cm^2
ounce-inch	oz-in.	square foot	sq ft
outside diameter	OD	square inch	sq in.
parts per million	ppm	square kilometer	sq km or km^2
peck	pk	square meter	sq m or m^2
penny (pence)	d	square micron	sq μ or sq mu or μ^2
pennyweight	dwt	square millimeter	sq mm or mm^2
per	(see Fundamental Rules)	square root of mean square	rms
peso	spell out	standard	std
pint	pt	steradian	sr
potential	spell out		
potential difference	spell out		
pound	lb		

TABLE 51.3 (Continued)

Name of Term	Abbreviation	Name of Term	Abbreviation
tangent	tan	volt-ampere	Va
temperature	temp	volt-coulomb	spell out
tensile strength	ts	watt	W
thousand	M	watthour	Whr
thousand foot-pounds	kip-ft	watts per candle	Wpc
thousand pound	kip	week	spell out
ton	spell out	weight	wt
ton-mile	spell out	yard	yd
versed sine	vers	year	yr
volt	V		

^aThese forms are recommended for readers whose familiarity with the terms used makes possible a maximum of abbreviations. For other classes of readers, editors may wish to use less contracted combinations made up from this list. For example, the list gives the abbreviation of the term “feet per second” “fps.” To some readers ft/sec will be more easily understood.

^bThis list of abbreviations is adapted from the recommendations of the American National Standards Institute [see ANSI Y1.1-1972 (R1984)].

^cAbbreviation recommended by the American Society of Mechanical Engineers (ASME) Power Test Codes Committee. B = 1 Btu, kB = 1000 Btu, mB = 1,000,000 Btu. The American Society of Heating, Refrigerating and Air-Conditioning Engineers (ASHRAE) recommends the use of Mb = 1000 Btu and Mbh = 1000 Btu/hr.

^dThere are circumstances under which one or the other of these forms is preferred. In general the sign° is used where space conditions make it necessary, as in tabular matter, and when abbreviations are cumbersome, as in some angular measurements, i.e., 59°23' 42". In the interest of simplicity and clarity the Committee has recommended that the abbreviation for the temperature scale, °F, °C, K, etc., always be included in expressions for numerical temperatures, but, wherever feasible, the abbreviation for “degree” be omitted; as 69 F.

TABLE 51.4 Symbols for Physical Quantities^{a,b}

Name of Quantity	Symbol	Name of Quantity	Symbol
Absorption factor	α	Dihedral	Γ
Acceleration		Helical angle of advance	ϕ
Angular	α	Of attack	α
Linear, general	a	Of downwash	ε
Acceleration due to gravity		Of radiation	θ
General	g	Of sideslip	β
International adopted standard	g_0	Of sidewash	σ
Local	g_L	Solid	ω
Gravitational conversion factor	g_c	Angular	
Activity	a	Acceleration	α
Activity coefficient, molal basis	γ	Displacements	δ
Adiabatic factor	X	Frequency	ω
Admittance	Y	Momentum	H
Advanced ratio of propeller	J	Velocity	ω
Altitude	h, z	Area	A
Amplitude	A	Area ^c	S
Angle	α	Aspect ratio	A, AR
Angle	$\beta \phi$	Atomic weight	A
Blade angle	β	Attack, angle of	α
Effective helix	ψ	Attenuation	a

(continued)

TABLE 51.4 (Continued)

Name of Quantity	Symbol	Name of Quantity	Symbol
Axes		Concentration factor, stress	K
Of aircraft (left handed)		Conductance	
Earth-bound coordinate system	x, y, z	Electrical	G
Lateral	Y	Thermal	$1/R$
Longitudinal	X	Per unit area	$1/RA$
Normal	Z	Conductivity	
Bazin's coefficient of roughness	m	Electrical	γ, σ
Blade width (propellers)	b	Equivalent	Λ
Boundary layer thickness	δ	Thermal	k
Breadth	b	Contraction, coefficient of	C_c
Capacitance, capacity	C	Correlation coefficient	R
Capacitivity	ε	Coupling coefficient	k
Of evacuated space	ε_v	Critical state or indicating critical value (subscript)	c
Relative	ε_r	Current ^d	I
Charge, electric or quantity of electricity	Q	Damping	
Charge density		Coefficient	c
Line density of charge	λ	Constant or coefficient	δ
Surface density of charge	σ	Factor	λ
Volume density of charge	ρ	Deflection	δ
Chézy's coefficient	C	Of beam, maximum	δ
Chord length	c	Density	ρ
Circular frequency ($2\pi f$)	ω	Relative to standard air density	σ
Circulation, strength of single vortex	Γ	Depth	h
Coefficient		Depth	y
Absolute	C	Of flow, channels	y
General	C	Diameter	D
Of contraction	C_c	Dielectric constant	ε
Of discharge	C	Difference between values	Δ
Of discharge	C_q	Difference of potential ^{d,e}	E, e
Of energy per unit weight in $C_e(V^2/2g)$	C_e	Diffusion coefficient	D_v
Of flow (Chézy)	C	Diffusivity	α
Of friction (Weisbach–Darcy)	f	Diffusivity, thermal	α
Of friction	μ, f	Of vapor	D_v
Of momentum per unit weight in $C_m(V/g)$	C_m	Discharge	
Of roughness (Bazin)	m	Coefficient of	C_q
Of roughness (Kutter and Manning)	n	Coefficient of	C
Of heat transfer overall	V	Rate of; or flow	Q
Of velocity	C_v	Per unit width	q
Compressibility factor	z	Displacement, electric	D
Concentrated load	F, P, Q	Distance	
Concentration	C, c	From center of gravity to center of pressure of horizontal tail surface	f
Concentration, volumetric	c	Linear	s
		Drag, absolute coefficient of	D
		Dynamic (or impact) pressure	q

TABLE 51.4 (Continued)

Name of Quantity	Symbol	Name of Quantity	Symbol
Eccentricity of application of load	e	Force	F
Efficiency	η	Electromotive ^e	E, e
Elastance	S	Magnetomotive	M, \mathcal{F}
Mutual	S_m, S_{rc}	Moment of	M
Self	S, S_{cc}	Normal	N
Elasticity		Shearing force in beam section	V
Bulk modulus, of liquids	K	Total load	F
Kinematic K/ρ	e	Forces or loads, concentrated	P, Q, F
Modulus of	E	Fraction	
Elastivity	σ	By volume	x_v
Electric potential ^{d,e}	E, e	By weight	x_w
Electricity, quantity of	Q, q	Free energy	
Electromotive force ^d	E, e	Gibbs	G
Electronic charge, absolute value	e	Helmholtz	A
Electrostatic flux	ψ	Frequency	f
Elevation		Circular ($2\pi f$)	θ
Above datum	Z	Of radiant energy	ν
Above stream bed	Z_0	Reduced (flutter)	k
Elongation, total	δ	Rotational	n
Emissivity, total	ε	Frequency, angular	ω
Energy	W	Friction	
Work total	E	Coefficient of sliding	f, μ
Energy	E	Factor used in expressing	f
Internal; intrinsic ^f	U, u	pipeloss	
Kinetic	E_k, T	In energy balance	F
Per unit time (power)	P	Fugacity	f
Potential	E_p, V	Gas constant	R
Enthalpy ^f	H or h	Gibbs function, total potential	G, g
Enthalpy	H	function	
Of dry saturated vapor	h_g	Gyration, radius of	k
Of saturated liquid	h_f	Head	
Per unit weight	h	Atmospheric	h_a
Entropy	S, s	Lost ^c	h
Error signal	ε	Potential	h_{pz}
Expansion, exponent of polytropic	n	Pressure	h_p
Cubical, thermal coefficient	β	Velocity	h_v
Linear, thermal coefficient	α	Heat	
Factor of safety	N	Content; enthalpy ^f	H, h
Film thickness, effective	B	Content of dry saturated	h_g
Flow rate	w	vapor; enthalpy of dry	
In pounds per unit of time	w	saturated vapor	
Volumetric	q	Content of saturated liquid;	h_f
Fluidity	$1/\mu$	enthalpy of saturated liquid	
Flux		Equivalent of work	$1/J$
Density, magnetic	B	Flow rate	q
Displacement	ψ	Across a boundary surface	h
Magnetic	Φ	Latent, of evaporation	λ, h_{fg}
Force	F		

(continued)

TABLE 51.4 (Continued)

Name of Quantity	Symbol	Name of Quantity	Symbol
Mechanical equivalent of	J	Inertia, moment of	
Of vaporization at constant pressure ^f	H_{fg} , λ , or h_{fg}	Polar	J
Specific, at constant pressure	c_p	Rectangular	I
Specific, at constant volume	c_v	Product moment of	I_{xy}
Ratio of specific heats	γ , κ , or k	Intensity	
Transfer, overall coefficient of	U	Electric	E, K
Transfer, surface coefficient of	h	Magnetic	H
Height	h	Isentropic factor	X
Crest, weirs	z	Joule–Thomson coefficient	μ
Helix, effective angle	Φ	Kutter coefficient of roughness	n
Helmholtz free energy; internal potential function ^f	A , a	Length	L
Humidity	H	Length	l
Density of water vapor; weight of water vapor per unit of volume of space	ρH	Lift	L
Density of water vapor at saturation	ρ_s	Linear expansion, coefficient	α
Enthalpy of the mixture minus the enthalpy of the liquid at the temperature of adiabatic saturation; carrier sigma function	h_Σ	Linear velocity	v
Humid volume, volume of mixture per unit of weight of dry air	v_H	Load	
Partial pressure of water vapor	p_H	Concentrated	F , P , Q
Percentage humidity by weight	w_H/w_s	Eccentricity of application of	e
Relative humidity; ratio of an actual partial pressure of water vapor in air to the saturation partial pressure	H_R	Factor	n
Saturation pressure of water vapor	p_s	Per unit distance	w , q
Saturation weight of water vapor per unit of weight of air	H_s , w_s	Total	W , P
Weight of water vapor per unit of weight of dry air	H , w_H	Mach	
Hydraulic radius	R_H	Angle	μ
Mean in a reach	R_m	Number	M
Of cross-sectional area	R	Magnetic	
Hydraulic slope	S_w	Flux	Φ
Impedance	Z	Intensity	H
Impulse	I	Magnetomotive force	M , \mathcal{F}
Inductance	L	Mass	m
Magnetic	B	Flow rate	w
Mutual	L_m	Velocity	G
Self	L , L_{cc}	Mean free path	λ
Inertia, moment of	I	Mechanical equivalent of heat	J
		Microscale (turbulence)	λ .
		Modulus	
		Bulk, of elasticity of liquids	K
		Of elasticity	E
		Of elasticity in shear	G
		Section	Z
		Shear	G
		Molecular weight	M
		Moment	
		Electric	p
		Magnetic	m
		Of any area about a given axis, statical	Q

TABLE 51.4 (Continued)

Name of Quantity	Symbol	Name of Quantity	Symbol
Of force, including bending moment	M	Pressure	
Of inertia, polar	J	Dynamic	q
Rectangular	I	Intensity; force per unit area	p
Mutual inductance	L_m	Relative	δ
Neutral axis, distance to extreme fiber	c	Saturation of water vapor	p_s
Nozzle divergence factor	λ	Propagation constant	γ
Number in general	N	Poynting vector	Π
Of conductors or turns	N	Q factor of a reactor	Q
Of moles, pound-moles, kilogram-moles, etc.	n	Quality of vapor	x
Of phases	m	Quantity	
Of poles	p	Of electricity	Q
Of revolutions per unit of time	n	Of heat per unit mass or unit weight	q
Perimeter, wetted, of a sectional area	P	Of heat per unit time	q
Period	T	Of matter	W
Permeability		Total, of a fluid, water, gas, heat (by volume)	Q
Magnetic	μ	Radiant density	u
Of evacuated space	μ_c	Radiant energy	U
Relative	μ_r	Radiant flux	Φ
Permeance	\mathcal{F}, Λ	Density	W
Permittivity	ϵ	Radiant intensity	J
Phase		Radiation, intensity of	N
Angle	ϕ	Radii	r, R
Constant	β	Radius	r
Displacement	ϕ	Of gyration	k
Pitch, geometric	p	Range	R
Planck constant	h	Reactance	X
Poisson ratio	μ, ν	Capacitive	X_c
Polarization, magnetic	B_i	Inductive	X_L
Pole strength	m	Mutual	X_m, X_{rc}
Potential		Self	X, X_{cc}
Electric ^{d,e}	V	Reactive factor	F_q
Function	ϕ	Recovery factor	η_r
Function, internal; Helmholtz free energy	A, a	Reduced frequency (flutter)	k
Function, total; Gibbs function	G, g	Reflection factor	ρ
Magnetic	\mathbf{M}, \mathcal{F}	Reluctance	\mathcal{R}
Magnetic vector	\mathbf{A}	Reluctivity	ν
Retarded vector	\mathbf{A}_r	Resistance	
Power		Electrical	R
Active	P	Temperature coefficient	α
Apparent	S	Thermal	R
Factor	F_p	Per unit area	RA
Reactive	Q	Resistivity	
		Electrical	ρ
		Thermal	$1/k$
		Revolutions per unit time	n

(continued)

TABLE 51.4 (Continued)

Name of Quantity	Symbol	Name of Quantity	Symbol
Reynolds number	R	Surface coefficient of heat transfer	h
Richness; equivalence ratio (combustion)	R	Surface per unit volume	a
Rotation		Surface tension	σ
Rate of	n	Kinematic σ/ρ	ω
Speed of	n	Susceptance	B
Safety factor	N	Susceptibility	
Saturation pressure of water vapor	p_s	Dielectric	η
Section modulus	Z	Magnetic	κ
Self-inductance	L, L_{cc}	Sweepback angle	Λ
Set of control surfaces, angle of ^c	δ	Taper ratio	λ
Shape factor	S	Temperature	
Shearing force in beam section	V	Absolute ^g (°F abs or K)	T or Θ
Slip	s	Ordinary ^g (°F or °C)	t or θ
Slope		Ratio	θ
Of channel bed	S_0	Thermal	
Of cuts and embankments	s	Conductance	$1/R$
Of energy grade line	S	Per unit area; "unit conductance"	$1/RA$
Of hydraulic grade line	S_w	Conductivity	k
Of lift curve	a	Diffusivity	α
Solidity, propellers	σ	Resistance	R
Span	b	Resistance of unit area	RA
Effectiveness	e	Resistivity	$1/k$
Specific		Transfer factor	j
Gravity	G	Transmission	q
Heat	c	Thickness	$d, t, \text{ or } h$
Heat at constant pressure	c_p	Thrust	
Molar	C_p	Stream	F
Heat at constant volume	c_v	Propeller	T
Molar	C_v	Time	t
Heats, ratio	γ or k or k	Time ^h	t or τ
Volume	v	Time constant	τ
Weight	γ	Torque	Q
Speed		Torque	T or M
Linear	V, v, u	Transmission	
Of rotation	n	Factor	τ
Spring constant	k	Thermal	q
Stefan–Boltzmann constant	σ	Turbulence exchange, coefficient	ε
Strain		Turbulence scale	L
Normal	ε	Vaporization, heat of, at constant pressure	H_{fg}, h_{fg}, λ
Shear	γ	Velocity	V
Stream function	ψ	Velocity	V or v
Stress		Acoustic	V_a
Concentration factor	K	Angular	ω
Normal	σ, s	Average	V
Shear	τ, s_s	Belanger critical	V_c
Supercompressibility factor	z	Components in x, y, z directions, respectively	u, v, w

TABLE 51.4 (Continued)

Name of Quantity	Symbol	Name of Quantity	Symbol
Linear	v	Weight	
Local	u	Molecular	M
Mass, mass flow, per unit cross-sectional area, per unit time	G	Per unit time per unit area of cross section; "mass velocity"	G
Mean (Q/A)	V	Per unit volume	γ
Of light	c	Rate; per unit of power; for unit of time	w
Of sound	a, c	Rate of flow per unit of breadth	Γ
Of uniform flow	V_0	Specific, with g_c	γ
Of wave celerity	c	Total	W
Relative	v	Weirs	
Temporal means of components	$\bar{u}, \bar{v}, \bar{w}$	Crest height	z
Vibration constant	p	Crest length	b
Viscosity		Degree of submergence	N
Absolute; coefficient of	μ	Wetted perimeter	L_p
Kinematic	ν	Width (same as breadth)	b
Relative (to absolute viscosity of water)	μ/μ_w	Of stream bed	b
Relative kinematic	use ν/ν_w	Width, channel surface	b_w
Voltage ^d	E, e	Wing setting, angle of (angle between the wing chord and the thrust line)	i_w
Volume	V	Work	W
Molar	V, V_m	External	W_e
Specific	v	Heat equivalent of	$1/J$ or A
Total	V, V_L	Per unit weight	w, w_k
Volume rate; discharge by volume, fluid rate of flow by volume	q, Q		
Wavelength	λ		
Constant	β		

^aThe most frequently used American Standard and Tentative Standard Symbols are included in this table. Sources used are publications of the American National Standards Institute shown in the bibliography below.

^bWhere possible, capital letters denote total quantities and small letters denote specific quantities, or quantities per unit.

^cUse with appropriate subscript.

^dWhere distinctions between maximum, instantaneous, effective (root-mean-square), and average values are necessary, E_m, I_m, P_m are recommended for maximum values; e, i, p for instantaneous values; E, I for effective (rms) values; and P for average value.

^eWhere a distinction between electromotive force and difference of electric potential is desirable, the symbols E, e , and V, v , respectively, may be used.

^fIn each instance uppercase italics may be used optionally for values in general or per mole. Molal values may have subscript M . Lowercase italics are to be used for specific values (per pound, gram, liter, etc.). Molecular values may be represented by lowercase italics or by lowercase italics with subscript m .

^g θ is preferable only when t is used for time in the same discussion. Θ is preferable only when θ is used for ordinary temperature.

^h τ should be used only when t is used for ordinary temperature in the same discussion.

BIBLIOGRAPHY FOR LETTER SYMBOLS

Acoustics, Letter Symbols and Abbreviations for Quantities Used in, ANSI/ASME Y10.11-1984.

Aeronautical Sciences, Letter Symbols for, ANSI Y10.7-1954.

Chemical Engineering, Letter Symbols for, ANSI Y10.12-1955(R1973).

Glossary of Terms Concerning Letter Symbols, ANSI Y10.1-1972.
Heat and Thermodynamics, Letter Symbols for, ANSI Y10.4-1982.
Hydraulics, Letter Symbols for, ANSI Y10.2-1958.
Illuminating Engineering, Letter Symbols for, ANSI Y10.18-1967(R1977).
Letter Symbols for SI Units and Certain Other Units of Measurement, ANSI/IEEE 260-1978.
Mathematic signs and Symbols for Use in Physical Sciences and Technology (includes supplement ANSI Y10.20a-1975), ANSI Y10.20-1975.
Mechanics and Time-Related Phenomena, ANSI/ASME Y10.3M-1984.
Meteorology, Letter Symbols for, ANSI Y10.10-1953(R1973).
Quantities Used in Electrical Science and Electrical Engineering, Letter Symbols for, ANSI/IEEE 280-1985.
Selecting Greek Letters Used as Letter Symbols for Engineering Mathematics, Guide for, ANSI Y10.17-1961 (R1973).

TABLE 51.5 Graphic Symbols (after Dreyfus, 1972)

Symbols are a graphical referrent to information and have been used for millennia as devices for convenient shorthand notation, to restrict interpretation only to *cognoscenti*, or for compression of data. Examples from several engineering fields are shown here. ISO recommendations are indicated by ¶ and ISO draft recommendations by ¶¶.

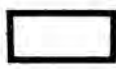













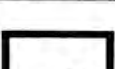

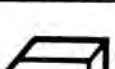
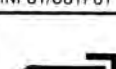


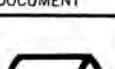

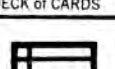
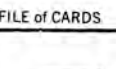
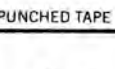

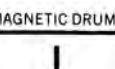
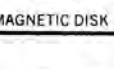
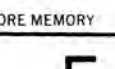
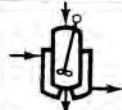









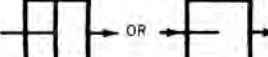
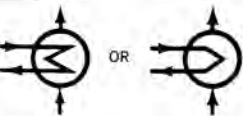


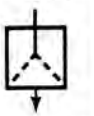




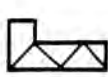
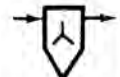
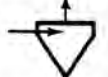
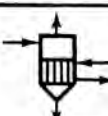
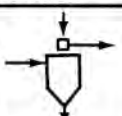
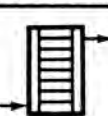
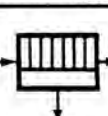
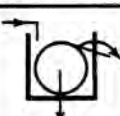
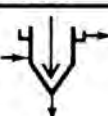
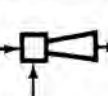






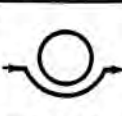


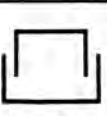
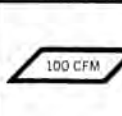



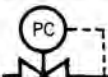
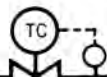

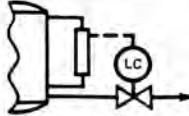
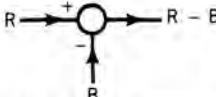

 PROCESS	 DECISION	 PREPARATION	 PREDEFINED PROCESS	 MANUAL OPERATION	
 AUXILIARY OPERATION	 MERGE	 EXTRACT	 COLLATE	 SORT	 MANUAL INPUT
 INPUT/OUTPUT	 ONLINE STORAGE	 OFFLINE STORAGE	 DOCUMENT	 PUNCHED CARD	 DECK of CARDS
 FILE of CARDS	 PUNCHED TAPE	 MAGNETIC TAPE	 MAGNETIC DRUM	 MAGNETIC DISK	 CORE MEMORY
 DISPLAY	 COMMUNICATION LINK	 ENTRANCE CONNECTOR	 EXIT CONNECTOR	 TERMINAL; INTERRUPTION	 COMMENT; ANNOTATION

TABLE 51.5 (Continued)

Chemical Engineering

 JACKETED REACTOR, Stirred	 NUCLEAR REACTOR	 PACKED COLUMN	 PLATE COLUMN	 SECTIONED COLUMN	 DISK and DONUT COLUMN
 FIXED BED REACTOR	 FLUIDIZED BED REACTOR	 AUTOCLAVE	 CENTRIFUGAL PUMP	 RECIPROCATING PUMP	
 REBOILER		 HEAT EXCHANGER	 WATER COOLER	 COOLING TOWER	 SPRAY DRYER
 BLOWER; FAN	 BELT CONVEYOR; SHAKER	 BUCKET CONVEYOR	 SCREW FEEDER	 CENTRIFUGE	 CYCLONE SEPARATOR
 SINGLE-EFFECT EVAPORATOR	 BAROMETRIC CONDENSER	 ELECTRICAL PRECIPITATOR	 PLATE and FRAME FILTER	 ROTARY VACUUM FILTER	 THICKENER
 JET MIXER; EJECTOR	 MIXER	 SCREENERS	 BALL MILL	 ROLLER CRUSHER	 JACKETED VESSEL
 ROTARY DRUM DRYER; KILN	 ROTARY FILM DRYER; FLAKER	 PRESSURE STORAGE TANK	 BULK STORAGE TANK	 GAS HOLDER STORAGE TANK	 GAS FLOW
 TEMPERATURE	 PRESSURE	 ALL CONTROL VALVES	 PRESSURE CONTROLLER	 TEMPERATURE CONTROLLER	 FLOW CONTROLLER
 LEVEL CONTROLLER		 SUMMATION POINT		 OPERATIONAL BLOCK	

(continued)

TABLE 51.5 (Continued)

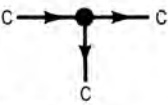



















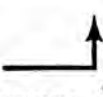
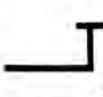



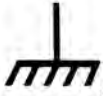





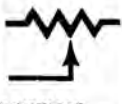







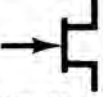

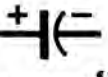

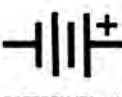

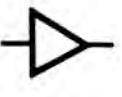





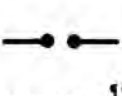
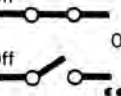





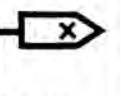

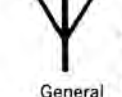
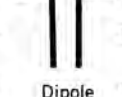


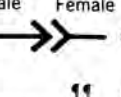





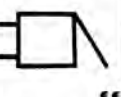
Chemical Engineering (continued)					
					
TAKE-OFF POINT					
Electrical Engineering					
					
DIRECT CURRENT (DC)	ALTERNATING CURRENT (AC)	AUDIO FREQUENCY AC	SUPRAAUDIO FREQUENCY AC	CROSSED CONDUCTORS	JOINED CONDUCTORS
					
SINGLE-PHASE	2-PHASE 3-WIRE	2-PHASE 4-WIRE	3-PHASE 3-WIRE (Delta)	3-PHASE 3-WIRE (Star)	3-PHASE 4-WIRE (Star)
					
2 and 3-PHASE TEE CONNECTED	3-PHASE, 3-WIRE VEE CONNECTED	6-PHASE; FORK with NEUTRAL	START of WINDING	VARIABLE CONTROL	VARIABLE CONTROL by STEPS
					
PRESET CONTROL	ADJUSTABLE TAPPING	PRESET TAPPING	NON-LINEAR VARIABILITY	SATURABLE PROPERTIES	EARTH (Ground)
					
CHASSIS of EQUIPMENT	INSULATED COUPLING	UNINSULATED COUPLING	SCREENED CONDUCTOR	RESISTOR	
					
NON-INDUCTIVE RESISTOR (Heater)	ADJUSTABLE CONTACT RESISTOR	INDUCTOR	TRANSFORMER	VACUUM TUBE (Triode)	DIODE
					
CONTROLLED RECTIFIER	TRANSISTOR (n---n type)	TRANSISTOR (p---p type)	TRANSISTOR, Field-effect (n-channel)	FIXED CAPACITOR	ELECTROLYTIC CAPACITOR











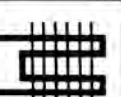
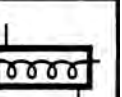
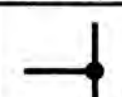
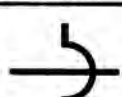
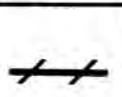
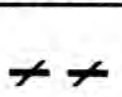
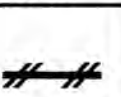
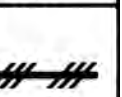
TABLE 51.5 (Continued)

Electrical Engineering (continued)

					
ALTERNATING CURRENT SOURCE	BATTERY (Direct Current Source)	PIEZOELECTRIC CRYSTAL UNIT	AMPLIFIER	LOUDSPEAKER	MICROPHONE
					
CATHODE RAY TUBE (TV)	LAMP BULB	INDICATOR	LIGHTNING ARRESTER	SWITCH	
					
FUSE	CIRCUIT BREAKER	RECORDING HEAD	PLAYBACK HEAD	ERASE HEAD	EQUIPMENT OUTLINE
					
General AERIAL (Antenna)	Dipole	Loop	Loop	CONNECTOR	
					
PERMANENT MAGNET	PHOTO- SENSITIVITY	BELL	BUZZER		

¶ Draft ISO Recommendation

Mechanical Engineering

					
COMPRESSOR	PNEUMATIC COMPRESSOR	HYDRAULIC MOTOR	OSCILLATING MOTOR	HYDRAULIC PUMP	PUMP, ROTARY and CENTRIFUGAL
					
ENGINE, Gas	BLOWER, Gas	TURBINE	HEAT EXCHANGER	AIR COOLED CONDENSER	WATER COOLED CONDENSER
					
PIPE LINE JUNCTION	CROSSED PIPE LINES	LOW PRESSURE STEAM SUPPLY	LOW PRESSURE STEAM RETURN	MEDIUM PRESSURE STEAM SUPPLY	HIGH PRESSURE STEAM SUPPLY

(continued)

TABLE 51.5 (Continued)

Mechanical Engineering (continued)

PNEUMATIC FLOW DIRECTION	HYDRAULIC FLOW DIRECTION	WASTE WATER	COLD WATER	HOT WATER SUPPLY	HOT WATER RETURN
VENT PIPE	CHILLED WATER LINE	FUEL LINE	GAS LINE	VACUUM LINE	THREADED PIPE JOINT
FLANGED PIPE JOINT	WELDED PIPE JOINT	BELL and SPIGOT PIPE JOINT	SOLDERED PIPE JOINT	UNION, Threaded	TEE JOINT, Threaded
CROSS JOINT, Threaded	90° ELBOW, Threaded	LATERAL JOINT, Threaded	EOCENTRIC REDUCER	CONCENTRIC REDUCER	THREADED BUSHING
EXPANSION JOINT FLANGE	CHECK VALVE	SHUT-OFF VALVE; GATE VALVE	GLOBE VALVE	COCK VALVE	DIAPHRAGM VALVE
SAFETY VALVE	STOP COCK	PRESSURE GAUGE	THERMOMETER	Welding	FILLET
PLUG; SLOT	ARC-SPOT; ARC SEAM	BACKING; BACK	MELT-THROUGH	EDGE FLANGE	CORNER FLANGE
SURFACING	SQUARE GROOVE	"V" GROOVE	"U" GROOVE	"J" GROOVE	FLARE "V" GROOVE
FLARE BEVEL GROOVE	BEVEL GROOVE	WELD ALL AROUND	FIELD WELD	FLUSH CONTOUR	CONVEX CONTOUR

TABLE 51.5 (Continued)

Mechanical Engineering (continued)

Geometric Tolerances	STRAIGHTNESS	FLATNESS	FLATNESS and STRAIGHTNESS	CIRCULARITY (Roundness)	CYLINDRICITY
PROFILE of any LINE	PROFILE of any SURFACE	PARALLELISM	SQUARENESS (Perpendicularity)	ANGULARITY	POSITION
COAXIALITY; CONCENTRICITY	SYMMETRY	RUN-OUT	SURFACE ROUGHNESS	SURFACE to be FINISHED (Machined)	

- ▲ Pneumatic machinery is indicated by \triangle , hydraulic machinery by \blacktriangle
- G indicates Gas. Different initial may be substituted to indicate other type of machine; e.g., D (diesel), M (motor), T (turbine), E (steam). Exception: (see Engine, Gas) Steam Engine is indicated by symbol without initial.
- ★ C indicates Circulating Water. Different initial indicates other type of machine or service e.g., D (concentrate), F (boiler feed), O (oil), S (service), V (air).
- "Return" indicated by broken line, as illustrated in Low Pressure Steam Return.
- ¶ Draft ISO Recommendation
- ▼ Flanged, Welded, Bell and Spigot, or Soldered Union indicated by substituting appropriate markings (see Joints). Example; $\times \text{---} \times$ Welded Union.
- ¶ ISO Recommendation

BIBLIOGRAPHY FOR GRAPHIC SYMBOLS

- Arnell, A. *Standard Graphical Symbols—A Comprehensive Guide for Use in Industry, Engineering and Science*, McGraw-Hill, New York, 1963.
- Dreyfus, H. *Symbol Sourcebook*, McGraw-Hill, New York, 1972.
- Electrical and Electronics Diagrams (Including Reference Designation Class Designation Letters), Graphic Symbols for, ANSI/IEEE 315-1975.
- Electrical and Electronics Diagrams, Graphic Symbols for (supplement to ANSI/IEEE 315-1975), ANSI/IEEE 315A-1986.
- Electrical and Electronics Parts and Equipments, Reference Designations for, ANSI/IEEE 200-1975.
- Electrical Wiring and Layout Diagrams Used in Architecture and Building Construction, Graphic Symbols for, ANSI Y32.9-1972.
- Fire Fighting Operations, Symbols for, ANSI/NFPA 178-1980.
- Fire Protection Symbols for risk Analysis Diagrams, ANSI/NFPA 174-1980.
- Fire-Protection Symbols for Architectural and Engineering Drawings, ANSI/NFPA 172-1980.
- Fluid Power Diagrams, Graphic Symbols for, ANSI Y32.10-1967(R1979).
- Grid and Mapping Used in Cable Television Systems, Graphic Symbols for, ANSI/IEEE 623-1976.
- Heat-Power Apparatus, Graphic Symbols for, ANSI Y32.2.6M-1984.
- Heating, Ventilating, and Air Conditioning, Graphic Symbols for, ANSI Y32.2.4M-1984.
- Polon, D. D. (Ed.), *Encyclopedia of Engineering Signs and Symbols*, Odyssey Press, New York, 1965.
- Shepard, W., *Shepard's Glossary of Graphic Signs and Symbols*, Dent, London, 1971.

TABLE 51.6 Personal Computer Numeric Codes for Characters and Symbols

IBM PC Character Set (00-7F) Quick Reference																IBM PC Character Set (80-FF) Quick Reference															
DECIMAL VALUE	0	16	32	48	64	80	96	112	DECIMAL VALUE	128	144	160	176	192	208	224	240	DECIMAL VALUE	128	144	160	176	192	208	224	240					
HEX DECIMAL VALUE	0	1	2	3	4	5	6	7	HEX DECIMAL VALUE	8	9	A	B	C	D	E	F	HEX DECIMAL VALUE	8	9	A	B	C	D	E	F					
0	0	BLANK (NULL)	BLANK (SPACE)	O	@	P	'	p	C	Ç	É	á	á	á	á	á	á	80	81	82	83	84	85	86	87	88					
1	1	☺	☹	1	A	Q	a	q	ü	ü	æ	í	í	í	í	í	í	89	8A	8B	8C	8D	8E	8F	90	91					
2	2	☺	"	2	B	R	b	r	é	é	Æ	ó	ó	ó	ó	ó	ó	92	93	94	95	96	97	98	99	9A					
3	3	♥	#	3	C	S	c	s	â	â	ô	ú	ú	ú	ú	ú	ú	9B	9C	9D	9E	9F	90	91	92	93					
4	4	♦	\$	4	D	T	d	t	ä	ä	ö	ñ	ñ	ñ	ñ	ñ	ñ	94	95	96	97	98	99	9A	9B	9C					
5	5	♣	%	5	E	U	e	u	à	à	ù	ä	ä	ä	ä	ä	ä	9D	9E	9F	90	91	92	93	94	95					
6	6	♠	&	6	F	V	f	v	ç	ç	û	ç	ç	ç	ç	ç	ç	9E	9F	90	91	92	93	94	95	96					
7	7	•	'	7	G	W	g	w	ê	ê	ÿ	ö	ö	ö	ö	ö	ö	9F	90	91	92	93	94	95	96	97					
8	8	•	(8	H	X	h	x	ë	ë	ÿ	ö	ö	ö	ö	ö	ö	90	91	92	93	94	95	96	97	98					
9	9	○)	9	I	Y	i	y	ë	ë	ÿ	ö	ö	ö	ö	ö	ö	91	92	93	94	95	96	97	98	99					
10	A	○	*	:	J	Z	j	z	è	è	ÿ	ö	ö	ö	ö	ö	ö	92	93	94	95	96	97	98	99	9A					
11	B	♂	+	:	K	[k	{	ï	ï	ÿ	ö	ö	ö	ö	ö	ö	93	94	95	96	97	98	99	9A	9B					
12	C	♀	,	<	L	\	l		ï	ï	ÿ	ö	ö	ö	ö	ö	ö	94	95	96	97	98	99	9A	9B	9C					
13	D	♂	—	=	M] m	m	}	ï	ï	ÿ	ö	ö	ö	ö	ö	ö	95	96	97	98	99	9A	9B	9C	9D					
14	E	♂	•	>	N	^ n	n	~	ï	ï	ÿ	ö	ö	ö	ö	ö	ö	96	97	98	99	9A	9B	9C	9D	9E					
15	F	☼	/	?	O	_ o	o	Δ	Ä	Ä	ß	«	»	«	»	«	»	97	98	99	9A	9B	9C	9D	9E	9F					
																		98	99	9A	9B	9C	9D	9E	9F	BLANK FF					

Source: Reprint courtesy of International Business Machines Corporation, copyright 1985 © by International Business Machines Corporation.

TABLE 51.7 Conversions for Number Systems of Different Bases

Radix 16 Hexadecimal	Radix 10 Decimal	Radix 8 Octal	Radix 2 Binary		Radix 16 Hexadecimal	Radix 10 Decimal	Radix 8 Octal	Radix 2 Binary	
			BIT					BIT	
			8765	4321				8765	4321
00	0	00	0000	0000	32	50	62	0011	0010
01	1	01	0000	0001	33	51	63	0011	0011
02	2	02	0000	0010	34	52	64	0011	0100
03	3	03	0000	0011	35	53	65	0011	0101
04	4	04	0000	0100	36	54	66	0011	0110
05	5	05	0000	0101	37	55	67	0011	0111
06	6	06	0000	0110	38	56	70	0011	1000
07	7	07	0000	0111	39	57	71	0011	1001
08	8	10	0000	1000	3A	58	72	0011	1010
09	9	11	0000	1001	3B	59	73	0011	1011
0A	10	12	0000	1010	3C	60	74	0011	1100
0B	11	13	0000	1011	3D	61	75	0011	1101
0C	12	14	0000	1100	3E	62	76	0011	1110
0D	13	15	0000	1101	3F	63	77	0011	1111
0E	14	16	0000	1110	40	64	100	0100	0000
0F	15	17	0000	1111	41	65	101	0100	0001
10	16	20	0001	0000	42	66	102	0100	0010
11	17	21	0001	0001	43	67	103	0100	0011
12	18	22	0001	0010	44	68	104	0100	0100
13	19	23	0001	0011	45	69	105	0100	0101
14	20	24	0001	0100	46	70	106	0100	0110
15	21	25	0001	0101	47	71	107	0100	0111
16	22	26	0001	0110	48	72	110	0100	1000
17	23	27	0001	0111	49	73	111	0100	1001
18	24	30	0001	1000	4A	74	112	0100	1010
19	25	31	0001	1001	4B	75	113	0100	1011
1A	26	32	0001	1010	4C	76	114	0100	1100
1B	27	33	0001	1011	4D	77	115	0100	1101
1C	28	34	0001	1100	4E	78	116	0100	1110
1D	29	35	0001	1101	4F	79	117	0100	1111
1E	30	36	0001	1110	50	80	120	0101	0000
1F	31	37	0001	1111	51	81	121	0101	0001
20	32	40	0010	0000	52	82	122	0101	0010
21	33	41	0010	0001	53	83	123	0101	0011
22	34	42	0010	0010	54	84	124	0101	0100
23	35	43	0010	0011	55	85	125	0101	0101
24	36	44	0010	0100	56	86	126	0101	0110
25	37	45	0010	0101	57	87	127	0101	0111
26	38	46	0010	0110	58	88	130	0101	1000
27	39	47	0010	0111	59	89	131	0101	1001
28	40	50	0010	1000	5A	90	132	0101	1010
29	41	51	0010	1001	5B	91	133	0101	1011
2A	42	52	0010	1010	5C	92	134	0101	1100
2B	43	53	0010	1011	5D	93	135	0101	1101
2C	44	54	0010	1100	5E	94	136	0101	1110
2D	45	55	0010	1101	5F	95	137	0101	1111
2E	46	56	0010	1110	60	96	140	0110	0000
2F	47	57	0010	1111	61	97	141	0110	0001
30	48	60	0011	0000	62	98	142	0110	0010
31	49	61	0011	0001	63	99	143	0110	0011

(continued)

TABLE 51.7 (Continued)

Radix 16 Hexadecimal	Radix 10 Decimal	Radix 8 Octal	Radix 2 Binary		Radix 16 Hexadecimal	Radix 10 Decimal	Radix 8 Octal	Radix 2 Binary	
			BIT					BIT	
			8765	4321				8765	4321
64	100	144	0110	0100	96	150	226	1001	0110
65	101	145	0110	0101	97	151	227	1001	0111
66	102	146	0110	0110	98	152	230	1001	1000
67	103	147	0110	0111	99	153	231	1001	1001
68	104	150	0110	1000	9A	154	232	1001	1010
69	105	151	0110	1001	9B	155	233	1001	1011
6A	106	152	0110	1010	9C	156	234	1001	1100
6B	107	153	0110	1011	9D	157	235	1001	1101
6C	108	154	0110	1100	9E	158	236	1001	1110
6D	109	155	0110	1101	9F	159	237	1001	1111
6E	110	156	0110	1110	A0	160	240	1010	0000
6F	111	157	0110	1111	A1	161	241	1010	0001
70	112	160	0111	0000	A2	162	242	1010	0010
71	113	161	0111	0001	A3	163	243	1010	0011
72	114	162	0111	0010	A4	164	244	1010	0100
73	115	163	0111	0011	A5	165	245	1010	0101
74	116	164	0111	0100	A6	166	246	1010	0110
75	117	165	0111	0101	A7	167	247	1010	0111
76	118	166	0111	0110	A8	168	250	1010	1000
77	119	167	0111	0111	A9	169	251	1010	1001
78	120	170	0111	1000	AA	170	252	1010	1010
79	121	171	0111	1001	D6	214	326	1101	0110
7A	122	172	0111	1010	D7	215	327	1101	0111
7B	123	173	0111	1011	D8	216	330	1101	1000
7C	124	174	0111	1100	D9	217	331	1101	1001
7D	125	175	0111	1101	DA	218	332	1101	1010
7E	126	176	0111	1110	DB	219	333	1101	1011
7F	127	177	0111	1111	DC	220	334	1101	1100
80	128	200	0000	0000	DD	221	335	1101	1101
81	129	201	1000	0001	DE	222	336	1101	1110
82	130	202	1000	0010	DF	223	337	1101	1111
83	131	203	1000	0011	E0	224	340	1110	0000
84	132	204	1000	0100	E1	225	341	1110	0001
85	133	205	1000	0101	E2	226	342	1110	0010
86	134	206	1000	0110	E3	227	343	1110	0011
87	135	207	1000	0111	E4	228	344	1110	0100
88	136	210	1000	1000	E5	229	345	1110	0101
89	137	211	1000	1001	E6	230	346	1110	0110
8A	138	212	1000	1010	E7	231	347	1110	0111
8B	139	213	1000	1011	E8	232	350	1110	1000
8C	140	214	1000	1100	E9	233	351	1110	1001
8D	141	215	1000	1101	EA	234	352	1110	1010
8E	142	216	1000	1110	EB	235	353	1110	1011
8F	143	217	1000	1111	EC	236	354	1110	1100
90	144	220	1001	0000	ED	237	355	1110	1101
91	145	221	1001	0001	EE	238	356	1110	1110
92	146	222	1001	0010	EF	239	357	1110	1111
93	147	223	1001	0011	F0	240	360	1111	0000
94	148	224	1001	0100	F1	241	361	1111	0001
95	149	225	1001	0101	F2	242	362	1111	0010

TABLE 51.7 (Continued)

Radix 16 Hexadecimal	Radix 10 Decimal	Radix 8 Octal	Radix 2 Binary		Radix 16 Hexadecimal	Radix 10 Decimal	Radix 8 Octal	Radix 2 Binary	
			BIT					BIT	
			8765	4321				8765	4321
F3	243	363	1111	0011	BA	186	272	1011	1010
F4	244	364	1111	0100	BB	187	273	1011	1011
F5	245	365	1111	0101	BC	188	274	1011	1100
F6	246	366	1111	0110	BD	189	275	1011	1101
F7	247	367	1111	0111	BE	190	276	1011	1110
F8	248	370	1111	1000	BF	191	277	1011	1111
F9	249	371	1111	1001	C0	192	300	1100	0000
FA	250	372	1111	1010	C1	193	301	1100	0001
FB	251	373	1111	1011	C2	194	302	1100	0010
FC	252	374	1111	1100	C3	195	303	1100	0011
FD	253	375	1111	1101	C4	196	304	1100	0100
FE	254	376	1111	1110	C5	197	305	1100	0101
FF	255	377	1111	1111	C6	198	306	1100	0110
AB	171	253	1010	1011	C7	199	307	1100	0111
AC	172	254	1010	1100	C8	200	310	1100	1000
AD	173	255	1010	1101	C9	201	311	1100	1001
AE	174	256	1010	1110	CA	202	312	1100	1010
AF	175	257	1010	1111	CB	203	313	1100	1011
B0	176	260	1011	0000	CC	204	314	1100	1100
B1	177	261	1011	0001	CD	205	315	1100	1101
B2	178	262	1011	0010	CE	206	316	1100	1110
B3	179	263	1011	0011	CF	207	317	1100	1111
B4	180	264	1011	0100	D0	208	320	1101	0000
B5	181	265	1011	0101	D1	209	321	1101	0001
B6	182	286	1011	0110	D2	210	322	1101	0010
B7	183	267	1011	0111	D3	211	323	1101	0011
B8	184	270	1011	1000	D4	212	324	1101	0100
B9	185	271	1011	1001	D5	213	325	1101	0101

TABLE 51.8 Computer Graphics Codes and Standards

Modern computer-aided design (CAD), computer-aided manufacturing (CAM), and computer-aided engineering (CAE) are heavily dependent on computer graphics. A standard computer graphics metafile (CGM) is necessary in order to:

1. Allow picture information to be stored in an organized way on a graphical software system.
2. Facilitate transfer of picture information between different graphical software systems.
3. Enable picture information to be transferred between graphical devices.
4. Enable picture information to be transferred between different computer graphics installations.

More particularly, the CGM should provide these capabilities in a device-independent manner. To accomplish this, the standard defines the form (syntax) and functional behavior (semantics) of a set of elements that may occur in the CGM. There are eight classes of elements:

1. Delimiter Elements—delimit significant structures within the metafile.
2. Metafile Descriptor Elements—describe the functional content, default conditions, identification, and characteristics of the CGM.
3. Picture Descriptor Elements—set the interpretation modes of attribute elements for each picture.

(continued)

TABLE 51.8 (Continued)

4. Control Elements—allow picture boundaries and coordinate representation to be modified.
5. Graphical Primitive Elements—describe the visual components of a picture in the CGM.
6. Attribute Elements—describe the appearance of graphical primitive elements.
7. Escape Elements—describe device- or system-dependent elements used to construct a picture; however, the elements are not otherwise standardized.
8. External Elements—communicate information not directly related to the generation of a graphical image.

A computer graphics metafile is a collection of elements from this standardized set. The BEGIN METAFILE and END METAFILE elements each occur exactly once in a complete metafile; as many or as few of the elements in the other classes may occur as are needed. A metafile needs to be interpreted in order to display its pictorial content on a graphics device. The descriptor elements give the interpreter sufficient data to interpret metafile elements and to make informed decisions concerning the resources needed for display.

A CGM contains delimiter elements; in addition it may include control elements for metafile interpretation, picture descriptor elements for declaring parameter modes of attribute elements, graphical primitive elements for defining graphical entities, attribute elements for defining the appearance of the graphical primitive elements, escape elements for accessing nonstandardized features of particular devices, and external elements for communication of information external to the definition of the pictures in the CGM.

Full description and depiction of all the elements thus far defined in this standardized set is beyond the scope of this handbook. The interested reader is referred to ANSI standard X3.122-1986 and Smith, B. M., et al. "Initial Graphics Exchange Specification (IGES), Version 2.0," NBS (R82-2631) (AF) Feb. (1983) 26 pp.

51.2 MATHEMATICAL TABLES

TABLE 51.9 Certain Constants Containing *e* and π^a

Powers of <i>e</i>			Multiples of π			Fractions of π		
<i>e</i> ^{<i>n</i>}	Value	Logarithm	$n\pi$	Value	Logarithm	π/n	Value	Logarithm
<i>e</i>	2.718282	0.434294	π	3.141593	0.497150	$\pi/2$	1.570780	0.196120
<i>e</i> ^{−1}	0.367879	̄1.565706	2π	6.283185	0.798180	$\pi/3$	1.047198	0.020029
<i>e</i> ²	7.389057	0.868589	3π	9.424778	0.974271	$\pi/4$	0.785398	̄1.895090
<i>e</i> ^{−2}	0.135335	̄1.131411	4π	12.566371	1.099210	$\pi/180$	0.017453 ^b	2.241877
<i>e</i> ^{1/2}	1.648721	0.217147	5π	15.707963	1.196120			
Reciprocals of π			Powers of π			Roots of π		
n/π	Value	Logarithm	$\pi^{\pm n}$	Value	Logarithm	$\pi^{\pm 1/n}$	Value	Logarithm
$1/\pi$	0.318310	̄1.502850	π^2	9.869604	0.994300	$\sqrt{\pi}$	1.772454	0.248575
$2/\pi$	0.636620	̄1.803880	$1/\pi^2$	0.101321	̄1.005700	$1/\sqrt{\pi}$	0.564190	̄1.751425
$3/\pi$	0.954930	̄1.979971	π^3	31.006277	1.491450	$\sqrt[3]{\pi}$	1.464592	0.165717
$180/\pi$	57.295780 ^c	1.758123	$1/\pi^3$	0.032252	̄2.508550	$1/\sqrt[3]{\pi}$	0.682784	̄1.834283

^a*e* = 2.7182818285; π = 3.1415926536; $M = \log_{10}e = 0.4342944819$; $M^{-1} = \log_e 10 = 2.3025850930$.

^bNumber of radians per degree.

^cNumber of degrees per radian.

TABLE 51.10 Factorials

n	$n! = 1 \cdot 2 \cdot 3 \cdots n$	$1/n!$	n	$n! = 1 \cdot 2 \cdot 3 \cdots n$	$1/n!$
1	1	1.	11	$399,168 \times 10^2$	0.250521×10^{-7}
2	2	0.5	12	$479,002 \times 10^3$	0.208768×10^{-8}
3	6	0.166667	13	$622,702 \times 10^4$	0.160590×10^{-9}
4	24	0.416667×10^{-1}	14	$871,783 \times 10^5$	0.114707×10^{-10}
5	120	0.833333×10^{-2}	15	$130,767 \times 10^7$	0.764716×10^{-12}
6	720	0.138889×10^{-2}	16	$209,228 \times 10^8$	0.477948×10^{-13}
7	5,040	0.198413×10^{-3}	17	$355,687 \times 10^9$	0.281146×10^{-14}
8	40,320	0.248016×10^{-4}	18	$640,237 \times 10^{10}$	0.156192×10^{-15}
9	362,880	0.275573×10^{-5}	19	$121,645 \times 10^{12}$	0.822064×10^{-17}
10	3,628,800	0.275573×10^{-6}	20	$243,290 \times 10^{13}$	0.411032×10^{-19}

TABLE 51.11 Common and Natural Logarithms of Numbers

The common logarithm of a number is the index of the power to which the base 10 must be raised in order to equal the number.

The common logarithm of every positive number not an integral power of 10 consists of an *integral* and a *decimal part*. The integral part or whole number is called the *characteristic* and may be either *positive* or *negative*. The decimal or fractional part is a positive number called the *mantissa* and is the same for all numbers which have the same sequential digits.

The characteristic of the logarithm of any positive number greater than 1 is positive and is 1 less than the number of digits before the decimal point.

The characteristic of the logarithm of any positive number less than 1 is negative and is 1 more than the number of ciphers immediately after the decimal point.

A negative number or number less than zero has no real logarithm.

Examples: $\log_{10} 25,400. = 4.404834$, $\log_{10} 0.0254 = \bar{2}.404834$, or $8.404834 - 10$

The two systems of logarithms in general use are the common or Briggsian logarithms, introduced in 1615 by Henry Briggs, a contemporary of John Napier, the inventor of logarithms, and the natural or less appropriately termed Napierian or hyperbolic logarithms, which developed somewhat accidentally from Napier's original work. The latter have a base denoted by e , an irrational number, which is

$$\lim_{u \rightarrow \infty} \left(1 + \frac{1}{u}\right)^u = 1 + 1 + \frac{1}{2!} + \frac{1}{3!} + \frac{1}{4!} + \cdots = 2.7182818$$

To obtain the natural logarithm, the common logarithm is multiplied by $\log_e 10$, which is 2.302585, or $\log_e N = 2.302585 \log_{10} N$.

The natural logarithm of a number is the index of the power to which the base e ($=2.7182818$) must be raised in order to equal the number.

Example: $\log_e 4.12 = \ln 4.12 = 1.4159$.

Natural logarithms of numbers from 1.00 to 9.99 may be obtained directly; the natural logarithms of numbers outside of that range by the addition or subtraction of the natural logarithms of powers of 10.

Examples: $\log_e 679. = \log_e 6.79 + \log_e 10^2 = 1.9155 + 4.6052 = 6.5207$.
 $\log_e 0.0679 = \log_e 6.79 - \log_e 10^2 = 1.9155 - 4.6052 = -2.6897$

(continued)

TABLE 51.11 (Continued)

Natural Logarithms of Powers of 10

$\log_e 10 = 2.302585 \quad \log_e 10^4 = 9.210340 \quad \log_e 10^7 = 16.118096$
 $\log_e 10^2 = 4.605170 \quad \log_e 10^5 = 11.512925 \quad \log_e 10^8 = 18.420681$
 $\log_e 10^3 = 6.907755 \quad \log_e 10^6 = 13.815511 \quad \log_e 10^9 = 20.723266$

To obtain the common logarithm, the natural logarithm is multiplied by $\log_{10} e$, which is 0.434294, or $\log_{10} N = 0.434294 \log_e N$.

A negative number or number less than zero has no real logarithm.

Tabulations of common and natural logarithms are no longer provided in this handbook because of ready access to them on modern pocket and desk calculators.

Values and Logarithms of Exponentials and Hyperbolic Functions

Many calculators directly give values of e^x , e^{-x} , $\sinh x$, $\cosh x$, and $\tanh x$ for any value of x . These quantities are therefore not tabulated here.

For values of x greater than 6, e^x may be computed from the relationship $e^x = \log^{-1}(x \log_{10} e) = \log^{-1} 0.43429x$; e^{-x} approaches zero; $\sinh x$ and $\cosh x$ are approximately equal and become $0.5e^x$; and $\tanh x$ and $\coth x$ have values approximately equal to unity.

Where more accurate values of the exponentials and functions are required they may be computed from the following relationships:

$e = 2.7182818285$	$\frac{1}{e} = 0.3678794412$	
$M = \log_{10} e = 0.4342944819$	$\frac{1}{M} = \log_e 10 = 2.3025850930$	
$e^x = \log^{-1} Mx$	$e^{-x} = \log^{-1} (-Mx)$	
$\sin hx = \frac{e^x - e^{-x}}{2}$	$\cos hx = \frac{e^x + e^{-x}}{2}$	$\tan hx = \frac{e^x - e^{-x}}{e^x + e^{-x}}$
$\operatorname{csch} x = \frac{1}{\sin hx}$	$\sec hx = \frac{1}{\cos hx}$	$\cot hx = \frac{1}{\tan hx}$

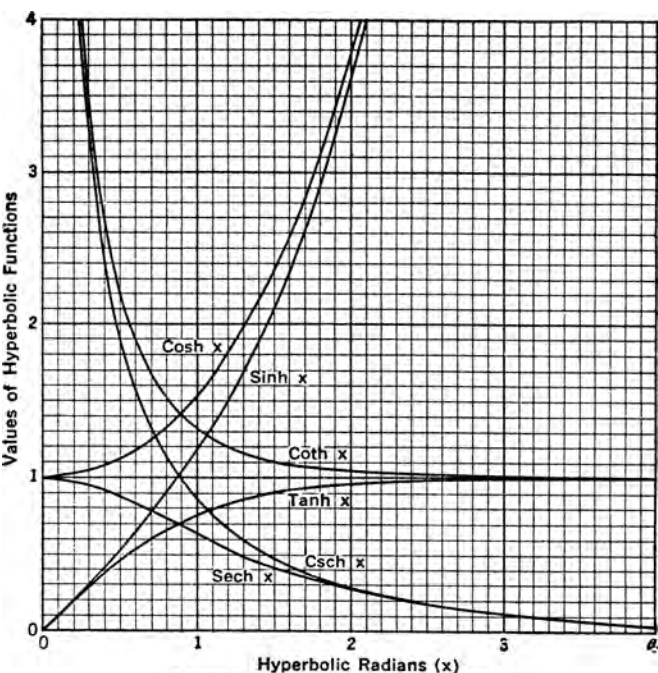


Chart of the Hyperbolic Functions.

Values of the hyperbolic functions are shown in the figure.

TABLE 51.12 Circular Arcs, Chords, and Segments

Central Angle in Degrees	Arc <i>R</i>	Height <i>R</i>	Chord <i>R</i>	Height Chord	Area <i>R</i> ²	Central Angle in Degrees	Arc <i>R</i>	Height <i>R</i>	Chord <i>R</i>	Height Chord	Area <i>R</i> ²
1	0.0175	0.0000	0.0175	0.0022	0.00000	51	0.8901	0.0974	0.8610	0.1131	0.05649
2	0.0349	0.0002	0.0349	0.0044	0.00000	52	0.9076	0.1012	0.8767	0.1154	0.05978
3	0.0524	0.0003	0.0524	0.0066	0.00001	53	0.9250	0.1051	0.8924	0.1177	0.06319
4	0.0698	0.0006	0.0698	0.0087	0.00003	54	0.9425	0.1090	0.9080	0.1200	0.06673
5	0.0873	0.0010	0.0872	0.0109	0.00006	55	0.9599	0.1130	0.9235	0.1223	0.07039
6	0.1047	0.0014	0.1047	0.0131	0.00010	56	0.9774	0.1171	0.9389	0.1247	0.07417
7	0.1222	0.0019	0.1221	0.0153	0.00015	57	0.9948	0.1212	0.9543	0.1270	0.07808
8	0.1396	0.0024	0.1395	0.0175	0.00023	58	1.0123	0.1254	0.9696	0.1293	0.08212
9	0.1571	0.0031	0.1569	0.0196	0.00032	59	1.0297	0.1296	0.9848	0.1316	0.08629
10	0.1745	0.0038	0.1743	0.0218	0.00044	60	1.0472	0.1340	1.0000	0.1340	0.09059
11	0.1920	0.0046	0.1917	0.0240	0.00059	61	1.0647	0.1384	1.015	0.1363	0.09502
12	0.2094	0.0055	0.2091	0.0262	0.00076	62	1.0821	0.1428	1.030	0.1387	0.09958
13	0.2296	0.0064	0.2264	0.0284	0.00097	63	1.0996	0.1474	1.045	0.1410	0.10428
14	0.2443	0.0075	0.2437	0.0306	0.00121	64	1.1170	0.1520	1.060	0.1434	0.10911
15	0.2618	0.0086	0.2611	0.0328	0.00149	65	1.1345	0.1566	1.075	0.1457	0.11408
16	0.2793	0.0097	0.2783	0.0350	0.00181	66	1.1519	0.1613	1.089	0.1481	0.11919
17	0.2967	0.0110	0.2956	0.0372	0.00217	67	1.1694	0.1661	1.104	0.1505	0.12443
18	0.3142	0.0123	0.3129	0.0394	0.00257	68	1.1868	0.1710	1.118	0.1529	0.12982
19	0.3316	0.0137	0.3301	0.0415	0.00302	69	1.2043	0.1759	1.133	0.13535	0.13535
20	0.3491	0.0152	0.3473	0.0437	0.00352	70	1.2217	0.1808	1.147	0.1576	0.14102
21	0.3665	0.0167	0.3645	0.0459	0.00408	71	1.2392	0.1859	1.161	0.1601	0.14683
22	0.3840	0.0184	0.3816	0.0481	0.00468	72	1.2566	0.1910	1.176	0.1625	0.15279
23	0.4014	0.0201	0.3987	0.0503	0.00535	73	1.2741	0.1961	1.190	0.1649	0.15889
24	0.4189	0.0219	0.4158	0.0526	0.00607	74	1.2915	0.2014	1.204	0.1673	0.16514
25	0.4363	0.0237	0.4329	0.0548	0.00686	75	1.3090	0.2066	1.218	0.1697	0.17154
26	0.4538	0.0256	0.4499	0.0570	0.00771	76	1.3265	0.2120	1.231	0.1722	0.17808
27	0.4712	0.0276	0.4669	0.0592	0.00862	77	1.3439	0.2174	1.245	0.1746	0.18477
28	0.4887	0.0297	0.4838	0.0614	0.00961	78	1.3614	0.2229	1.259	0.1771	0.19160
29	0.5061	0.0319	0.5008	0.0636	0.01067	79	1.3788	0.2284	1.272	0.1795	0.19859
30	0.5236	0.0341	0.5176	0.0658	0.1180	80	1.3963	0.2340	1.286	0.1820	0.20573
31	0.5411	0.0364	0.5345	0.0680	0.01301	81	1.4137	0.2396	1.299	0.1845	0.21301
32	0.5585	0.0387	0.5513	0.0703	0.01429	82	1.4312	0.2453	1.312	0.1869	0.22045
33	0.5760	0.0412	0.5680	0.0725	0.01566	83	1.4486	0.2510	1.325	0.1894	0.22804
34	0.5934	0.0437	0.5847	0.747	0.01711	84	1.4661	0.2569	1.338	0.1919	0.23578
35	0.6109	0.0463	0.6014	0.0770	0.01864	85	1.4835	0.2627	1.351	0.1944	0.24367
36	0.6283	0.0489	0.6180	0.0792	0.02027	86	1.5010	0.2686	1.364	0.1970	0.25171
37	0.6458	0.0517	0.6346	0.0814	0.02198	87	1.5184	0.2746	1.377	0.1995	0.25990
38	0.6632	0.0545	0.6511	0.0837	0.02378	88	1.5359	0.2807	1.389	0.2020	0.26825
39	0.6807	0.0574	0.6676	0.0859	0.02568	89	1.5533	0.2867	1.402	0.2046	0.27675
40	0.6981	0.0603	0.6840	0.0882	0.02767	90	1.5708	0.2929	1.414	0.2071	0.28540
41	0.7156	0.0633	0.7004	0.0904	0.02976	91	1.5882	0.2991	1.427	0.2097	0.29420
42	0.7330	0.0664	0.7167	0.0927	0.03195	92	1.6057	0.3053	1.439	0.2122	0.30316
43	0.7505	0.0696	0.7330	0.0949	0.03425	93	1.6232	0.3116	1.451	0.2148	0.31226
44	0.7679	0.0728	0.7492	0.0972	0.03664	94	1.6406	0.3180	1.463	0.2174	0.32152
45	0.7854	0.0761	0.7654	0.0995	0.03915	95	1.6581	0.3244	1.475	0.2200	0.33093
46	0.8029	0.0795	0.7815	0.1017	0.04176	96	1.6755	0.3309	1.486	0.2226	0.34050
47	0.8203	0.0829	0.7975	0.1040	0.04448	97	1.6930	0.3374	1.498	0.2252	0.35021
48	0.8378	0.0865	0.8135	0.1063	0.04731	98	1.7104	0.3439	1.509	0.2279	0.36008
49	0.8552	0.0900	0.8294	0.1086	0.05025	99	1.7279	0.3506	1.521	0.2305	0.37009
50	0.8727	0.0937	0.8452	0.1108	0.05331	100	1.7453	0.3572	1.532	0.2332	0.38026

(continued)

TABLE 51.12 (Continued)

Central Angle in Degrees	Arc <i>R</i>	Height <i>R</i>	Chord <i>R</i>	Height Chord	Area <i>R</i> ²	Central Angle in Degrees	Arc <i>R</i>	Height <i>R</i>	Chord <i>R</i>	Height Chord	Area <i>R</i> ²
101	1.7628	0.3639	1.543	0.2358	0.39058	141	2.4609	0.6662	1.885	0.3534	0.91580
102	1.7802	0.3707	1.554	0.2385	0.40104	142	2.4784	0.6744	1.891	0.3566	0.93135
103	1.7977	0.3775	1.565	0.2412	0.41166	143	2.4958	0.6827	1.897	0.3599	0.94700
104	1.8151	0.3843	1.576	0.2439	0.42242	144	2.5133	0.6910	1.902	0.3633	0.96274
105	1.8326	0.3912	1.587	0.2466	0.43333	145	2.5307	0.6993	1.907	0.3666	0.97858
106	1.8500	0.3982	1.597	0.2493	0.44439	146	2.5482	0.7076	1.913	0.3700	0.99449
107	1.8675	0.4052	1.608	0.2520	0.45560	147	2.5656	0.7160	1.918	0.3734	1.0105
108	1.8850	0.4122	1.618	0.2548	0.46695	148	2.5831	0.7244	1.923	0.3768	1.0266
109	1.9024	0.4193	1.628	0.2575	0.47844	149	2.6005	0.7328	1.927	0.3802	1.0428
110	1.9199	0.4264	1.638	0.2603	0.49008	150	2.6180	0.7412	1.932	0.3837	1.0590
111	1.9373	0.4336	1.648	0.2631	0.50187	151	2.6354	0.7496	1.936	0.3871	1.0753
112	1.9548	0.4408	1.658	0.2659	0.51379	152	2.6529	0.7581	1.941	0.3906	1.0917
113	1.9722	0.4481	1.668	0.2687	0.52586	153	2.6704	0.7666	1.945	0.3942	1.1082
114	1.9897	0.4554	1.677	0.2715	0.53807	154	2.6878	0.7750	1.949	0.3977	1.1247
115	2.0071	0.4627	1.687	0.2743	0.55041	155	2.7053	0.7836	1.953	0.4013	1.1413
116	2.0246	0.4701	1.696	0.2772	0.56289	156	2.7227	0.7921	1.956	0.4049	1.1580
117	2.0420	0.4775	1.705	0.2800	0.57551	157	2.7402	0.8006	1.960	0.4085	1.1747
118	2.0595	0.4850	1.714	0.2829	0.58827	158	2.7576	0.8092	1.963	0.4122	1.1915
119	2.0769	0.4925	1.723	0.2858	0.60116	159	2.7751	0.8178	1.967	0.4158	1.2084
120	2.0944	0.5000	1.732	0.2887	0.61418	160	2.7925	0.8264	1.970	0.4195	1.2253
121	2.1118	0.5076	1.741	0.2916	0.62734	161	2.8100	0.8350	1.973	0.4233	1.2422
122	2.1293	0.5152	1.749	0.2945	0.64063	162	2.8274	0.8436	1.975	0.4270	1.2592
123	2.1468	0.5228	1.758	0.2975	0.65404	163	2.8449	0.8522	1.978	0.4308	1.2763
124	2.1642	0.5305	1.766	0.3004	0.66759	164	2.8623	0.8608	1.981	0.4346	1.2934
125	2.1817	0.5383	1.774	0.3034	0.68125	165	2.8798	0.8695	1.983	0.4385	1.3105
126	2.1991	0.5460	1.782	0.3064	0.69505	166	2.8972	0.8781	1.985	0.4424	1.3277
127	2.2166	0.5538	1.790	0.3094	0.70897	167	2.9147	0.8868	1.987	0.4463	1.3449
128	2.2340	0.5616	1.798	0.3124	0.72301	168	2.9322	0.8955	1.989	0.4502	1.3621
129	2.2515	0.5695	1.805	0.3155	0.73716	169	2.9496	0.9042	1.991	0.4542	1.3794
130	2.2689	0.5774	1.813	0.3185	0.75143	170	2.9671	0.9128	1.992	0.4582	1.3967
131	2.2864	0.5853	1.820	0.3216	0.76584	171	2.9845	0.9215	1.994	0.4622	1.4140
132	2.3038	0.5933	1.827	0.3247	0.78034	172	3.0020	0.9302	1.995	0.4663	1.4314
133	2.3213	0.6013	1.834	0.3278	0.79497	173	3.0194	0.9390	1.996	0.4704	1.4488
134	2.3387	0.6093	1.841	0.3309	0.80970	174	3.0369	0.9477	1.997	0.4745	1.4662
135	2.3562	0.6173	1.848	0.3341	0.82454	175	3.0543	0.9564	1.998	0.4786	1.4836
136	2.3736	0.6254	1.854	0.3373	0.83949	176	3.0718	0.9651	1.999	0.4828	1.5010
137	2.3911	0.6335	1.861	0.3404	0.85455	177	3.0892	0.9738	1.999	0.4871	1.5184
138	2.4086	0.6416	1.867	0.3436	0.86971	178	3.1067	0.9825	2.000	0.4914	1.5359
139	2.4260	0.6498	1.873	0.3469	0.88497	179	3.1241	0.9913	2.000	0.4957	1.5533
140	2.4435	0.6580	1.879	0.3501	0.90034	180	3.1416	1.0000	2.000	0.5000	1.5708

TABLE 51.13 Values of Degrees, Minutes, and Seconds in Radians^a

Degrees	Radians Arc Length $R = 1$	Degrees	Radians Arc Length $R = 1$	Radians Arc Length $R = 1$	
				Minutes	Seconds
0		48	0.83775804	0	
1	0.01745329	49	0.85521133	1	0.00029089
2	0.03490659	50	0.87266463	2	0.00058178
3	0.05235988	51	0.89011792	3	0.00087266
4	0.06981317	52	0.90757121	4	0.00116355
5	0.08726646	53	0.92502450	5	0.00145444
6	0.10471976	54	0.94247780	6	0.00174533
7	0.12217305	55	0.95993109	7	0.00203622
8	0.13962634	56	0.97738438	8	0.00232711
9	0.15707963	57	0.99483767	9	0.00261799
10	0.17453293	58	1.01229097	10	0.00290888
11	0.19198622	59	1.02974426	11	0.00319977
12	0.20943951	60	1.04719755	12	0.00349066
13	0.22689280	61	1.06465084	13	0.00378155
14	0.24434610	62	1.08210414	14	0.00407243
15	0.26179939	63	1.09955743	15	0.00436332
16	0.27925268	64	1.11701072	16	0.00465421
17	0.29670597	65	1.13446401	17	0.00494510
18	0.31415927	66	1.15191731	18	0.00523599
19	0.33161256	67	1.16937060	19	0.00552688
20	0.34906585	68	1.18682389	20	0.00581776
21	0.36651914	69	1.20427718	21	0.00610865
22	0.38397244	70	1.22173048	22	0.00639954
23	0.40142573	71	1.23918377	23	0.00669043
24	0.41887902	72	1.25663706	24	0.00698132
25	0.43633231	73	1.27409035	25	0.00727221
26	0.45378561	74	1.29154365	26	0.00756309
27	0.47123890	75	1.30899694	27	0.00785398
28	0.48869219	76	1.32645023	28	0.00814487
29	0.50614548	77	1.34390352	29	0.00843576
30	0.52359878	78	1.36135682	30	0.00872665
31	0.54105207	79	1.37881011	31	0.00901753
32	0.55850536	80	1.39626340	32	0.00930842
33	0.57595865	81	1.41371669	33	0.00959931
34	0.59341195	82	1.43116999	34	0.00989020
35	0.61086524	83	1.44862328	35	0.01018109
36	0.62831853	84	1.46607657	36	0.01047198
37	0.64577182	85	1.48352986	37	0.01076286
38	0.66322512	86	1.50098316	38	0.01105375
39	0.68067841	87	1.51843645	39	0.01134464
40	0.69813170	88	1.53588974	40	0.01163553
41	0.71558499	89	1.55334303	41	0.01192642
42	0.73303829	90	1.57079633	42	0.01221730
43	0.75049158	91	1.58824962	43	0.01250819
44	0.76794487	92	1.60570291	44	0.01279908
45	0.78539816	93	1.62315620	45	0.01308997
46	0.80285146	94	1.64060950	46	0.01338086
47	0.82030475	95	1.65806279	47	0.01367175

(continued)

TABLE 51.13 (Continued)

Degrees	Radians Arc Length $R = 1$	Degrees	Radians Arc Length $R = 1$	Radians Arc Length $R = 1$		
				Minutes	Seconds	
96	1.67551608	139	2.42600766	48	0.01396263	0.00023271
97	1.69296937	140	2.44346095	49	0.01425352	0.00023756
98	1.71042267	141	2.46091424	50	0.01454441	0.00024241
99	1.72787596	142	2.47836754	51	0.01483530	0.00024725
100	1.74532925	143	2.49582083	52	0.01512619	0.00025210
101	1.76278254	144	2.51327413	53	0.01541707	0.00025695
102	1.78023584	145	2.53072742	54	0.01570796	0.00026180
103	1.79768913	146	2.54818071	55	0.01599885	0.00026665
104	1.81514242	147	2.56563401	56	0.01628974	0.00027150
105	1.83259571	148	2.58308729	57	0.01658063	0.00027634
106	1.85004901	149	2.60054058	58	0.01687152	0.00028119
107	1.86750230	150	2.61799388	59	0.01716240	0.00028604
108	1.88495559	151	2.63544717			
109	1.90240888	152	2.65290046			
110	1.91986218	153	2.67035375			
111	1.93731547	154	2.68780705			
112	1.95476876	155	2.70526034			
113	1.97222205	156	2.72271363			
114	1.98967535	157	2.74016693			
115	2.00712864	158	2.75762022			
116	2.02458193	159	2.77507351			
117	2.04203522	160	2.79252680			
118	2.05948852	161	2.80998009			
119	2.07694181	162	2.82743338			
120	2.09439510	163	2.84488668			
121	2.11184840	164	2.86233997			
122	2.12930169	165	2.87979327			
123	2.14675498	166	2.89724655			
124	2.16420828	167	2.91469985			
125	2.18166157	168	2.93215314			
126	2.19911486	169	2.94960643			
127	2.21656815	170	2.96705972			
128	2.23402145	171	2.98451302			
129	2.25147474	172	3.00196631			
130	2.26892803	173	3.01941961			
131	2.28638133	174	3.03687289			
132	2.30383462	175	3.05432619			
133	2.32128791	176	3.07177948			
134	2.33874121	177	3.08923277			
135	2.35619450	178	3.10668607			
136	2.37364780	179	3.12413962			
137	2.39110107	180	3.14159265			
138	2.40855436					

^aLengths of circular arcs, radius unity, for example:

$$\Theta = 30^{\circ}20'10''$$

$$30^{\circ} = 0.52359878$$

$$20' = 0.00581776$$

$$10'' = 0.00004848$$

$$\text{Arc length} = 0.52946502 \text{ radians}$$

TABLE 51.14 Values of Radians in Degrees

Radian	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.0000	0.5730	1.1459	1.7189	2.2918	2.8648	3.4377	4.0107	4.5837	5.1566
0.1	5.7296	6.3025	6.8755	7.4485	8.0214	8.5944	9.1673	9.7403	10.3132	10.8862
0.2	11.4591	12.0321	12.6051	13.1780	13.7510	14.3239	14.8969	15.4699	16.0428	16.6158
0.3	17.1887	17.7617	18.3346	18.9076	19.4806	20.0535	20.6265	21.1994	21.7724	22.3454
0.4	22.9183	23.4913	24.0642	24.6372	25.2101	25.7831	26.3561	26.9290	27.5020	28.0749
0.5	28.6479	29.2208	29.7938	30.3668	30.9397	31.5127	32.0856	32.6586	33.2316	33.8045
0.6	34.3775	34.9504	35.5234	36.0963	36.6693	37.2423	37.8152	38.3882	38.9611	39.5341
0.7	40.1070	40.6800	41.2530	41.8259	42.3989	42.9718	43.5448	44.1178	44.6907	45.2637
0.8	45.8366	46.4096	46.9825	47.5555	48.1285	48.7014	49.2744	49.8473	50.4203	50.9932
0.9	51.5662	52.1392	52.7121	53.2851	53.8580	54.4310	55.0039	55.5769	56.1499	56.7228
<div>1 rad = 57.29578°2 rads = 114.59156°3 rads = 171.88734°</div>										

TABLE 51.15 Decimals of a Degree in Minutes and Seconds

Decimal	0.00		0.01		0.02		0.03		0.04		0.05		0.06		0.07		0.08		0.09	
	Min	Sec	Min	Sec	Min	Sec	Min	Sec	Min	Sec	Min	Sec	Min	Sec	Min	Sec	Min	Sec	Min	Sec
0.0	0	0	0	36	1	12	1	48	2	24	3	0	3	36	4	12	4	48	5	24
0.1	6	0	6	36	7	12	7	48	8	24	9	0	9	36	10	12	10	48	11	24
0.2	12	0	12	36	13	12	13	48	14	24	15	0	15	36	16	12	16	48	17	24
0.3	18	0	18	36	19	12	19	48	20	24	21	0	21	36	22	12	22	48	23	24
0.4	24	0	24	36	25	12	25	48	26	24	27	0	27	36	28	12	28	48	29	24
0.5	30	0	30	36	31	12	31	48	32	24	33	0	33	36	34	12	34	48	35	24
0.6	36	0	36	36	37	12	37	48	38	24	39	0	39	36	40	12	40	48	41	24
0.7	42	0	42	36	43	12	43	48	44	24	45	0	45	36	46	12	46	48	47	24
0.8	48	0	48	36	49	12	49	48	50	24	51	0	51	36	52	12	52	48	53	24
0.9	54	0	54	36	55	12	55	48	56	24	57	0	57	36	58	12	58	48	59	24

TABLE 51.16 Minutes in Decimals of a Degree

Minutes	0	1	2	3	4	5	6	7	8	9
0	0.00000	0.01667	0.03333	0.05000	0.06667	0.08333	0.10000	0.11667	0.13333	0.15000
10	0.16667	0.18333	0.20000	0.21667	0.23333	0.25000	0.26667	0.28333	0.30000	0.31667
20	0.33333	0.35000	0.36667	0.38333	0.40000	0.41667	0.43333	0.45000	0.46667	0.48333
30	0.50000	0.51667	0.53333	0.55000	0.56667	0.58333	0.60000	0.61667	0.63333	0.65000
40	0.66667	0.68333	0.70000	0.71667	0.73333	0.75000	0.76667	0.78333	0.80000	0.81667
50	0.83333	0.85000	0.86667	0.88333	0.90000	0.91667	0.93333	0.95000	0.96667	0.98333

TABLE 51.17 Seconds in Decimals of a Degree

Seconds	0	1	2	3	4
0	0	0.0002778	0.0005555	0.0008333	0.0011111
10	0.0027778	0.0030555	0.0033333	0.0036111	0.0038888
20	0.0055555	0.0058333	0.0061111	0.0063888	0.0066667
30	0.0083333	0.0086111	0.0088888	0.0091667	0.0094444
40	0.0111111	0.0113888	0.0116667	0.0119444	0.0122222
50	0.0138888	0.0141667	0.0144444	0.0147222	0.0150000
Seconds	5	6	7	8	9
0	0.0013888	0.0016667	0.0019444	0.0022222	0.0024999
10	0.0041667	0.0044444	0.0047222	0.0050000	0.0052778
20	0.0069444	0.0072222	0.0075000	0.0077778	0.0080555
30	0.0097222	0.0100000	0.0102778	0.0105555	0.0108333
40	0.0125000	0.0127778	0.0130555	0.0133333	0.0136111
50	0.0152778	0.0155555	0.0158333	0.0161111	0.0163888

TABLE 51.18 Table of Integrals

<i>Elementary Indefinite Integrals</i>	
1.	$\int a \, dx = ax$
2.	$\int (u + v + w + \cdots) dx = \int u \, dx + \int v \, dx + \int w \, dx + \cdots$
3.	$\int u \, dv = uv - \int v \, du$, integration by parts
4.	$\int f(x) dx = \int f[\phi(y)] \phi'(y) dy$, $x = \phi(y)$, change of variable
5.	$\int x^n dx = \frac{x^{n+1}}{n+1}$ ($n \neq -1$)
6.	$\int \frac{dx}{x} = \log_e x + c = \log_e c_1 x$, [$\log_e x = \log_e(-x) + (2k+1)\pi i$]
7.	$\int e^{ax} dx = \frac{1}{a} e^{ax}$
8.	$\int a^x dx = \frac{a^x}{\log_e a}$
9.	$\int a^x \log_e a \, dx = a^x$.
10.	$\int \sin ax \, dx = -\frac{1}{a} \cos ax$
11.	$\int \cos ax \, dx = \frac{1}{a} \sin ax$
12.	$\int \tan ax \, dx = -\frac{1}{a} \log_e \cos ax = \frac{1}{a} \log_e \sec ax$

TABLE 51.18 (Continued)

-
13. $\int \cot ax \, dx = \frac{1}{a} \log_e \sin ax = -\frac{1}{a} \log_e \csc ax$
 14. $\int \sec ax \, dx = \frac{1}{a} \log_e (\sec ax + \tan ax) = \frac{1}{a} \log_e \tan \left(\frac{ax}{2} + \frac{\pi}{4} \right)$
 15. $\int \csc ax \, dx = \frac{1}{a} \log_e (\csc ax - \cot ax) = \frac{1}{a} \log_e \tan \frac{ax}{2}$
 16. $\int \frac{dx}{\sqrt{a^2 - x^2}} = \sin^{-1} \frac{x}{a} = -\cos^{-1} \frac{x}{a} \quad (x^2 < a^2)$
 17. $\int \frac{dx}{a^2 + x^2} = \frac{1}{a} \tan^{-1} \frac{x}{a} = -\frac{1}{a} \cot^{-1} \frac{x}{a}$
 18. $\int \sinh ax \, dx = \frac{1}{a} \cosh ax$
 19. $\int \cosh ax \, dx = \frac{1}{a} \sinh ax$
 20. $\int \tanh ax \, dx = \frac{1}{a} \log_e (\cosh ax)$
 21. $\int \coth ax \, dx = \frac{1}{a} \log_e (\sinh ax)$
 22. $\int \operatorname{sech} ax \, dx = \frac{1}{a} \sin^{-1} (\tanh ax) = \frac{1}{a} \tan^{-1} (\sinh ax)$
 23. $\int \operatorname{csch} ax \, dx = \frac{1}{a} \log_e \left(\tanh \frac{ax}{2} \right)$
 24. $\int \sin^2 ax \, dx = \frac{1}{2} x - \frac{1}{2a} \sin ax \cos ax = \frac{1}{2} x - \frac{1}{4a} \sin 2ax$
 25. $\int \cos^2 ax \, dx = \frac{1}{2} x + \frac{1}{2a} \sin ax \cos ax = \frac{1}{2} x + \frac{1}{4a} \sin 2ax$
 26. $\int \tan^2 ax \, dx = \frac{1}{a} \tan ax - x$
 27. $\int \cot^2 ax \, dx = -\frac{1}{a} \cot ax - x$
 28. $\int \sec^2 ax \, dx = \frac{1}{a} \tan ax$
 29. $\int \csc^2 ax \, dx = -\frac{1}{a} \cot ax$
 30. $\int \sin^{-1} ax \, dx = x \sin^{-1} ax + \frac{1}{a} \sqrt{1 - a^2 x^2}$
 31. $\int \cos^{-1} ax \, dx = x \cos^{-1} ax - \frac{1}{a} \sqrt{1 - a^2 x^2}$
 32. $\int \tan^{-1} ax \, dx = x \tan^{-1} ax - \frac{1}{2a} \log_e (1 + a^2 x^2)$
 33. $\int \cot^{-1} ax \, dx = x \cot^{-1} ax + \frac{1}{2a} \log_e (1 + a^2 x^2)$

(continued)

TABLE 51.18 (Continued)

$$34. \int \sec^{-1} ax \, dx = x \sec^{-1} ax - \frac{1}{a} \log_e(ax + \sqrt{a^2 x^2 - 1})$$

$$35. \int \csc^{-1} ax \, dx = x \csc^{-1} ax + \frac{1}{a} \log_e(ax + \sqrt{a^2 x^2 - 1})$$

Integrals Involving $ax + b$

$$36. \int (ax + b)^n \, dx = \frac{1}{a(n+1)} (ax + b)^{n+1} \quad (n \neq -1)$$

$$37. \int \frac{dx}{ax + b} = \frac{1}{a} \log_e(ax + b)$$

$$38. \int x(ax + b)^n \, dx = \frac{1}{a^2(n+2)} (ax + b)^{n+2} - \frac{b}{a^2(n+1)} (ax + b)^{n+1} \quad (n \neq -1, -2)$$

$$39. \int \frac{x \, dx}{ax + b} = \frac{x}{a} - \frac{b}{a^2} \log_e(ax + b)$$

$$40. \int \frac{x \, dx}{(ax + b)^2} = \frac{b}{a^2(ax + b)} + \frac{1}{a^2} \log_e(ax + b)$$

$$41. \int \frac{x^2 \, dx}{ax + b} = \frac{1}{a^3} \left[\frac{1}{2} (ax + b)^2 - 2b(ax + b) + b^2 \log_e(ax + b) \right]$$

$$42. \int \frac{x^2 \, dx}{(ax + b)^2} = \frac{1}{a^3} \left[(ax + b) - 2b \log_e(ax + b) - \frac{b^2}{ax + b} \right]$$

$$43. \int \frac{x^2 \, dx}{(ax + b)^3} = \frac{1}{a^3} \left[\log_e(ax + b) + \frac{2b}{ax + b} - \frac{b^2}{2(ax + b)^2} \right]$$

$$44. \int \frac{dx}{x(ax + b)} = \frac{1}{b} \log_e \frac{x}{ax + b}$$

$$45. \int \frac{dx}{x^2(ax + b)} = -\frac{1}{bx} + \frac{a}{b^2} \log_e \frac{ax + b}{x}$$

$$46. \int \frac{dx}{x(ax + b)^2} = \frac{1}{b(ax + b)} - \frac{1}{b^2} \log_e \frac{ax + b}{x}$$

$$47. \int \frac{dx}{x^2(ax + b)^2} = -\frac{b + 2ax}{b^2 x(ax + b)} + \frac{2a}{b^3} \log_e \frac{ax + b}{x}$$

$$48. \int \frac{dx}{x\sqrt{ax + b}} = \frac{1}{\sqrt{b}} \log_e \frac{\sqrt{ax + b} - \sqrt{b}}{\sqrt{ax + b} + \sqrt{b}} \quad (b \text{ positive})$$

$$49. \int \frac{dx}{x\sqrt{ax + b}} = \frac{2}{\sqrt{-b}} \tan^{-1} \frac{\sqrt{ax + b}}{-b} \quad (b \text{ negative})$$

$$50. \int \frac{\sqrt{ax + b}}{x} \, dx = 2\sqrt{ax + b} + \sqrt{b} \log_e \frac{\sqrt{ax + b} - \sqrt{b}}{\sqrt{ax + b} + \sqrt{b}} \quad (b \text{ positive})$$

$$51. \int \frac{\sqrt{ax + b}}{x} \, dx = 2\sqrt{ax + b} - 2\sqrt{-b} \tan^{-1} \sqrt{\frac{ax + b}{-b}} \quad (b \text{ negative})$$

$$52. \int \frac{dx}{x^2\sqrt{ax + b}} = -\frac{\sqrt{ax + b}}{bx} - \frac{a}{2b\sqrt{b}} \log_e \frac{\sqrt{ax + b} - \sqrt{b}}{\sqrt{ax + b} + \sqrt{b}} \quad (b \text{ positive})$$

TABLE 51.18 (Continued)

- $$53. \int \frac{dx}{x^2\sqrt{ax+b}} = -\frac{\sqrt{ax+b}}{bx} - \frac{a}{b\sqrt{-b}} \tan^{-1} \sqrt{\frac{ax+b}{-b}} \quad (b \text{ negative})$$
- $$54. \int \frac{ax+b}{fx+g} dx = \frac{ax}{f} + \frac{bf-cg}{f^2} \log_e(fx+g)$$
- $$55. \int \frac{dx}{(ax+b)(fx+g)} = \frac{1}{bf-cg} \log_e \left(\frac{fx+g}{ax+b} \right) \quad (ag \neq bf)$$
- $$56. \int \frac{x dx}{(ax+b)(fx+g)} = \frac{1}{bf-ag} \left(\frac{b}{a} \log_e(ax+b) - \frac{g}{f} \log_e(fx+g) \right) \quad (ag \neq bf)$$
- $$57. \int \frac{dx}{(ax+b)^2(fx+g)} = \frac{1}{bf-ag} \left(\frac{1}{ax+b} + \frac{f}{bf-ag} \log_e \frac{fx+g}{ax+b} \right) \quad (ag \neq bf)$$

Integrals Involving $ax^n + b$

- $$58. \int (ax^2 + b)^n x dx = \frac{1}{2a} \frac{(ax^2 + b)^{n+1}}{n+1} \quad (n \neq -1)$$
- $$59. \int \frac{dx}{ax^2 + b} = \frac{1}{\sqrt{ab}} \tan^{-1} \left(x \sqrt{\frac{a}{b}} \right) \quad (a \text{ and } b \text{ positive})$$
- $$60. \int \frac{dx}{ax^2 + b} = \frac{1}{2\sqrt{-ab}} \log_e \frac{x\sqrt{a} - \sqrt{-b}}{x\sqrt{a} + \sqrt{-b}} \quad (a \text{ positive, } b \text{ negative})$$
- $$= \frac{1}{2\sqrt{-ab}} \log_e \frac{\sqrt{b} + x\sqrt{-a}}{\sqrt{b} - x\sqrt{-a}} \quad (a \text{ negative, } b \text{ positive})$$
- $$61. \int \frac{dx}{x(ax^2 + b)} = \frac{1}{2b} \log_e \frac{x^2}{ax^2 + b}$$
- $$62. \int \frac{dx}{(ax^2 + b)^n} = \frac{1}{2(n-1)b} \frac{x}{(ax^2 + b)^{n-1}} + \frac{2n-3}{2(n-1)b} \int \frac{dx}{(ax^2 + b)^{n-1}} \quad (n \text{ integer } > 1)$$
- $$63. \int \frac{x^2 dx}{ax^2 + b} = \frac{x}{a} - \frac{b}{a} \int \frac{dx}{ax^2 + b}$$
- $$64. \int \frac{x^2 dx}{(ax^2 + b)^n} = \frac{1}{2(n-1)a} \frac{x}{(ax^2 + b)^{n-1}} + \frac{1}{2(n-1)a} \int \frac{dx}{(ax^2 + b)^{n-1}} \quad (n \text{ integer } > 1)$$
- $$65. \int \frac{dx}{x^2(ax^2 + b)^n} = \frac{1}{b} \int \frac{dx}{x^2(ax^2 + b)^{n-1}} - \frac{a}{b} \int \frac{dx}{(ax^2 + b)^n} \quad (n = \text{positive integer})$$
- $$66. \int \sqrt{ax^2 + b} dx = \frac{x}{2} \sqrt{ax^2 + b} + \frac{b}{2\sqrt{a}} \log_e \frac{x\sqrt{a} + \sqrt{ax^2 + b}}{\sqrt{b}} \quad (a \text{ positive})$$
- $$67. \int \sqrt{ax^2 + b} dx = \frac{x}{2} \sqrt{ax^2 + b} + \frac{b}{2\sqrt{-a}} \sin^{-1} \left(x \sqrt{-\frac{a}{b}} \right) \quad (a \text{ negative})$$

(continued)

TABLE 51.18 (Continued)

-
68. $\int \frac{dx}{\sqrt{ax^2 + b}} = \frac{1}{\sqrt{a}} \log_e(x\sqrt{a} + \sqrt{ax^2 + b})_n \quad (a \text{ positive})$
69. $\int \frac{dx}{\sqrt{ax^2 + b}} = \frac{1}{\sqrt{-a}} \sin^{-1}\left(x\sqrt{-\frac{a}{b}}\right) \quad (a \text{ negative})$
70. $\int \frac{x dx}{\sqrt{ax^2 + b}} = \frac{1}{a} \sqrt{ax^2 + b}$
71. $\int \frac{\sqrt{ax^2 + b}}{x} dx = \sqrt{ax^2 + b} + \sqrt{b} \log_e \frac{\sqrt{ax^2 + b} - \sqrt{b}}{x} \quad (b \text{ positive})$
72. $\int \frac{\sqrt{ax^2 + b}}{x} dx = \sqrt{ax^2 + b} - \sqrt{-b} \tan^{-1} \frac{\sqrt{ax^2 + b}}{\sqrt{-b}} \quad (b \text{ negative})$
73. $\int x\sqrt{ax^2 + b} dx = \frac{1}{3a} (ax^2 + b)^{3/2}$
74. $\int x^2 \sqrt{ax^2 + b} dx = \frac{x}{4a} (ax^2 + b)^{3/2} - \frac{bx}{8a} \sqrt{ax^2 + b} - \frac{b^2}{8a\sqrt{a}} \log_e(x\sqrt{a} + \sqrt{ax^2 + b}) \quad (a \text{ positive})$
75. $\int x^2 \sqrt{ax^2 + b} dx = \frac{x}{4a} (ax^2 + b)^{3/2} - \frac{bx}{8a} \sqrt{ax^2 + b} - \frac{b^2}{8a\sqrt{-a}} \sin^{-1}\left(x\sqrt{-\frac{a}{b}}\right) \quad (a \text{ negative})$
76. $\int \frac{dx}{x\sqrt{ax^2 + b}} = \frac{1}{\sqrt{b}} \log_e \left(\frac{\sqrt{ax^2 + b} - \sqrt{b}}{x} \right) \quad (b \text{ positive})$
77. $\int \frac{dx}{x\sqrt{ax^2 + b}} = \frac{1}{\sqrt{-b}} \sec^{-1}\left(x\sqrt{-\frac{a}{b}}\right) \quad (b \text{ negative})$
78. $\int \frac{x^2 dx}{\sqrt{ax^2 + b}} = \frac{x}{2a} \sqrt{ax^2 + b} - \frac{b}{2a\sqrt{a}} \log_e(x\sqrt{a} + \sqrt{ax^2 + b}) \quad (a \text{ positive})$
79. $\int \frac{x^2 dx}{\sqrt{ax^2 + b}} = \frac{x}{2a} \sqrt{ax^2 + b} - \frac{b}{2a\sqrt{-a}} \sin^{-1}\left(x\sqrt{-\frac{a}{b}}\right) \quad (a \text{ negative})$
80. $\int \frac{\sqrt{ax^2 + b}}{x^2} dx = -\frac{\sqrt{ax^2 + b}}{x} + \sqrt{a} \log_e(x\sqrt{a} + \sqrt{ax^2 + b}) \quad (a \text{ positive})$
81. $\int \frac{\sqrt{ax^2 + b}}{x^2} dx = -\frac{\sqrt{ax^2 + b}}{x} - \sqrt{-a} \sin^{-1}\left(x\sqrt{-\frac{a}{b}}\right) \quad (a \text{ negative})$
82. $\int \frac{dx}{x(ax^n + b)} = \frac{1}{bn} \log_e \frac{x^n}{ax^n + b}$
83. $\int \frac{dx}{x\sqrt{ax^n + b}} = \frac{1}{n\sqrt{b}} \log_e \frac{\sqrt{ax^n + b} - \sqrt{b}}{\sqrt{ax^n + b} + \sqrt{b}} \quad (b \text{ positive})$
84. $\int \frac{dx}{x\sqrt{ax^n + b}} = \frac{2}{n\sqrt{-b}} \sec^{-1} \sqrt{-\frac{ax^n}{b}} \quad (b \text{ negative})$

TABLE 51.18 (Continued)

 Integrals Involving $ax^2 + bx + d$

85. $\int \frac{dx}{ax^2 + bx + d} = \frac{1}{\sqrt{b^2 - 4ad}} \log_e \frac{2ax + b - \sqrt{b^2 - 4ad}}{2ax + b + \sqrt{b^2 - 4ad}} \quad (b^2 > 4ad)$
86. $\int \frac{dx}{ax^2 + bx + d} = \frac{2}{\sqrt{4ad - b^2}} \tan^{-1} \frac{2ax + b}{\sqrt{4ad - b^2}} \quad (b^2 < 4ad)$
87. $\int \frac{dx}{ax^2 + bx + d} = \frac{2}{2ax + b} \quad (b^2 = 4ad)$
88. $\int \frac{dx}{\sqrt{ax^2 + bx + d}} = \frac{1}{\sqrt{a}} \log_e (2ax + b + 2\sqrt{a(ax^2 + bx + d)}) \quad (a \text{ positive})$
89. $\int \frac{dx}{\sqrt{ax^2 + bx + d}} = \frac{1}{\sqrt{-a}} \sin^{-1} \frac{-2ax - b}{\sqrt{b^2 - 4ad}} \quad (a \text{ negative})$
90. $\int \frac{x dx}{ax^2 + bx + d} = \frac{1}{2a} \log_e (ax^2 + bx + d) - \frac{b}{2a} \int \frac{dx}{ax^2 + bx + d}$
91. $\int \frac{x dx}{\sqrt{ax^2 + bx + d}} = \frac{\sqrt{ax^2 + bx + d}}{a} - \frac{b}{2a} \int \frac{dx}{\sqrt{ax^2 + bx + d}}$
92. $\int \frac{dx}{x\sqrt{ax^2 + bx + d}} = -\frac{1}{\sqrt{d}} \log_e \left(\frac{\sqrt{ax^2 + bx + d} + \sqrt{d}}{x} + \frac{b}{2\sqrt{d}} \right) \quad (d \text{ positive})$
93. $\int \frac{dx}{x\sqrt{ax^2 + bx + d}} = \frac{1}{\sqrt{-d}} \sin^{-1} \frac{bx + 2d}{x\sqrt{b^2 - 4ad}} \quad (d \text{ negative})$
94. $\int \frac{dx}{x\sqrt{ax^2 + bx}} = -\frac{2}{bx} \sqrt{ax^2 + bx}$
95. $\int \sqrt{ax^2 + bx + d} dx = \frac{2ax + b}{4a} \sqrt{ax^2 + bx + d} + \frac{4ad - b^2}{8a} \int \frac{dx}{\sqrt{ax^2 + bx + d}}$
96. $\int x\sqrt{ax^2 + bx + d} dx = \frac{(ax^2 + bx + d)^{3/2}}{3a} - \frac{b}{2a} \int \sqrt{ax^2 + bx + d} dx$

 Integrals Involving $\sin^n ax$

97. $\int \sin^3 ax dx = -\frac{1}{a} \cos ax + \frac{1}{3a} \cos^3 ax$
98. $\int \sin^4 ax dx = \frac{3}{8}x - \frac{1}{4a} \sin 2ax + \frac{1}{32a} \sin 4ax$
99. $\int \sin^n ax dx = -\frac{\sin^{n-1} ax \cos ax}{na} + \frac{n-1}{n} \int \sin^{n-2} ax dx \quad (n = \text{positive integer})$
100. $\int x \sin ax dx = \frac{\sin ax}{a^2} - \frac{x \cos ax}{a}$
101. $\int x^2 \sin ax dx = \frac{2x}{a^2} \sin ax - \left(\frac{x^2}{a} - \frac{2}{a^3} \right) \cos ax$

(continued)

TABLE 51.18 (Continued)

$$102. \int x^3 \sin ax \, dx = \left(\frac{3x^2}{a^2} - \frac{6}{a^4} \right) \sin ax - \left(\frac{x^3}{a} - \frac{6x}{a^3} \right) \cos ax$$

$$103. \int x^n \sin ax \, dx = -\frac{x^n}{a} \cos ax + \frac{n}{a} \int x^{n-1} \cos ax \, dx \quad (n > 0)$$

$$104. \int \frac{\sin ax}{x^n} dx = -\frac{1}{n-1} \frac{\sin ax}{x^{n-1}} + \frac{a}{n-1} \int \frac{\cos ax}{x^{n-1}} dx$$

$$105. \int \frac{dx}{\sin^n ax} = -\frac{1}{a(n-1)} \frac{\cos ax}{\sin^{n-1} ax} + \frac{n-2}{n-1} \int \frac{dx}{\sin^{n-2} ax} \quad (n \text{ integer} > 1)$$

$$106. \int \frac{x \, dx}{\sin^2 ax} = -\frac{x}{a} \cot ax + \frac{1}{a^2} \log_e \sin ax$$

$$107. \int \frac{dx}{1 + \sin ax} = -\frac{1}{a} \tan\left(\frac{\pi}{4} - \frac{ax}{2}\right)$$

$$108. \int \frac{dx}{1 - \sin ax} = \frac{1}{a} \cot\left(\frac{\pi}{4} - \frac{ax}{2}\right)$$

$$109. \int \frac{x \, dx}{1 + \sin ax} = -\frac{x}{a} \tan\left(\frac{\pi}{4} - \frac{ax}{2}\right) + \frac{2}{a^2} \log_e \cos\left(\frac{\pi}{4} - \frac{ax}{2}\right)$$

$$110. \int \frac{x \, dx}{1 - \sin ax} = \frac{x}{a} \cot\left(\frac{\pi}{4} - \frac{ax}{2}\right) + \frac{2}{a^2} \log_e \sin\left(\frac{\pi}{4} - \frac{ax}{2}\right)$$

$$111. \int \frac{dx}{b + d \sin ax} = \frac{-2}{a\sqrt{b^2 - d^2}} \tan^{-1} \left[\sqrt{\frac{b-d}{b+d}} \tan\left(\frac{\pi}{4} - \frac{ax}{2}\right) \right] \quad (b^2 > d^2)$$

$$112. \int \frac{dx}{b + d \sin ax} = \frac{-1}{a\sqrt{d^2 - b^2}} \log_e \frac{d + b \sin ax + \sqrt{d^2 - b^2} \cos ax}{b + d \sin ax} \quad (d^2 > b^2)$$

$$113. \int \sin ax \sin bx \, dx = \frac{\sin(a-b)x}{2(a-b)} - \frac{\sin(a+b)x}{2(a+b)} \quad (a^2 \neq b^2)$$

Integrals Involving $\cos^n ax$

$$114. \int \cos^3 ax \, dx = \frac{1}{a} \sin ax - \frac{1}{3a} \sin^3 ax$$

$$115. \int \cos^4 ax \, dx = \frac{3}{8} x + \frac{1}{4a} \sin 2ax + \frac{1}{32a} \sin 4ax$$

$$116. \int \cos^n ax \, dx = \frac{\cos^{n-1} ax \sin ax}{na} + \frac{n-1}{n} \int \cos^{n-2} ax \, dx \quad (n = \text{positive integer})$$

$$117. \int x \cos ax \, dx = \frac{\cos ax}{a^2} + \frac{x \sin ax}{a}$$

$$118. \int x^2 \cos ax \, dx = \frac{2x}{a^2} \cos ax + \left(\frac{x^2}{a} - \frac{2}{a^3} \right) \sin ax$$

$$119. \int x^3 \cos ax \, dx = \left(\frac{3x^2}{a^2} - \frac{6}{a^4} \right) \cos ax + \left(\frac{x^3}{a} - \frac{6x}{a^3} \right) \sin ax$$

$$120. \int x^n \cos ax \, dx = \frac{x^n \sin ax}{a} - \frac{n}{a} \int x^{n-1} \sin ax \, dx \quad (n > 0)$$

TABLE 51.18 (Continued)

- $$121. \int \frac{\cos ax}{x^n} dx = \frac{1}{n-1} \frac{\cos ax}{x^{n-1}} - \frac{a}{n-1} \int \frac{\sin ax}{x^{n-1}} dx$$
- $$122. \int \frac{dx}{\cos^n ax} = \frac{1}{a(n-1)} \frac{\sin ax}{\cos^{n-1} ax} + \frac{n-2}{n-1} \int \frac{dx}{\cos^{n-2} ax} \quad (n \text{ integer} > 1)$$
- $$123. \int \frac{x dx}{\cos^2 ax} = \frac{x}{a} \tan ax + \frac{1}{a^2} \log_e \cos ax$$
- $$124. \int \frac{dx}{1 + \cos ax} = \frac{1}{a} \tan \frac{ax}{2}$$
- $$125. \int \frac{dx}{1 - \cos ax} = -\frac{1}{a} \cot \frac{ax}{2}$$
- $$126. \int \frac{x dx}{1 + \cos ax} = \frac{x}{a} \tan \frac{ax}{2} + \frac{2}{a^2} \log_e \cos \frac{ax}{2}$$
- $$127. \int \frac{x dx}{1 - \cos ax} = \frac{x}{a} \cot \frac{ax}{2} + \frac{2}{a^2} \log_e \sin \frac{ax}{2}$$
- $$128. \int \frac{dx}{b + d \cos ax} = \frac{2}{a\sqrt{b^2 - d^2}} \tan^{-1} \left(\sqrt{\frac{b-d}{b+d}} \tan \frac{ax}{2} \right) \quad (b^2 > d^2)$$
- $$129. \int \frac{dx}{b + d \cos ax} = \frac{1}{a\sqrt{d^2 - b^2}} \log_e \frac{d + b \cos ax + \sqrt{d^2 - b^2} \sin ax}{b + d \cos ax} \quad (d^2 > b^2)$$
- $$130. \int \cos ax \cos bx dx = \frac{\sin(a-b)x}{2(a-b)} + \frac{\sin(a+b)x}{2(a+b)} \quad (a^2 \neq b^2)$$

 Integrals Involving $\sin^n ax, \cos^n ax$

- $$131. \int \sin ax \cos bx dx = -\frac{1}{2} \left[\frac{\cos(a-b)x}{a-b} + \frac{\cos(a+b)x}{a+b} \right] \quad (a^2 \neq b^2)$$
- $$132. \int \sin^2 ax \cos^2 ax dx = \frac{x}{8} - \frac{\sin 4ax}{32a}$$
- $$133. \int \sin^n ax \cos ax dx = \frac{1}{a(n+1)} \sin^{n+1} ax \quad (n \neq -1)$$
- $$134. \int \sin ax \cos^n ax dx = \frac{1}{a(n+1)} \cos^{n+1} ax \quad (n \neq -1)$$
- $$135. \int \sin^n ax \cos^m ax dx = -\frac{\sin^{n-1} ax \cos^{m+1} ax}{a(n+m)} + \frac{n-1}{n+m} \int \sin^{n-2} ax \cos^m ax dx \quad (m, n \text{ pos})$$
- $$136. \int \frac{\sin^n ax}{\cos^m ax} dx = \frac{\sin^{n+1} ax}{a(m-1) \cos^{m-1} ax} - \frac{n-m+2}{m-1} \int \frac{\sin^n ax}{\cos^{m-2} ax} dx \quad (m, n \text{ pos}, m \neq 1)$$
- $$137. \int \frac{\cos^m ax}{\sin^n ax} dx = \frac{-\cos^{m+1} ax}{a(n-1) \sin^{n-1} ax} + \frac{n-m-2}{n-1} \int \frac{\cos^m ax}{\sin^{n-2} ax} dx \quad (m, n \text{ pos}, n \neq 1)$$
- $$138. \int \frac{dx}{\sin ax \cos ax} = \frac{1}{a} \log_e \tan ax$$

(continued)

TABLE 51.18 (Continued)

$$139. \int \frac{dx}{b \sin ax + d \cos ax} = \frac{1}{a\sqrt{b^2 + d^2}} \log_e \tan \frac{1}{2} \left(ax + \tan^{-1} \frac{d}{b} \right)$$

$$140. \int \frac{\sin ax}{b + d \cos ax} dx = -\frac{1}{ad} \log_e (b + d \cos ax)$$

$$141. \int \frac{\cos ax}{b + d \sin ax} dx = \frac{1}{ad} \log_e (b + d \sin ax)$$

Integrals Involving $\tan^n ax$, $\cot^n ax$, $\sec^n ax$, $\csc^n ax$

$$142. \int \tan^n ax dx = \frac{1}{a(n-1)} \tan^{n-1} ax - \int \tan^{n-2} ax dx \quad (n \text{ integer} > 1)$$

$$143. \int \cot^n ax dx = -\frac{1}{a(n-1)} \cot^{n-1} ax - \int \cot^{n-2} ax dx \quad (n \text{ integer} > 1)$$

$$144. \int \sec^n ax dx = \frac{1}{a(n-1)} \frac{\sin ax}{\cos^{n-1} ax} + \frac{n-2}{n-1} \int \sec^{n-2} ax dx \quad (n \text{ integer} > 1)$$

$$145. \int \csc^n ax dx = -\frac{1}{a(n-1)} \frac{\cos ax}{\sin^{n-1} ax} + \frac{n-2}{n-1} \int \csc^{n-2} ax dx \quad (n \text{ integer} > 1)$$

$$146. \int \frac{dx}{b + d \tan ax} = \frac{1}{b^2 + d^2} \left[bx + \frac{d}{a} \log_e (b \cos ax + d \sin ax) \right]$$

$$147. \int \frac{dx}{\sqrt{b + d \tan^2 ax}} = \frac{1}{a\sqrt{b-d}} \sin^{-1} \left[\sqrt{\frac{b-d}{b}} \sin ax \right] \quad (b \text{ pos}, b^2 > d^2)$$

$$148. \int \tan ax \sec ax dx = \frac{1}{a} \sec ax$$

$$149. \int \tan^n ax \sec^2 ax dx = \frac{1}{a(n+1)} \tan^{n+1} ax \quad (n \neq -1)$$

$$150. \int \frac{\sec^2 ax dx}{\tan ax} = \frac{1}{a} \log_e \tan ax$$

$$151. \int \cot ax \csc ax dx = -\frac{1}{a} \csc ax$$

$$152. \int \cot^n ax \csc^2 ax dx = -\frac{1}{a(n+1)} \cot^{n+1} ax \quad (n \neq -1)$$

$$153. \int \frac{\csc^2 ax}{\cot ax} dx = -\frac{1}{a} \log_e \cot ax$$

Integrals Involving b^{ax} , e^{ax} , $\sin bx$, $\cos bx$

$$154. \int x b^{ax} dx = \frac{x b^{ax}}{a \log_e b} - \frac{b^{ax}}{a^2 (\log_e b)^2}$$

$$155. \int x e^{ax} dx = \frac{e^{ax}}{a^2} (ax - 1)$$

$$156. \int x^n b^{ax} dx = \frac{x^n b^{ax}}{a \log_e b} - \frac{n}{a \log_e b} \int x^{n-1} b^{ax} dx \quad (n \text{ positive})$$

$$157. \int x^n e^{ax} dx = \frac{1}{a} x^n e^{ax} - \frac{n}{a} \int x^{n-1} e^{ax} dx \quad (n \text{ positive})$$

TABLE 51.18 (Continued)

$$\begin{aligned}
 158. \quad & \int \frac{dx}{b + de^{ax}} = \frac{1}{ab} [ax - \log_e(b + de^{ax})] \\
 159. \quad & \int \frac{e^{ax} dx}{b + de^{ax}} = \frac{1}{ad} \log_e(b + de^{ax}) \\
 160. \quad & \int \frac{dx}{be^{ax} + de^{-ax}} = \frac{1}{a\sqrt{bd}} \tan^{-1} \left(e^{ax} \sqrt{\frac{b}{d}} \right) \quad (b \text{ and } d \text{ positive}) \\
 161. \quad & \int \frac{e^{ax}}{x} dx = \log_e x + ax + \frac{(ax)^2}{2 \cdot 2!} + \frac{(ax)^3}{3 \cdot 3!} + \dots \\
 162. \quad & \int \frac{e^{ax}}{x^n} dx = \frac{1}{n-1} \left(-\frac{e^{ax}}{x^{n-1}} + a \int \frac{e^{ax}}{x^{n-1}} dx \right) \quad (n \text{ integer} > 1) \\
 163. \quad & \int e^{ax} \sin bx \, dx = \frac{e^{ax}}{a^2 + b^2} (a \sin bx - b \cos bx) \\
 164. \quad & \int e^{ax} \cos bx \, dx = \frac{e^{ax}}{a^2 + b^2} (a \cos bx + b \sin bx) \\
 165. \quad & \int xe^{ax} \sin bx \, dx = \frac{xe^{ax}}{a^2 + b^2} (a \sin bx - b \cos bx) \\
 & \quad - \frac{e^{ax}}{(a^2 + b^2)^2} [(a^2 - b^2) \sin bx - 2ab \cos bx] \\
 166. \quad & \int xe^{ax} \cos bx \, dx = \frac{xe^{ax}}{a^2 + b^2} (a \cos bx + b \sin bx) \\
 & \quad - \frac{e^{ax}}{(a^2 + b^2)^2} [(a^2 - b^2) \cos bx + 2ab \sin bx]
 \end{aligned}$$

Integrals Involving $\log_e ax$

$$\begin{aligned}
 167. \quad & \int \log_e ax \, dx = x \log_e ax - x \\
 168. \quad & \int (\log_e ax)^n dx = x(\log_e ax)^n - n(\log_e ax)^{n-1} dx \quad (n \text{ positive}) \\
 169. \quad & \int x^n \log_e ax \, dx = x^{n+1} \left(\frac{\log_e ax}{n+1} \right) - \frac{1}{(n+1)^2} \quad (n \neq -1) \\
 170. \quad & \int \frac{(\log_e ax)^n}{x} dx = \frac{(\log_e ax)^{n+1}}{n+1} \quad (n \neq -1) \\
 171. \quad & \int \frac{dx}{x \log_e ax} = \log_e (\log_e x) \\
 172. \quad & \int \frac{dx}{\log_e ax} = \frac{1}{a} \left[\log_e (\log_e ax) + \log_e ax + \frac{(\log_e ax)^2}{2 \cdot 2!} + \dots \right] \\
 173. \quad & \int x^m (\log_e ax)^n dx = \frac{x^{m+1} (\log_e ax)^n}{m+1} - \frac{n}{m+1} \int x^m (\log_e ax)^{n-1} dx \quad (m, n \neq -1) \\
 174. \quad & \int \frac{x^m dx}{(\log_e ax)^n} = -\frac{x^{m+1}}{(n-1)(\log_e ax)^{n-1}} + \frac{m+1}{n-1} \int \frac{x^m dx}{(\log_e ax)^{n-1}}
 \end{aligned}$$

(continued)

TABLE 51.18 (Continued)

Some Definite Integrals

1. $\int_0^a \sqrt{a^2 - x^2} dx = \frac{\pi a^2}{4}$
2. $\int_0^a \sqrt{2ax - x^2} dx = \frac{\pi a^2}{4}$
3. $\int_0^\infty \frac{dx}{a + bx^2} = \frac{\pi}{2\sqrt{ab}} \quad (a \text{ and } b \text{ positive})$
4. $\int_0^{\sqrt{a/b}} \frac{dx}{a + bx^2} = \int_{\sqrt{a/b}}^\infty \frac{dx}{a + bx^2} = \frac{\pi}{4\sqrt{ab}} \quad (a \text{ and } b \text{ positive})$
5. $\int_0^{\sqrt{a/b}} \frac{dx}{\sqrt{a - bx^2}} = \frac{\pi}{2\sqrt{b}} \quad (a \text{ and } b \text{ positive})$
6. $\int_0^\infty \frac{\sin bx}{x} dx = \begin{cases} \frac{\pi}{2} & (b > 0) \\ 0 & (b = 0) \\ -\frac{\pi}{2} & (b < 0) \end{cases}$
7. $\int_0^\infty \frac{\tan x}{x} dx = \frac{\pi}{2}$
8. $\int_0^{\pi/2} \sin^{2n+1} x dx = \int_0^{\pi/2} \cos^{2n+1} x dx = \frac{2 \cdot 4 \cdot 6 \cdots 2n}{3 \cdot 5 \cdot 7 \cdots (2n+1)} \quad (n > 0)$
9. $\int_0^{\pi/2} \sin^{2n} x dx = \int_0^{\pi/2} \cos^{2n} x dx = \frac{1 \cdot 3 \cdot 5 \cdots (2n-1)}{2 \cdot 4 \cdot 6 \cdots 2n} \cdot \frac{\pi}{2} \quad (n > 0)$
10. $\int_0^\pi \sin ax \sin bx dx = \int_0^\pi \cos ax \cos bx dx = 0 \quad (a \neq b)$
11. $\int_0^\pi \sin^2 ax dx = \int_0^\pi \cos^2 ax dx = \frac{\pi}{2}$
12. $\int_0^{\pi/2} \log_e \cos x dx = \int_0^{\pi/2} \log_e \sin x dx = -\frac{\pi}{2} \log_e 2$
13. $\int_0^\infty e^{-ax^2} dx = \frac{1}{2} \sqrt{\frac{\pi}{a}}$
14. $\int_0^\infty x^n e^{-ax} dx = \frac{n!}{a^{n+1}} \quad (a > 0, n = 1, 2, 3, \dots)$
15. $\int_0^1 \frac{\log_e x}{1-x} dx = -\frac{\pi^2}{6}$
16. $\int_0^1 \frac{\log_e x}{1+x} dx = -\frac{\pi^2}{12}$
17. $\int_0^1 \frac{\log_e x}{1-x^2} dx = \frac{\pi^2}{8}$

TABLE 51.19 Haversines^a

θ°	Value	Log	θ°	Value	Log	θ°	Value	Log	θ°	Value	Log
0	0.00000	—	45	0.14645	0.16568	90	0.50000	0.69897	135	0.85355	0.93123
1	0.00008	0.88168	46	0.15267	0.18376	91	0.50873	0.70648	136	0.85967	0.93433
2	0.00030	0.48371	47	0.15900	0.20140	92	0.51745	0.71387	137	0.86568	0.93736
3	0.00069	0.83584	48	0.16543	0.21863	93	0.52617	0.72112	138	0.87157	0.94030
4	0.00122	0.08564	49	0.17197	0.23545	94	0.53488	0.72825	139	0.87735	0.94318
5	0.00190	0.27936	50	0.17861	0.25190	95	0.54358	0.73526	140	0.88302	0.94597
6	0.00274	0.43760	51	0.18534	0.26797	96	0.55226	0.74215	141	0.88857	0.94869
7	0.00373	0.57135	52	0.19217	0.28368	97	0.56093	0.74891	142	0.89401	0.95134
8	0.00487	0.68717	53	0.19909	0.29905	98	0.56959	0.75556	143	0.89932	0.95391
9	0.00616	0.78929	54	0.20611	0.31409	99	0.57822	0.76209	144	0.90451	0.95641
10	0.00760	0.88059	55	0.21321	0.32281	100	0.58682	0.76851	145	0.90958	0.95884
11	0.00919	0.96315	56	0.22040	0.34322	101	0.59540	0.77481	146	0.91452	0.96119
12	0.01093	0.03847	57	0.22768	0.35733	102	0.60396	0.78101	147	0.91934	0.96347
13	0.01281	0.10772	58	0.23504	0.37114	103	0.61248	0.78709	148	0.92402	0.96568
14	0.01485	0.17179	59	0.24248	0.38468	104	0.62096	0.79306	149	0.92858	0.96782
15	0.01704	0.23140	60	0.25000	0.39794	105	0.62941	0.79893	150	0.93301	0.96989
16	0.01937	0.28711	61	0.25760	0.41094	106	0.63782	0.80470	151	0.93731	0.97188
17	0.02185	0.33940	62	0.26526	0.42368	107	0.64619	0.81036	152	0.94147	0.97381
18	0.02447	0.38867	63	0.27300	0.43617	108	0.65451	0.81592	153	0.94550	0.97566
19	0.02724	0.43522	64	0.28081	0.44842	109	0.66278	0.82137	154	0.94940	0.97745
20	0.03015	0.47934	65	0.28869	0.46043	110	0.67101	0.82673	155	0.95315	0.97016
21	0.03321	0.52127	66	0.29663	0.47222	111	0.67918	0.83199	156	0.95677	0.98081
22	0.03641	0.56120	67	0.30463	0.48378	112	0.68730	0.83715	157	0.96025	0.98239
23	0.03975	0.59931	68	0.31270	0.49512	113	0.69537	0.84221	158	0.96359	0.98389
24	0.04323	0.63576	69	0.32082	0.50625	114	0.70337	0.84718	159	0.96679	0.98533
25	0.04685	0.67067	70	0.32899	0.51718	115	0.71131	0.85206	160	0.96985	0.98670
26	0.05060	0.70418	71	0.33722	0.52791	116	0.71919	0.85684	161	0.97276	0.98801
27	0.05450	0.73637	72	0.34549	0.53844	117	0.72700	0.86153	162	0.97553	0.98924
28	0.05853	0.76735	73	0.35381	0.54878	118	0.73474	0.86613	163	0.97815	0.99041
29	0.06269	0.79720	74	0.36218	0.55893	119	0.74240	0.87064	164	0.98063	0.99151
30	0.06699	0.82599	75	0.37059	0.56889	120	0.75000	0.87506	165	0.98296	0.99254
31	0.07142	0.85380	76	0.37904	0.57868	121	0.75752	0.87939	166	0.98515	0.99350
32	0.07598	0.88068	77	0.38752	0.58830	122	0.76496	0.88364	167	0.98719	0.99440
33	0.08066	0.90668	78	0.39604	0.59774	123	0.77232	0.88780	168	0.98907	0.99523
34	0.08548	0.93187	79	0.40460	0.60702	124	0.77960	0.89187	169	0.99081	0.99599
35	0.09042	0.95628	80	0.41318	0.61613	125	0.78679	0.89586	170	0.99240	0.99669
36	0.09549	0.97996	81	0.42178	0.62509	126	0.79389	0.89976	171	0.99384	0.99732
37	0.10068	0.00295	82	0.43041	0.63389	127	0.80091	0.90358	172	0.99513	0.99788
38	0.10599	0.02528	83	0.43907	0.64253	128	0.80783	0.90732	173	0.99627	0.99838
39	0.11143	0.04699	84	0.44774	0.65102	129	0.81466	0.91098	174	0.99726	0.99881
40	0.11698	0.06810	85	0.45642	0.65937	130	0.82139	0.91455	175	0.99810	0.99917
41	0.12265	0.08865	86	0.46512	0.66757	131	0.82803	0.91805	176	0.99878	0.99947
42	0.12843	0.10866	87	0.47383	0.67562	132	0.83457	0.92146	177	0.99931	0.99970
43	0.13432	0.12815	88	0.48255	0.68354	133	0.84100	0.92480	178	0.99970	0.99987
44	0.14033	0.14715	89	0.49127	0.69132	134	0.84733	0.92805	179	0.99992	0.99997
									180	1.00000	0.00000

^ahav $\theta = \frac{1}{2}$ vers $\theta = \frac{1}{2}(1 - \cos \theta) = \sin^2 \frac{1}{2} \theta a$

hav($-\theta$) = hav θ

hav($180^\circ - \theta$) = hav($180^\circ + \theta$) = 1 - hav θ

Characteristics of the logarithms are omitted.

TABLE 51.20 Complete Elliptic Integrals^a

$\sin^{-1}k$	K	$\log K$	E	$\log E$	$\sin^{-1}k$	K	$\log K$	E	$\log E$
0°	1.5708	0.196120	1.5708	0.196120	45	1.8541	0.268127	1.3506	0.130541
1	1.5709	0.196153	1.5707	0.196087	46	1.8691	0.271644	1.3418	0.127690
2	1.5713	0.196252	1.5703	0.195988	47	1.8848	0.275267	1.3329	0.124788
3	1.5719	0.196418	1.5697	0.195822	48	1.9011	0.279001	1.3238	0.121836
4	1.5727	0.196649	1.5689	0.195591	49	1.9180	0.282848	1.3147	0.118836
5°	1.5738	0.196947	1.5678	0.195293	50°	1.9356	0.286811	1.3055	0.115790
6	1.5751	0.197312	1.5665	0.194930	51	1.9539	0.290895	1.2963	0.112698
7	1.5767	0.197743	1.5649	0.194500	52	1.9729	0.295101	1.2870	0.109563
8	1.5785	0.198241	1.5632	0.194004	53	1.9927	0.299435	1.2776	0.106386
9	1.5805	0.198806	1.5611	0.193442	54	2.0133	0.303901	1.2681	0.103169
10	1.5828	0.199438	1.5589	0.192815	55	2.0347	0.308504	1.2587	0.099915
11	1.5854	0.200137	1.5564	0.192121	56	2.0571	0.313247	1.2492	0.096626
12	1.5882	0.200904	1.5537	0.191362	57	2.0804	0.318138	1.2397	0.093303
13	1.5913	0.201740	1.5507	0.190537	58	2.1047	0.323182	1.2301	0.089950
14	1.5946	0.202643	1.5476	0.189646	59	2.1300	0.328384	1.2206	0.086569
15	1.5981	0.203615	1.5442	0.188690	60	2.1565	0.333753	1.2111	0.083164
16	1.6020	0.204657	1.5405	0.187668	61	2.1842	0.339295	1.2015	0.079738
17	1.6061	0.205768	1.5367	0.186581	62	2.2132	0.345020	1.1920	0.076293
18	1.6105	0.206948	1.5326	0.185428	63	2.2435	0.350936	1.1826	0.072834
19	1.6151	0.208200	1.5283	0.184210	64	2.2754	0.357053	1.1732	0.069364
20	1.6200	0.209522	1.5238	0.182928	65	2.3088	0.363384	1.1638	0.065889
21	1.6252	0.210916	1.5191	0.181580	66	2.3439	0.369940	1.1545	0.062412
22	1.6307	0.212382	1.5141	0.180168	67	2.3809	0.376736	1.1453	0.058937
23	1.6365	0.213921	1.5090	0.178691	68	2.4198	0.383787	1.1362	0.055472
24	1.6426	0.215533	1.5037	0.177150	69	2.4610	0.391112	1.1272	0.052020
25	1.6490	0.217219	1.4981	0.175545	70	2.5046	0.398730	1.1184	0.048589
26	1.6557	0.218981	1.4924	0.173876	71	2.5507	0.406665	1.1096	0.045183
27	1.6627	0.220818	1.4864	0.172144	72	2.5998	0.414943	1.1011	0.041812
28	1.6701	0.222732	1.4803	0.170348	73	2.6521	0.423596	1.0927	0.038481
29	1.6777	0.224723	1.4740	0.168489	74	2.7081	0.432660	1.0844	0.035200
30	1.6858	0.226793	1.4675	0.166567	75	2.7681	0.442176	1.0764	0.031976
31	1.6941	0.228943	1.4608	0.164583	76	2.8327	0.452196	1.0686	0.028819
32	1.7028	0.231173	1.4539	0.162537	77	2.9026	0.462782	1.0611	0.025740
33	1.7119	0.233485	1.4469	0.160429	78	2.9786	0.474008	1.0538	0.022749
34	1.7214	0.235880	1.4397	0.158261	79	3.0617	0.485967	1.0468	0.019858
35	1.7312	0.238359	1.4323	0.156031	80	3.1534	0.498777	1.0401	0.017081
36	1.7415	0.240923	1.4248	0.153742	81	3.2553	0.512591	1.0338	0.014432
37	1.7552	0.243575	1.4171	0.151393	82	3.3699	0.527613	1.0278	0.011927
38	1.7633	0.246315	1.4092	0.148985	83	3.5004	9.544120	1.0223	0.009584
39	1.7748	0.249146	1.4013	0.146519	84	3.6519	0.562514	1.0172	0.007422
40	1.7868	0.252068	1.3931	0.143995	85	3.8317	0.583396	1.0127	0.005465
41	1.7992	0.255085	1.3849	0.141414	86	4.0528	0.607751	1.0086	0.003740
42	1.8122	0.258197	1.3765	0.138778	87	4.3387	0.637355	1.0053	0.002278
43	1.8256	0.261406	1.3680	0.136086	88	4.7427	0.676027	1.0026	0.001121
44	1.8396	0.264716	1.3594	0.133340	89	5.4349	0.735192	1.0008	0.000326
					90	∞	∞	1.0000	0.000000

TABLE 51.20 (Continued)

$\sin^{-1}k$	k	K	$\log K$	$\sin^{-1}k$	k	K	$\log K$	$\sin^{-1}k$	k	K	$\log K$
89	20	5.840	0.76641	89	40	6.533	0.81511	89	50	7.226	0.85890
89	22	5.891	0.77019	89	41	6.584	0.81849	89	51	7.332	0.86522
89	24	5.946	0.77422	89	42	6.639	0.82210	89	52	7.449	0.87210
89	26	6.003	0.77837	89	43	6.696	0.82582	89	53	7.583	0.87984
89	28	6.063	0.78269	89	44	6.756	0.82969	89	54	7.737	0.88857
89	30	6.128	0.78732	89	45	6.821	0.83385	89	55	7.919	0.89867
89	32	6.197	0.79218	89	46	6.890	0.83822	89	56	8.143	0.91078
89	34	6.271	0.79734	89	47	6.964	0.84286	89	57	8.430	0.92583
89	36	6.351	0.80284	89	48	7.044	0.84782	89	58	8.836	0.94626
89	38	6.438	0.80875	89	49	7.131	0.85315	89	59	9.529	0.97905
								90	0	∞	∞

TABLE 51.21 Gamma Functions^a

n	$\Gamma(n)$	n	$\Gamma(n)$	n	$\Gamma(n)$	n	$\Gamma(n)$
1.00	1.00000	1.25	0.90640	1.50	0.88623	1.75	0.91906
1.01	0.99433	1.26	0.90440	1.51	0.88659	1.76	0.92137
1.02	0.98884	1.27	0.90250	1.52	0.88704	1.77	0.92376
1.03	0.98355	1.28	0.90072	1.53	0.88757	1.78	0.92623
1.04	0.97844	1.29	0.89904	1.54	0.88818	1.79	0.92877
1.05	0.97350	1.30	0.89747	1.55	0.88887	1.80	0.93138
1.06	0.96874	1.31	0.89600	1.56	0.88964	1.81	0.93408
1.07	0.96415	1.32	0.89464	1.57	0.89049	1.82	0.93685
1.08	0.95973	1.33	0.89338	1.58	0.89142	1.83	0.93969
1.09	0.95546	1.34	0.89222	1.59	0.89243	1.84	0.94261
1.10	0.95135	1.35	0.89115	1.60	0.89352	1.85	0.94561
1.11	0.94739	1.36	0.89018	1.61	0.89468	1.86	0.94869
1.12	0.94359	1.37	0.88931	1.62	0.89592	1.87	0.95184
1.13	0.93993	1.38	0.88854	1.63	0.89724	1.88	0.95507
1.14	0.93642	1.39	0.88785	1.64	0.89864	1.89	0.95838
1.15	0.93304	1.40	0.88726	1.65	0.90012	1.90	0.96177
1.16	0.92980	1.41	0.88676	1.66	0.90167	1.91	0.96523
1.17	0.92670	1.42	0.88636	1.67	0.90330	1.92	0.96878
1.18	0.92373	1.43	0.88604	1.68	0.90500	1.93	0.97240
1.19	0.92088	1.44	0.88580	1.69	0.90678	1.94	0.97610
1.20	0.91817	1.45	0.88565	1.70	0.90864	1.95	0.97988
1.21	0.91558	1.46	0.88560	1.71	0.91057	1.96	0.98374
1.22	0.91311	1.47	0.88563	1.72	0.91258	1.97	0.98768
1.23	0.91075	1.48	0.88575	1.73	0.91466	1.98	0.99171
1.24	0.90852	1.49	0.88595	1.74	0.91683	1.99	0.99581
						2.00	1.00000

^aValues of $\Gamma(n) = \int_0^\infty e^{-x} x^{n-1} dx$; $\Gamma(n+1) = n\Gamma(n)$.

For large positive integers, Stirling's formula gives an approximation in which the relative error decreases as n increases:

$$\Gamma(n+1) = (2\pi n)^{1/2} \left(\frac{n}{e}\right)^n$$

Source: From *CRC Standard Mathematical Tables*, Chemical Rubber Publishing Co., 12th ed., 1959. Used by permission.

TABLE 51.22 Bessel Functions

<i>J</i> ₀ (<i>x</i>) and <i>J</i> ₁ (<i>x</i>) ^{<i>a</i>}								
<i>x</i>	<i>J</i> ₀ (<i>x</i>)	<i>J</i> ₁ (<i>x</i>)	<i>x</i>	<i>J</i> ₀ (<i>x</i>)	<i>J</i> ₁ (<i>x</i>)	<i>x</i>	<i>J</i> ₀ (<i>x</i>)	<i>J</i> ₁ (<i>x</i>)
0.0	1.0000	0.0000	3.0	−0.2601	0.3391	6.0	0.1506	−0.2767
0.1	0.9975	0.0499	3.1	−0.2921	0.3009	6.1	0.1773	−0.2559
0.2	0.9900	0.0995	3.2	−0.3202	0.2613	6.2	0.2017	−0.2329
0.3	0.9776	0.1483	3.3	−0.3443	0.2207	6.3	0.2238	−0.2081
0.4	0.9604	0.1960	3.4	−0.3643	0.1792	6.4	0.2433	−0.1816
0.5	0.9385	0.2423	3.5	−0.3801	0.1374	6.5	0.2601	−0.1538
0.6	0.9120	0.2867	3.6	−0.3918	0.0955	6.6	0.2740	−0.1250
0.7	0.8812	0.3290	3.7	−0.3992	0.0538	6.7	0.2851	−0.0953
0.8	0.8463	0.3668	3.8	−0.4026	0.0128	6.8	0.2931	−0.0652
0.9	0.8075	0.4059	3.9	−0.4018	−0.0272	6.9	0.2981	−0.0349
1.0	0.7652	0.4401	4.0	−0.3971	−0.0660	7.0	0.3001	−0.0047
1.1	0.7196	0.4709	4.1	−0.3887	−0.1033	7.1	0.2991	0.0252
1.2	0.6711	0.4983	4.2	−0.3766	−0.1386	7.2	0.2951	0.0543
1.3	0.6201	0.5220	4.3	−0.3610	−0.1719	7.3	0.2882	0.0826
1.4	0.5669	0.5419	4.4	−0.3423	−0.2028	7.4	0.2786	0.1096
1.5	0.5118	0.5579	4.5	−0.3205	−0.2311	7.5	0.2663	0.1352
1.6	0.4554	0.5699	4.6	−0.2961	−0.2566	7.6	0.2516	0.1592
1.7	0.3980	0.5778	4.7	−0.2693	−0.2791	7.7	0.2346	0.1813
1.8	0.3400	0.5815	4.8	−0.2404	−0.2985	7.8	0.2154	0.2014
1.9	0.2818	0.5812	4.9	−0.2097	−0.3147	7.9	0.1944	0.2192
2.0	0.2239	0.5767	5.0	−0.1776	−0.3276	8.0	0.1717	0.2346
2.1	0.1666	0.5683	5.1	−0.1443	−0.3371	8.1	0.1475	0.2476
2.2	0.1104	0.5560	5.2	−0.1103	−0.3432	8.2	0.1222	0.2580
2.3	0.0555	0.5399	5.3	−0.0758	−0.3460	8.3	0.0960	0.2657
2.4	0.0025	0.5202	5.4	−0.0412	−0.3453	8.4	0.0692	0.2708
2.5	−0.0484	0.4971	5.5	−0.0068	−0.3414	8.5	0.0419	0.2731
2.6	−0.0968	0.4708	5.6	0.0270	−0.3343	8.6	0.0146	0.2728
2.7	−0.1424	0.4416	5.7	0.0599	−0.3241	8.7	−0.0125	0.2697
2.8	−0.1850	0.4097	5.8	0.0917	−0.3110	8.8	−0.0392	0.2641
2.9	−0.2243	0.3754	5.9	0.1220	−0.2951	8.9	−0.0653	0.2559
<i>Y</i> ₀ (<i>x</i>) and <i>Y</i> ₁ (<i>x</i>)								
<i>x</i>	<i>Y</i> ₀ (<i>x</i>)	<i>Y</i> ₁ (<i>x</i>)	<i>x</i>	<i>Y</i> ₀ (<i>x</i>)	<i>Y</i> ₁ (<i>x</i>)	<i>x</i>	<i>Y</i> ₀ (<i>x</i>)	<i>Y</i> ₁ (<i>x</i>)
0.0	(−∞)	(−∞)	2.5	0.498	0.146	5.0	−0.309	0.148
0.5	−0.445	−1.471	3.0	0.377	0.325	5.5	−0.340	−0.024
1.0	0.088	−0.781	3.5	0.189	0.410	6.0	−0.288	−0.175
1.5	0.382	−0.412	4.0	−0.017	0.398	6.5	−0.173	−0.274
2.0	0.510	−0.107	4.5	−0.195	0.301	7.0	−0.026	−0.303

^{*a*}*J*₁(*x*) = 0 for *x* = 0, 3.832, 7.016, 10.173, 13.324, . . .
*J*₀(*x*) = 0 for *x* = 2.405, 5.520, 8.654, 11.792, . . .

51.3 STATISTICAL TABLES¹

TABLE 51.23 Binomial Coefficients

n	$\binom{n}{0}$	$\binom{n}{1}$	$\binom{n}{2}$	$\binom{n}{3}$	$\binom{n}{4}$	$\binom{n}{5}$	$\binom{n}{6}$	$\binom{n}{7}$	$\binom{n}{8}$	$\binom{n}{9}$	$\binom{n}{10}$
0	1										
1	1	1									
2	1	2	1								
3	1	3	3	1							
4	1	4	6	4	1						
5	1	5	10	10	5	1					
6	1	6	15	20	15	6	1				
7	1	7	21	35	35	21	7	1			
8	1	8	28	56	70	56	28	8	1		
9	1	9	36	84	126	126	84	36	9	1	
10	1	10	45	120	210	252	210	120	45	10	1
11	1	11	55	165	330	462	462	330	165	55	11
12	1	12	66	220	495	792	924	792	495	220	66
13	1	13	78	286	715	1287	1716	1716	1287	715	286
14	1	14	91	364	1001	2002	3003	3432	3003	2002	1001
15	1	15	105	455	1365	3003	5005	6435	6435	5005	3003
16	1	16	120	560	1820	4368	8008	11440	12870	11440	8008
17	1	17	136	680	2380	6188	12376	19448	24310	24310	19448
18	1	18	153	816	3060	8568	18564	31824	43758	48620	43758
19	1	19	171	969	3876	11628	27132	50388	75582	92378	92378
20	1	20	190	1140	4845	15504	38760	77520	125970	167960	184756

$$nC_m = \binom{n}{m} = \frac{n!}{[(n-m)!m!]} = \binom{n}{n-m} \cdot \binom{n}{0} = 1.$$

$$(p+q)^n = p^n + \binom{n}{1}p^{n-1}q + \cdots + \binom{n}{s}p^s q^t + \cdots + q^n, s+t=n.$$

Probability Let p be the probability of an event e in one trial and q the probability of failure of e . The probability that, in n trials, the event e will occur exactly $n-t$ times is $\binom{n}{t}p^{n-t}q^t$. The probability that an event e will happen at least r times in n trials is $\sum_{t=0}^{n-r} \binom{n}{t}p^{n-t}q^t$; at most r times in n trials is $\sum_{t=n-r}^t \binom{n}{t}p^{n-t}q^t$.

In a *point binomial*, $(p+q)^n$, distribution, the mean number of favorable events is np ; the mean number of unfavorable events is nq ; the *standard deviation* is $\sigma = \sqrt{pqn}$; and $a_3 = (p-q)/\sigma$. The *mean deviation from the mean* MD is $\sigma\sqrt{2/\pi} = 0.7979\sigma$; the *semiquartile deviation from the mean* is $0.6745\sigma = 0.845$ MD.

The probability that a deviation of an individual measure from the average lies between $y = -a$ and $y = a$ is

$$\frac{1}{\sqrt{\pi}} \int_{y=-a}^{y=a} h e^{-h^2 y^2} dy = \frac{1}{\sigma\sqrt{2\pi}} \int_{y=-a}^{y=a} e^{-y^2/2\sigma^2} dy = \frac{1}{\sqrt{2\pi}} \int_{x=-b}^{x=b} e^{-x^2/2} dx$$

where $x = hy\sqrt{2}$, $b = ha\sqrt{2}$, and $\sigma = 1/h\sqrt{2}$; h is called the *modulus of precision* and σ the *standard (quadratic mean) deviation*.

¹ Tables 51.23–51.25 from Burington, *Handbook of Math Tables and Formulas*, published by McGraw-Hill.

TABLE 51.24 Probability Functions

$\frac{1}{2}(1 + \alpha) = \int_{-\infty}^x \Phi(x) dx = \text{area under } \Phi(x) \text{ from } -\infty \text{ to } x$											
$\alpha = \int_{-x}^x \Phi(x)dx, \qquad \Phi(x) = \frac{1}{\sqrt{2\pi}}e^{-x^2/2} = \text{normal function}$											
$\Phi^{(2)}(x) = (x^2 - 1) \qquad \Phi(x) = \text{second derivative of } \Phi(x)$											
$\Phi^{(3)}(x) = (3x - x^3) \qquad \Phi(x) = \text{third derivative of } \Phi(x)$											
$\Phi^{(4)}(x) = (x^4 - 6x^2 + 3) \qquad \Phi(x) = \text{fourth derivative of } \Phi(x)$											
x	$\frac{1}{2}(1 + \alpha)$	$\Phi(x)$	$\Phi^{(2)}(x)$	$\Phi^{(3)}(x)$	$\Phi^{(4)}(x)$	x	$\frac{1}{2}(1 + \alpha)$	$\Phi(x)$	$\Phi^{(2)}(x)$	$\Phi^{(3)}(x)$	$\Phi^{(4)}(x)$
0.00	0.5000	0.3989	−0.3989	0.0000	1.1968	0.35	0.6368	0.3752	−0.3293	0.3779	0.8556
0.01	0.5040	0.3989	−0.3989	0.0120	1.1965	0.36	0.6406	0.3739	−0.3255	0.3864	0.8373
0.02	0.5080	0.3989	−0.3987	0.0239	1.1956	0.37	0.6443	0.3726	−0.3216	0.3947	0.8186
0.03	0.5120	0.3988	−0.3984	0.0359	1.1941	0.38	0.6480	0.3712	−0.3176	0.4028	0.7996
0.04	0.5160	0.3986	−0.3980	0.0478	1.1920	0.39	0.6517	0.3697	−0.3135	0.4107	0.7803
0.05	0.5199	0.3984	−0.3975	0.0597	1.1894	0.40	0.6554	0.3683	−0.3094	0.4184	0.7607
0.06	0.5239	0.3982	−0.3968	0.0716	1.1861	0.41	0.6591	0.3668	−0.3059	0.4259	0.7408
0.07	0.5279	0.3980	−0.3960	0.0834	1.1822	0.42	0.6628	0.3653	−0.3008	0.4332	0.7206
0.08	0.5319	0.3977	−0.3951	0.0952	1.1778	0.43	0.6664	0.3637	−0.2965	0.4403	0.7001
0.09	0.5359	0.3973	−0.3941	0.1070	1.1727	0.44	0.6700	0.3621	−0.2920	0.4472	0.6793
0.10	0.5398	0.3970	−0.3930	0.1187	1.1671	0.45	0.6736	0.3605	−0.2875	0.4539	0.6583
0.11	0.5438	0.3965	−0.3917	0.1303	1.1609	0.46	0.6772	0.3589	−0.2830	0.4603	0.6371
0.12	0.5478	0.3961	−0.3904	0.1419	1.1541	0.47	0.6808	0.3572	−0.2783	0.4666	0.6156
0.13	0.5517	0.3956	−0.3889	0.1534	1.1468	0.48	0.6844	0.3555	−0.2736	0.4727	0.5940
0.14	0.5557	0.3951	−0.3873	0.1648	1.1389	0.49	0.6879	0.3538	−0.2689	0.4785	0.5721
0.15	0.5596	0.3945	−0.3856	0.1762	1.1304	0.50	0.6915	0.3521	−0.2641	0.4841	0.5501
0.16	0.5636	0.3939	−0.3838	0.1874	1.1214	0.51	0.6950	0.3503	−0.2592	0.4895	0.5279
0.17	0.5675	0.3932	−0.3819	0.1986	1.1118	0.52	0.6985	0.3485	−0.2543	0.4947	0.5056
0.18	0.5714	0.3925	−0.3798	0.2097	1.1017	0.53	0.7019	0.3467	−0.2493	0.4996	0.4831
0.19	0.5753	0.3918	−0.3777	0.2206	1.0911	0.54	0.7054	0.3448	−0.2443	0.5043	0.4605
0.20	0.5793	0.3910	−0.3754	0.2315	1.0799	0.55	0.7088	0.3429	−0.2392	0.5088	0.4378
0.21	0.5832	0.3902	−0.3730	0.2422	1.0682	0.56	0.7123	0.3410	−0.2341	0.5131	0.4150
0.22	0.5871	0.3894	−0.3706	0.2529	1.0560	0.57	0.7157	0.3391	−0.2289	0.5171	0.3921
0.23	0.5910	0.3885	−0.3680	0.2634	1.0434	0.58	0.7190	0.3372	−0.2238	0.5209	0.3691
0.24	0.5948	0.3876	−0.3653	0.2737	1.0302	0.59	0.7224	0.3352	−0.2185	0.5245	0.3461
0.25	0.5987	0.3867	−0.3625	0.2840	1.0165	0.60	0.7257	0.3332	−0.2133	0.5278	0.3231
0.26	0.6026	0.3857	−0.3596	0.2941	1.0024	0.61	0.7291	0.3312	−0.2080	0.5309	0.3000
0.27	0.6064	0.3847	−0.3566	0.3040	0.9878	0.62	0.7324	0.3292	−0.2027	0.5338	0.2770
0.28	0.6103	0.3836	−0.3535	0.3138	0.9727	0.63	0.7357	0.3271	−0.1973	0.5365	0.2539
0.29	0.6141	0.3825	−0.3504	0.3235	0.9572	0.64	0.7389	0.3251	−0.1919	0.5389	0.2309
0.30	0.6179	0.3814	−0.3471	0.3330	0.9413	0.65	0.7422	0.3230	−0.1865	0.5411	0.2078
0.31	0.6217	0.3802	−0.3437	0.3423	0.9250	0.66	0.7454	0.3209	−0.1811	0.5431	0.1849
0.32	0.6255	0.3790	−0.3402	0.3515	0.9082	0.67	0.7486	0.3187	−0.1757	0.5448	0.1620
0.33	0.6293	0.3778	−0.3367	0.3605	0.8910	0.68	0.7517	0.3166	−0.1702	0.5463	0.1391
0.34	0.6331	0.3765	−0.3330	0.3693	0.8735	0.69	0.7549	0.3144	−0.1647	0.5476	0.1164

TABLE 51.24 (Continued)

x	$\frac{1}{2}(1 + \alpha)$	$\Phi(x)$	$\Phi^{(2)}(x)$	$\Phi^{(3)}(x)$	$\Phi^{(4)}(x)$	x	$\frac{1}{2}(1 + \alpha)$	$\Phi(x)$	$\Phi^{(2)}(x)$	$\Phi^{(3)}(x)$	$\Phi^{(4)}(x)$
0.70	0.7580	0.3123	−0.1593	0.5486	0.0937	1.15	0.8749	0.2059	0.0664	0.3973	−0.6561
0.71	0.7611	0.3101	−0.1538	0.5495	0.0712	1.16	0.8770	0.2036	0.0704	0.3907	−0.6643
0.72	0.7642	0.3079	−0.1483	0.5501	0.0487	1.17	0.8790	0.2012	0.0742	0.3840	−0.6720
0.73	0.7673	0.3056	−0.1428	0.5504	0.0265	1.18	0.8810	0.1989	0.0780	0.3772	−0.6792
0.74	0.7704	0.3034	−0.1373	0.5506	0.0043	1.19	0.8830	0.1965	0.0818	0.3704	−0.6861
0.75	0.7734	0.3011	−0.1318	0.5505	−0.0176	1.20	0.8849	0.1942	0.0854	0.3635	−0.6926
0.76	0.7764	0.2989	−0.1262	0.5502	−0.0394	1.21	0.8869	0.1919	0.0890	0.3566	−0.6986
0.77	0.7794	0.2966	−0.1207	0.5497	−0.0611	1.22	0.8888	0.1919	0.0890	0.3566	−0.6986
0.78	0.7823	0.2943	−0.1153	0.5490	−0.0825	1.23	0.8907	0.1872	0.0960	0.3425	−0.7094
0.79	0.7852	0.2920	−0.1098	0.5481	−0.1037	1.24	0.8925	0.1849	0.0994	0.3354	−0.7141
0.80	0.7881	0.2897	−0.1043	0.5469	−0.1247	1.25	0.8944	0.1826	0.1027	0.3282	−0.7185
0.81	0.7910	0.2874	−0.0988	0.5456	−0.1455	1.26	0.8962	0.1804	0.1060	0.3210	−0.7224
0.82	0.7939	0.2850	−0.0934	0.5440	−0.1660	1.27	0.8980	0.1781	0.1092	0.3138	−0.7259
0.83	0.7967	0.2827	−0.0880	0.5423	−0.1862	1.28	0.8997	0.1758	0.1123	0.3065	−0.7291
0.84	0.7995	0.2803	−0.0825	0.5403	−0.2063	1.29	0.9015	0.1736	0.1153	0.2992	−0.7318
0.85	0.8023	0.2780	−0.0771	0.5381	−0.2260	1.30	0.9032	0.1714	0.1182	0.2918	−0.7341
0.86	0.8051	0.2756	−0.0718	0.5358	−0.2455	1.31	0.9049	0.1691	0.1211	0.2845	−0.7361
0.87	0.8078	0.2732	−0.0664	0.5332	−0.2646	1.32	0.9066	0.1669	0.1239	0.2771	−0.7376
0.88	0.8106	0.2709	−0.0611	0.5305	−0.2835	1.33	0.9082	0.1647	0.1267	0.2697	−0.7388
0.89	0.8133	0.2685	−0.0558	0.5276	−0.3021	1.34	0.9099	0.1626	0.1293	0.2624	−0.7395
0.90	0.8159	0.2661	−0.0506	0.5245	−0.3203	1.35	0.9115	0.1604	0.1319	0.2550	−0.7399
0.91	0.8186	0.2637	−0.0453	0.5212	−0.3383	1.36	0.9131	0.1582	0.1344	0.2476	−0.7400
0.92	0.8212	0.2613	−0.0401	0.5177	−0.3559	1.37	0.9147	0.1561	0.1369	0.2402	−0.7396
0.93	0.8238	0.2589	−0.0350	0.5140	−0.3731	1.38	0.9162	0.1539	0.1392	0.2328	−0.7389
0.94	0.8264	0.2565	−0.0299	0.5102	−0.3901	1.39	0.9177	0.1518	0.1415	0.2254	−0.7378
0.95	0.8289	0.2541	−0.0248	0.5062	−0.4066	1.40	0.9192	0.1497	0.1437	0.2180	−0.7364
0.96	0.8315	0.2516	−0.0197	−0.521	−0.4228	1.41	0.9207	0.1476	0.1459	0.2107	−0.7347
0.97	0.8340	0.2492	−0.0147	0.4978	−0.4387	1.42	0.9222	0.1456	0.1480	0.2033	−0.7326
0.98	0.8365	0.2468	−0.0098	0.4933	−0.4541	1.43	0.9236	0.1435	0.1500	0.1960	−0.7301
0.99	0.8389	0.2444	−0.0049	0.4887	−0.4692	1.44	0.9251	0.1415	0.1519	0.1887	−0.7274
1.00	0.8413	0.2420	0.0000	0.4839	−0.4839	1.45	0.9265	0.1394	0.1537	0.1815	−0.7243
1.01	0.8438	0.2396	0.0048	0.4790	−0.4983	1.46	0.9279	0.1374	0.1555	0.1742	−0.7209
1.02	0.8461	0.2371	0.0096	0.4740	−0.5122	1.47	0.9292	0.1354	0.1572	0.1670	−0.7172
1.03	0.8485	0.2347	0.0143	0.4688	−0.5257	1.48	0.9306	0.1344	0.1588	0.1599	−0.7132
1.04	0.8508	0.2323	0.0190	0.4635	−0.5389	1.49	0.9319	0.1315	0.1604	0.1528	−0.7089
1.05	0.8531	0.2299	0.0236	0.4580	−0.5516	1.50	0.9332	0.1295	0.1619	0.1457	−0.7043
1.06	0.8554	0.2275	0.0281	0.4524	−0.5639	1.51	0.9345	0.1276	0.1633	0.1387	−0.6994
1.07	0.8577	0.2251	0.0326	0.4467	−0.5758	1.52	0.9357	0.1257	0.1647	0.1317	−0.6942
1.08	0.8599	0.2227	0.0371	0.4409	−0.5873	1.53	0.9370	0.1238	0.1660	0.1248	−0.6888
1.09	0.8621	0.2203	0.0414	0.4350	−0.5984	1.54	0.9382	0.1219	0.1672	0.1180	−0.6831
1.10	0.8643	0.2179	0.0458	0.4290	−0.6091	1.55	0.9394	0.1200	0.1683	0.1111	−0.6772
1.11	0.8665	0.2155	0.0500	0.4228	−0.6193	1.56	0.9406	0.1182	0.1694	0.1044	−0.6710
1.12	0.8686	0.2131	0.0542	0.4166	−0.6292	1.57	0.9418	0.1163	0.1704	0.0977	−0.6646
1.13	0.8708	0.2107	0.0583	0.4102	−0.6386	1.58	0.9429	0.1145	0.1714	0.0911	−0.6580
1.14	0.8729	0.2083	0.0624	0.4038	−0.6476	1.59	0.9441	0.1127	0.1722	0.0846	−0.6511

(continued)

TABLE 51.24 (Continued)

x	$\frac{1}{2}(1+\alpha)$	$\Phi(x)$	$\Phi^{(2)}(x)$	$\Phi^{(3)}(x)$	$\Phi^{(4)}(x)$	x	$\frac{1}{2}(1+\alpha)$	$\Phi(x)$	$\Phi^{(2)}(x)$	$\Phi^{(3)}(x)$	$\Phi^{(4)}(x)$
1.60	0.9452	0.1109	0.1730	0.0781	-0.6441	2.05	0.9798	0.0468	0.1563	-0.1203	-0.2222
1.61	0.9463	0.1092	0.1738	0.0717	-0.6368	2.06	0.9803	0.0478	0.1550	-0.1225	-0.2129
1.62	0.9474	0.1074	0.1745	0.0654	-0.6293	2.07	0.9808	0.0468	0.1538	-0.1245	-0.2036
1.63	0.9484	0.1057	0.1751	0.0591	-0.6216	2.08	0.9812	0.0459	0.1526	-0.1265	-0.1945
1.64	0.9495	0.1040	0.1757	0.0529	-0.6138	2.09	0.9817	0.0449	0.1513	-0.1284	-0.1854
1.65	0.9505	0.1023	0.1762	0.0468	-0.6057	2.10	0.9821	0.0440	0.1500	-0.1302	-0.1765
1.66	0.9515	0.1006	0.1766	0.0408	-0.5975	2.11	0.9821	0.0440	0.1500	-0.1302	-0.1765
1.67	0.9525	0.0989	0.1770	0.0349	-0.5891	2.12	0.9830	0.0422	0.1474	-0.1336	-0.1588
1.68	0.9535	0.0973	0.1773	0.0290	-0.5806	2.13	0.9834	0.0413	0.1460	-0.1351	-0.1502
1.69	0.9545	0.0957	0.1776	0.0233	-0.5720	2.14	0.9838	0.0404	0.1446	-0.1366	-0.1416
1.70	0.9554	0.0940	0.1778	0.0176	-0.5632	2.15	0.9842	0.0395	0.1433	-0.1380	-0.1332
1.71	0.9564	0.0925	0.1779	0.0120	-0.5542	2.16	0.9846	0.0387	0.1419	-0.1393	-0.1249
1.72	0.9573	0.0909	0.1780	0.0065	-0.5452	2.17	0.9850	0.0379	0.1405	-0.1405	-0.1167
1.73	0.9582	0.0893	0.1780	0.0011	-0.5360	2.18	0.9854	0.0371	0.1391	-0.1416	-0.1086
1.74	0.9591	0.0878	0.1780	-0.0042	-0.5267	2.19	0.9857	0.0363	0.1377	-0.1426	-0.1006
1.75	0.9599	0.0863	0.1780	-0.0094	-0.5173	2.20	0.9861	0.0355	0.1362	-0.1436	-0.0927
1.76	0.9608	0.0848	0.1778	-0.0146	-0.5079	2.21	0.9864	0.0347	0.1348	-0.1445	-0.0850
1.77	0.9616	0.0833	0.1777	-0.0196	-0.4983	2.22	0.9868	0.0339	0.1333	-0.1453	-0.0774
1.78	0.9625	0.0818	0.1774	-0.0245	-0.4887	2.23	0.9871	0.0332	0.1319	-0.1460	-0.0700
1.79	0.9633	0.0804	0.1772	-0.0294	-0.4789	2.24	0.9875	0.0325	0.1304	-0.1467	-0.0626
1.80	0.9641	0.0790	0.1769	-0.0341	-0.4692	2.25	0.9878	0.0317	0.1289	-0.1473	-0.0554
1.81	0.9649	0.0775	0.1765	-0.0388	-0.4593	2.26	0.9881	0.0310	0.1275	-0.1478	-0.0484
1.82	0.9656	0.0761	0.1761	-0.0433	-0.4494	2.27	0.9884	0.0303	0.1260	-0.1483	-0.0414
1.83	0.9664	0.0748	0.1756	-0.0477	-0.4395	2.28	0.9887	0.0297	0.1245	-0.1486	-0.0346
1.84	0.9671	0.0734	0.1751	-0.0521	-0.4295	2.29	0.9890	0.0290	0.1230	-0.1490	-0.0279
1.85	0.9678	0.0721	0.1746	-0.0563	-0.4195	2.30	0.9893	0.0283	0.1215	-0.1492	-0.0214
1.86	0.9686	0.0707	0.1740	-0.0605	-0.4095	2.31	0.9896	0.0277	0.1200	-0.1494	-0.0150
1.87	0.9693	0.0694	0.1734	-0.0645	-0.3995	2.32	0.9898	0.0270	0.1185	-0.1495	-0.0088
1.88	0.9699	0.0681	0.1727	-0.0685	-0.3894	2.33	0.9901	0.0264	0.1170	-0.1496	-0.0027
1.89	0.9706	0.0669	0.1720	-0.0723	-0.3793	2.34	0.9904	0.0258	0.1155	-0.1496	0.0033
1.90	0.9713	0.0656	0.1713	-0.0761	-0.3693	2.35	0.9906	0.0252	0.1141	-0.1495	0.0092
1.91	0.9719	0.0644	0.1705	-0.0797	-0.3592	2.36	0.9909	0.0246	0.1126	-0.1494	0.0149
1.92	0.9726	0.0632	0.1697	-0.0832	-0.3492	2.37	0.9911	0.0241	0.1111	-0.1492	0.0204
1.93	0.9732	0.0620	0.1688	-0.0867	-0.3392	2.38	0.9913	0.0235	0.1096	-0.1490	0.0258
1.94	0.9738	0.0608	0.1679	-0.0900	-0.3292	2.39	0.9916	0.0229	0.1081	-0.1487	0.0311
1.95	0.9744	0.0596	0.1670	-0.0933	-0.3192	2.40	0.9918	0.0224	0.1066	-0.1483	0.0362
1.96	0.9750	0.0584	0.1661	-0.0964	-0.3093	2.41	0.9920	0.0219	0.1051	-0.1480	0.0412
1.97	0.9756	0.0573	0.1651	-0.0994	-0.2994	2.42	0.9922	0.0213	0.1036	-0.1475	0.0461
1.98	0.9761	0.0562	0.1641	-0.1024	-0.2895	2.43	0.9925	0.0208	0.1022	-0.1470	0.0508
1.99	0.9767	0.0551	0.1630	-0.1052	-0.2797	2.44	0.9927	0.0203	0.1007	-0.1465	0.0554
2.00	0.9772	0.0540	0.1620	-0.1080	-0.2700	2.45	0.9929	0.0198	0.0992	-0.1459	0.0598
2.01	0.9778	0.0529	0.1609	-0.1106	-0.2603	2.46	0.9931	0.0194	0.0978	-0.1453	0.0641
2.02	0.9783	0.0519	0.1598	-0.1132	-0.2506	2.47	0.9932	0.0189	0.0963	-0.1446	0.0683
2.03	0.9788	0.0508	0.1586	-0.1157	-0.2411	2.48	0.9934	0.0184	0.0949	-0.1439	0.0723
2.04	0.9793	0.0498	0.1575	-0.1180	-0.2316	2.49	0.9936	0.0180	0.0935	-0.1432	0.0762

TABLE 51.24 (Continued)

x	$\frac{1}{2}(1 + \alpha)$	$\Phi(x)$	$\Phi^{(2)}(x)$	$\Phi^{(3)}(x)$	$\Phi^{(4)}(x)$	x	$\frac{1}{2}(1 + \alpha)$	$\Phi(x)$	$\Phi^{(2)}(x)$	$\Phi^{(3)}(x)$	$\Phi^{(4)}(x)$
2.50	0.9938	0.0175	0.0920	-0.1424	0.0800	2.95	0.9984	0.0051	0.0396	-0.0865	0.1364
2.51	0.9940	0.0171	0.0906	-0.1416	0.0836	2.96	0.9985	0.0050	0.0388	-0.0852	0.1358
2.52	0.9941	0.0167	0.0892	-0.1408	0.0871	2.97	0.9985	0.0048	0.0379	-0.0838	0.1352
2.53	0.9943	0.0163	0.0878	-0.1399	0.0905	2.98	0.9986	0.0047	0.0371	-0.0825	0.1345
2.54	0.9945	0.0158	0.0864	-0.1389	0.0937	2.99	0.9986	0.0046	0.0363	-0.0811	0.1337
2.55	0.9946	0.0154	0.0850	-0.1380	0.0968	3.00	0.9987	0.0044	0.0355	-0.0798	0.1330
2.56	0.9948	0.0151	0.0836	-0.1370	0.0998	3.01	0.9987	0.0043	0.0347	-0.0785	0.1321
2.57	0.9949	0.0147	0.0823	-0.1360	0.1027	3.02	0.9987	0.0042	0.0339	-0.0771	0.1313
2.58	0.9951	0.0143	0.0809	-0.1350	0.1054	3.03	0.9988	0.0040	0.0331	-0.0758	0.1304
2.59	0.9952	0.0319	0.0796	-0.1339	0.1080	3.04	0.9988	0.0039	0.0324	-0.0745	0.1294
2.60	0.9953	0.0136	0.0782	-0.1328	0.1105	3.05	0.9989	0.0038	0.0316	-0.0732	0.1285
2.60	0.9953	0.0136	0.0782	-0.1328	0.1105	3.06	0.9989	0.0037	0.0309	-0.0720	0.1275
2.62	0.9956	0.0129	0.0756	-0.1305	0.1152	3.07	0.9989	0.0036	0.0302	-0.0707	0.1264
2.63	0.9957	0.0126	0.0743	-0.1294	0.1173	3.08	0.9990	0.0035	0.0295	-0.0694	0.1254
2.64	0.9959	0.0122	0.0730	-0.1282	0.1194	3.09	0.9990	0.0034	0.0288	-0.0682	0.1243
2.65	0.9960	0.0119	0.0717	-0.1270	0.1213	3.10	0.9990	0.0033	0.0281	-0.0669	0.1231
2.66	0.9961	0.0116	0.0705	-0.1258	0.1231	3.11	0.9991	0.0032	0.0275	-0.0657	0.1220
2.67	0.9962	0.0113	0.0692	-0.1245	0.1248	3.12	0.9991	0.0031	0.0268	-0.0645	0.1208
2.68	0.9963	0.0110	0.0680	-0.1233	0.1264	3.13	0.9991	0.0030	0.0262	-0.0633	0.1196
2.69	0.9964	0.0107	0.0668	-0.1220	0.1279	3.14	0.9992	0.0029	0.0256	-0.0621	0.1184
2.70	0.9965	0.0104	0.0656	-0.1207	0.1293	3.15	0.9992	0.0028	0.0249	-0.0609	0.1171
2.71	0.9966	0.0101	0.0644	-0.1194	0.1306	3.16	0.9992	0.0027	0.0243	-0.0598	0.1159
2.72	0.9967	0.0099	0.0632	-0.1181	0.1317	3.17	0.9992	0.0026	0.0237	-0.0586	0.1146
2.73	0.9968	0.0096	0.0620	-0.1168	0.1328	3.18	0.9993	0.0025	0.0232	-0.0575	0.1133
2.74	0.9969	0.0093	0.0608	-0.1154	0.1338	3.19	0.9993	0.0025	0.0226	-0.0564	0.1120
2.75	0.9970	0.0091	0.0597	-0.1141	0.1347	3.20	0.9993	0.0024	0.0220	-0.0552	0.1107
2.76	0.9971	0.0088	0.0585	-0.1127	0.1356	3.21	0.9993	0.0023	0.0215	-0.0541	0.1093
2.77	0.9972	0.0086	0.0574	-0.1114	0.1363	3.22	0.9994	0.0022	0.0210	-0.0531	0.1080
2.78	0.9973	0.0084	0.0563	-0.1100	0.1369	3.23	0.9994	0.0022	0.0204	-0.0520	0.1066
2.79	0.9974	0.0081	0.0552	-0.1087	0.1375	3.24	0.9994	0.0021	0.0199	-0.0509	0.1053
2.80	0.9974	0.0079	0.0541	-0.1073	0.1379	3.25	0.9994	0.0020	0.0194	-0.0499	0.1039
2.81	0.9975	0.0077	0.0531	-0.1059	0.1383	3.26	0.9994	0.0020	0.0189	-0.0488	0.1025
2.82	0.9976	0.0075	0.0520	-0.1045	0.1386	3.27	0.9995	0.0019	0.0184	-0.0478	0.1011
2.83	0.9977	0.0073	0.0510	-0.1031	0.1389	3.28	0.9995	0.0018	0.0180	-0.0468	0.0997
2.84	0.9977	0.0071	0.0500	-0.1017	0.1390	3.29	0.9995	0.0018	0.0175	-0.0458	0.0983
2.85	0.9978	0.0069	0.0490	-0.1003	0.1391	3.30	0.9995	0.0017	0.0170	-0.0449	0.0969
2.86	0.9979	0.0067	0.0480	-0.0990	0.1391	3.31	0.9995	0.0017	0.0166	-0.0439	0.0955
2.87	0.9979	0.0065	0.0470	-0.0976	0.1391	3.32	0.9996	0.0016	0.0162	-0.0429	0.0941
2.88	0.9980	0.0063	0.0460	-0.0962	0.1389	3.33	0.9996	0.0016	0.0157	-0.0420	0.0927
2.89	0.9981	0.0061	0.0451	-0.0948	0.1388	3.34	0.9996	0.0015	0.0153	-0.0411	0.0913
2.90	0.9981	0.0060	0.0441	-0.0934	0.1385	3.35	0.9996	0.0015	0.0149	-0.0402	0.0899
2.91	0.9982	0.0058	0.0432	-0.0920	0.1382	3.36	0.9996	0.0014	0.0145	-0.0393	0.0885
2.92	0.9982	0.0056	0.0423	-0.0906	0.1378	3.37	0.9996	0.0014	0.0141	-0.0384	0.0871
2.93	0.9983	0.0055	0.0414	-0.0893	0.1374	3.38	0.9996	0.0013	0.0138	-0.0376	0.0857
2.94	0.9984	0.0053	0.0405	-0.0879	0.1369	3.39	0.9997	0.0013	0.0134	-0.0367	0.0843

(continued)

TABLE 51.24 (Continued)

x	$\frac{1}{2}(1+\alpha)$	$\Phi(x)$	$\Phi^{(2)}(x)$	$\Phi^{(3)}(x)$	$\Phi^{(4)}(x)$	x	$\frac{1}{2}(1+\alpha)$	$\Phi(x)$	$\Phi^{(2)}(x)$	$\Phi^{(3)}(x)$	$\Phi^{(4)}(x)$
3.40	0.9997	0.0012	0.0130	-0.0359	0.0829	3.80	0.9999	0.0003	0.0039	-0.0127	0.0365
3.41	0.9997	0.0012	0.0127	-0.0350	0.0815	3.81	0.9999	0.0003	0.0038	-0.0123	0.0356
3.42	0.9997	0.0012	0.0123	-0.0342	0.0801	3.82	0.9999	0.0003	0.0037	-0.0120	0.0347
3.43	0.9997	0.0011	0.0120	-0.0334	0.0788	3.83	0.9999	0.0003	0.0036	-0.0116	0.0339
3.44	0.9997	0.0011	0.0116	-0.0327	0.0774	3.84	0.9999	0.0003	0.0034	-0.0113	0.0331
3.45	0.9997	0.0010	0.0113	-0.0319	0.0761	3.85	0.9999	0.0002	0.0033	-0.0110	0.0323
3.46	0.9997	0.0010	0.0110	-0.0311	0.0747	3.86	0.9999	0.0002	0.0032	-0.0107	0.0315
3.47	0.9997	0.0010	0.0107	-0.0304	0.0734	3.87	1.0000	0.0002	0.0031	-0.0104	0.0307
3.48	0.9998	0.0009	0.0104	-0.0297	0.0721	3.88	1.0000	0.0002	0.0030	-0.0100	0.0299
3.49	0.9998	0.0009	0.0101	-0.0290	0.0707	3.89	1.0000	0.0002	0.0029	-0.0098	0.0292
3.50	0.9998	0.0009	0.0098	-0.0283	0.0694	3.90	1.0000	0.0002	0.0028	-0.0095	0.0284
3.51	0.9998	0.0008	0.0095	-0.0276	0.0681	3.91	1.0000	0.0002	0.0027	-0.0092	0.0277
3.52	0.9998	0.0008	0.0093	-0.0269	0.0669	3.92	1.0000	0.0002	0.0026	-0.0089	0.0270
3.53	0.9998	0.0008	0.0090	-0.0262	0.0656	3.93	1.0000	0.0002	0.0026	-0.0086	0.0263
3.54	0.9998	0.0008	0.0087	-0.0256	0.0643	3.94	1.0000	0.0002	0.0025	-0.0084	0.0256
3.55	0.9998	0.0007	0.0085	-0.0249	0.0631	3.95	1.0000	0.0002	0.0024	-0.0081	0.0250
3.56	0.9998	0.0007	0.0082	-0.0243	0.0618	3.96	1.0000	0.0002	0.0023	-0.0079	0.0243
3.57	0.9998	0.0007	0.0080	-0.0237	0.0606	3.97	1.0000	0.0002	0.0022	-0.0076	0.0237
3.58	0.9998	0.0007	0.0078	-0.0231	0.0594	3.98	1.0000	0.0001	0.0022	-0.0074	0.0230
3.59	0.9998	0.0006	0.0075	-0.0225	0.0582	3.99	1.0000	0.0001	0.0021	-0.0072	0.0224
3.60	0.9998	0.0006	0.0073	-0.0219	0.0570	4.00	1.0000	0.0001	0.0020	-0.0070	0.0218
3.61	0.9999	0.0006	0.0071	-0.0214	0.0559	4.05	1.0000	0.0001	0.0017	-0.0059	0.0190
3.62	0.9999	0.0006	0.0069	-0.0208	0.0547	4.10	1.0000	0.0001	0.0014	-0.0051	0.0165
3.63	0.9999	0.0006	0.0067	-0.0203	0.0536	4.15	1.0000	0.0001	0.0012	-0.0043	0.0143
3.64	0.9999	0.0005	0.0065	-0.0198	0.0524	4.20	1.0000	0.0001	0.0010	-0.0036	0.0123
3.65	0.9999	0.0005	0.0063	-0.0192	0.0513	4.25	1.0000	0.0001	0.0008	-0.0031	0.0105
3.66	0.9999	0.0005	0.0061	-0.0187	0.0502	4.30	1.0000	0.0000	0.0007	-0.0026	0.0090
3.67	0.9999	0.0005	0.0059	-0.0182	0.0492	4.35	1.0000	0.0000	0.0006	-0.0022	0.0077
3.68	0.9999	0.0005	0.0057	-0.0177	0.0481	4.40	1.0000	0.0000	0.0005	-0.0018	0.0065
3.69	0.9999	0.0004	0.0056	-0.0173	0.0470	4.45	1.0000	0.0000	0.0004	-0.0015	0.0055
3.70	0.9999	0.0004	0.0054	-0.0168	0.0460	4.50	1.0000	0.0000	0.0003	-0.0012	0.0047
3.71	0.9999	0.0004	0.0052	-0.0164	0.0450	4.55	1.0000	0.0000	0.0003	-0.0010	0.0039
3.72	0.9999	0.0004	0.0051	-0.0159	0.0440	4.60	1.0000	0.0000	0.0002	-0.0009	0.0033
3.73	0.9999	0.0004	0.0049	-0.0155	0.0430	4.65	1.0000	0.0000	0.0002	-0.0007	0.0027
3.74	0.9999	0.0004	0.0048	-0.0150	0.0420	4.70	1.0000	0.0000	0.0001	-0.0006	0.0023
3.75	0.9999	0.0004	0.0046	-0.0146	0.0410	4.75	1.0000	0.0000	0.0001	-0.0005	0.0019
3.76	0.9999	0.0003	0.0045	-0.0142	0.0401	4.80	1.0000	0.0000	0.0001	-0.0004	0.0016
3.77	0.9999	0.0003	0.0043	-0.0138	0.0392	4.85	1.0000	0.0000	0.0001	-0.0003	0.0013
3.78	0.9999	0.0003	0.0042	-0.0134	0.0382	4.90	1.0000	0.0000	0.0001	-0.0003	0.0011
3.79	0.9999	0.0003	0.0041	-0.0131	0.0373	4.95	1.0000	0.0000	0.0000	-0.0002	0.0009

The sum of those terms of

$$(p+q)^n \equiv \sum_{t=0}^n \binom{n}{t} p^{n-t} q^t \quad p+q=1$$

in which t ranges from a to b inclusive, a and b being integers ($a \leq t \leq b$), is (if n is large enough) approximately

$$\int_{x_1}^{x_2} \phi(x) dx + \left[\frac{q-p}{6\sigma} \phi^2(x) + \frac{1}{24} \left(\frac{1}{\sigma^2} - \frac{6}{n} \right) \phi^{(3)}(x) \right]_{x_1}^{x_2}$$

where $x_1 = (a - \frac{1}{2} - qn)/\sigma$, $x_2 = (b + \frac{1}{2} - qn)/\sigma$.

The sum of the first $t+1$ terms of

$$(p+q)^n \equiv \sum_{t=0}^n \binom{n}{t} p^{n-t} q^t \quad p+q=1$$

is approximately

$$\int_x^\infty \phi(x) dx + \frac{q-p}{6\sigma} \phi^2(x) - \frac{1}{24} \left(\frac{1}{\sigma^2} - \frac{6}{n} \right) \phi^{(3)}(x)$$

where $x = (s - \frac{1}{2} - np)/\sigma$, $s = n - t$. The sum of the last $s+1$ terms is approximately

$$\int_x^\infty \phi(x) dx - \frac{q-p}{6\sigma} \phi^2(x) - \frac{1}{24} \left(\frac{1}{\sigma^2} - \frac{6}{n} \right) \phi^{(3)}(x)$$

where $x = (t - \frac{1}{2} - nq)/\sigma$.

The *probable error* of a single observation in a series of n measures, t_1, t_2, \dots, t_n , the arithmetic mean of which is m , is

$$e = \frac{0.6745}{\sqrt{n-1}} \sqrt{(m-t_1)^2 + (m-t_2)^2 + \dots + (m-t_n)^2}$$

the probable error of the mean is

$$E = \frac{0.6745}{\sqrt{n(n-1)}} \sqrt{(m-t_1)^2 + (m-t_2)^2 + \dots + (m-t_n)^2}$$

Approximate values of e and E are

$$e = 0.8453 \frac{\sum_{i=1}^n d_i}{\sqrt{n(n-1)}} \quad E = 0.8453 \frac{\sum_{i=1}^n d_i}{\sqrt{n(n-1)}}$$

where $\sum_{i=1}^n d_i$ is the sum of the deviations $d_i = |t_i - m|$.

TABLE 51.25 Factors for Computing Probable Errors

n	$\frac{1}{\sqrt{n}}$	$\frac{1}{\sqrt{n(n-1)}}$	$\frac{0.6745}{\sqrt{n-1}}$	$\frac{0.6745}{\sqrt{n(n-1)}}$	$\frac{0.8453}{n\sqrt{n-1}}$	$\frac{0.8453}{\sqrt{n(n-1)}}$
2	0.707 107	0.707 107	0.6745	0.4769	0.4227	0.5978
3	0.577 350	0.408 248	0.4769	0.2754	0.1993	0.3451
4	0.500 000	0.288 675	0.3894	0.1947	0.1220	0.2440
5	0.447 214	0.223 607	0.3372	0.1508	0.0845	0.1890
6	0.408 248	0.182 574	0.3016	0.1231	0.0630	0.1543
7	0.377 964	0.154 303	0.2754	0.1041	0.0493	0.1304
8	0.353 553	0.133 631	0.2549	0.0901	0.0399	0.1130
9	0.333 333	0.117 851	0.2385	0.0795	0.0332	0.0996
10	0.316 228	0.105 409	0.2248	0.0711	0.0282	0.0891
11	0.301 511	0.095 346	0.2133	0.0643	0.0243	0.0806
12	0.288 675	0.087 039	0.2034	0.0587	0.0212	0.0736
13	0.277 350	0.080 064	0.1947	0.0540	0.0188	0.0677
14	0.267 261	0.074 125	0.1871	0.0500	0.0167	0.0627
15	0.258 199	0.069 007	0.1803	0.0465	0.0151	0.0583
16	0.250 000	0.064 550	0.1742	0.0435	0.0136	0.0546
17	0.242 536	0.060 634	0.1686	0.0409	0.0124	0.0513
18	0.235 702	0.057 166	0.1636	0.0386	0.0114	0.0483
19	0.229 416	0.054 074	0.1590	0.0365	0.0105	0.0457
20	0.223 607	0.051 299	0.1547	0.0346	0.0097	0.0434
21	0.218 218	0.048 795	0.1508	0.0329	0.0090	0.0412
22	0.213 201	0.046 524	0.1472	0.0314	0.0084	0.0393
23	0.208 514	0.044 455	0.1438	0.0300	0.0078	0.0376
24	0.204 124	0.042 563	0.1406	0.0287	0.0073	0.0360
25	0.200 000	0.040 825	0.1377	0.0275	0.0069	0.0345
26	0.196 116	0.039 223	0.1349	0.0265	0.0065	0.0332
27	0.192 450	0.037 743	0.1323	0.0255	0.0061	0.0319
28	0.188 982	0.036 370	0.1298	0.0245	0.0058	0.0307
29	0.185 695	0.035 093	0.1275	0.0237	0.0055	0.0297
30	0.182 574	0.033 903	0.1252	0.0229	0.0052	0.0287
31	0.179 605	0.032 791	0.1231	0.0221	0.0050	0.0277
32	0.176 777	0.031 750	0.1211	0.0214	0.0047	0.0268
33	0.174 078	0.030 773	0.1192	0.0208	0.0045	0.0260
34	0.171 499	0.029 854	0.1174	0.0201	0.0043	0.0252
35	0.169 031	0.028 989	0.1157	0.0196	0.0041	0.0245
36	0.166 667	0.028 172	0.1140	0.0190	0.0040	0.0238
37	0.164 399	0.027 400	0.1124	0.0185	0.0038	0.0232
38	0.162 221	0.026 669	0.1109	0.0180	0.0037	0.0225
39	0.160 128	0.025 976	0.1094	0.0175	0.0035	0.0220
40	0.158 114	0.025 318	0.1080	0.0171	0.0034	0.0214
41	0.156 174	0.024 693	0.1066	0.0167	0.0033	0.0209
42	0.154 303	0.024 098	0.1053	0.0163	0.0031	0.0204
43	0.152 499	0.023 531	0.1041	0.0159	0.0030	0.0199
44	0.150 756	0.022 990	0.1029	0.0155	0.0029	0.0194

TABLE 51.25 (Continued)

n	$\frac{1}{\sqrt{n}}$	$\frac{1}{\sqrt{n(n-1)}}$	$\frac{0.6745}{\sqrt{n-1}}$	$\frac{0.6745}{\sqrt{n(n-1)}}$	$\frac{0.8453}{n\sqrt{n-1}}$	$\frac{0.8453}{\sqrt{n(n-1)}}$
45	0.149 071	0.022 473	0.1017	0.0152	0.0028	0.0190
46	0.147 442	0.021 979	0.1005	0.0148	0.0027	0.0186
47	0.145 865	0.021 507	0.0994	0.0145	0.0027	0.0182
48	0.144 338	0.021 054	0.0984	0.0142	0.0026	0.0178
49	0.142 857	0.020 620	0.0974	0.0139	0.0025	0.0174
50	0.141 421	0.020 203	0.0964	0.0136	0.0024	0.0171
51	0.140 028	0.019 803	0.0954	0.0134	0.0023	0.0167
52	0.138 675	0.019 418	0.0945	0.0131	0.0023	0.0164
53	0.137 361	0.019 048	0.0935	0.0129	0.0022	0.0161
54	0.136 083	0.018 692	0.0927	0.0126	0.0022	0.0158
55	0.134 840	0.018 349	0.0918	0.0124	0.0021	0.0155
56	0.133 631	0.018 019	0.0910	0.0122	0.0020	0.0152
57	0.132 453	0.017 700	0.0901	0.0119	0.0020	0.0150
58	0.131 306	0.017 392	0.0893	0.0117	0.0019	0.0147
59	0.130 189	0.017 095	0.0886	0.0115	0.0019	0.0145
60	0.129 099	0.016 807	0.0878	0.0113	0.0018	0.0142
61	0.128 037	0.016 529	0.0871	0.0112	0.0018	0.0140
62	0.127 000	0.016 261	0.0864	0.0110	0.0018	0.0138
63	0.125 988	0.016 001	0.0857	0.0108	0.0017	0.0135
64	0.125 000	0.015 749	0.0850	0.0106	0.0017	0.0133
65	0.124 035	0.015 504	0.0843	0.0105	0.0016	0.0131
66	0.123 091	0.015 268	0.0837	0.0103	0.0016	0.0129
67	0.122 169	0.015 038	0.0830	0.0101	0.0016	0.0127
68	0.121 268	0.014 815	0.0824	0.0100	0.0015	0.0125
69	0.120 386	0.014 599	0.0818	0.0099	0.0015	0.0123
70	0.119 523	0.014 389	0.0812	0.0097	0.0015	0.0122
71	0.118 678	0.014 185	0.0806	0.0096	0.0014	0.0120
72	0.117 851	0.013 986	0.0801	0.0094	0.0014	0.0118
73	0.117 041	0.013 793	0.0795	0.0093	0.0014	0.0117
74	0.116 248	0.013 606	0.0789	0.0092	0.0013	0.0115
75	0.115 470	0.013 423	0.0784	0.0091	0.0013	0.0113
76	0.114 708	0.013 245	0.0779	0.0089	0.0013	0.0112
77	0.113 961	0.013 072	0.0773	0.0088	0.0013	0.0111
78	0.113 228	0.012 904	0.0769	0.0087	0.0012	0.0109
79	0.112 509	0.012 739	0.0764	0.0086	0.0012	0.0108
80	0.111 803	0.012 579	0.0759	0.0085	0.0012	0.0106
81	0.111 111	0.012 423	0.0754	0.0084	0.0012	0.0105
82	0.110 432	0.012 270	0.0749	0.0083	0.0012	0.0104
83	0.109 764	0.012 121	0.0745	0.0082	0.0011	0.0103
84	0.109 109	0.011 976	0.0740	0.0081	0.0011	0.0101
85	0.108 465	0.011 835	0.0736	0.0080	0.0011	0.0100
86	0.107 833	0.011 696	0.0732	0.0079	0.0011	0.0099
87	0.107 211	0.011 561	0.0727	0.0078	0.0011	0.0098
88	0.106 600	0.011 429	0.0723	0.0077	0.0010	0.0097
89	0.106 000	0.011 300	0.0719	0.0076	0.0010	0.0096

(continued)

TABLE 51.25 (Continued)

n	$\frac{1}{\sqrt{n}}$	$\frac{1}{\sqrt{n(n-1)}}$	$\frac{0.6745}{\sqrt{n-1}}$	$\frac{0.6745}{\sqrt{n(n-1)}}$	$\frac{0.8453}{n\sqrt{n-1}}$	$\frac{0.8453}{\sqrt{n(n-1)}}$
90	0.105 409	0.011 173	0.0715	0.0075	0.0010	0.0094
91	0.104 828	0.011 050	0.0711	0.0075	0.0010	0.0093
92	0.104 257	0.010 929	0.0707	0.0074	0.0010	0.0092
93	0.103 695	0.010 811	0.0703	0.0073	0.0010	0.0091
94	0.103 142	0.010 695	0.0699	0.0072	0.0009	0.0090
95	0.102 598	0.010 582	0.0696	0.0071	0.0009	0.0089
96	0.102 062	0.010 471	0.0692	0.0071	0.0009	0.0089
97	0.101 535	0.010 363	0.0688	0.0070	0.0009	0.0088
98	0.101 015	0.010 257	0.0685	0.0069	0.0009	0.0087
99	0.100 504	0.010 152	0.0681	0.0069	0.0009	0.0086
100	0.100 000	0.010 050	0.0678	0.0068	0.0008	0.0085

TABLE 51.26 Statistics and Probability Formulas

$p(x) = dP(x)/dx$	Differential probability (density) function of random variable x ; univariate frequency function
$P(x) = \int_{-\infty}^x P(x') dx'$	Cumulative probability function of random variable x ; univariate distribution function
$P(A < x < B)$	Cumulative probability that x is between A and B
$P(E \cap F)$	Probability of simultaneous (joint) occurrence of E and F
$P(E \cup F)$	Probability of occurrence of E or F or both
$P(E F) = P(E \cap F)/P(F)$	Conditional probability; probability of occurrence of E provided F has occurred
$E[f(x)] = \int_{-\infty}^{\infty} f(x)p(x) dx$	Expected value of function of a random variable x
$E(x) = \bar{x} = \int_{-\infty}^{\infty} xp(x)dx$	Expected (mean) value of random variable x
$\alpha_r = E(x^r)$	r th moment of random variable x ; r th moment about the origin
$\mu_r = E(x - \bar{x})^r$	r th moment of random variable x from mean value; r th central moment
$\text{Var } x = \frac{E[(x - \bar{x})^2]}{x^2 - \bar{x}^2} = \overline{(x - \bar{x})^2}$	Variance value of random variable x
$\sigma = (\text{Var } x)^{1/2}$	Standard deviation of random variable x
$M_x(s) = E(e^{sx})$	Moment generating function associated with random variable x
$\psi_x(q) = E(e^{jqx})$	Characteristic function associated with random variable x
$\psi_g(q) = E[e^{jqg(x)}]$	Characteristic function of $g(x)$ with random variable x
$p(x, y) = d^2P(x, y)/dx dy$	Differential probability (density) function of random variables x and y ; bivariate frequency function
$P(x, y) = \int_{-\infty}^x \int_{-\infty}^y p(x', y') dx' dy'$	Cumulative probability function of random variables x and y ; bivariate distribution function
$P(A < x < B, C < y < D)$	Cumulative probability that x is between A and B and that also y is between C and D ; cumulative joint probability
$\text{Cov}(x, y) = E[(x - \bar{x})(y - \bar{y})] = (x - \bar{x})(y - \bar{y})$	Covariance value of random variables x and y
$\rho(x, y) = \text{Cov}(x, y)/\sigma_x \sigma_y$	Correlation coefficient of random variables x and y

Source: Giacoletto, *Electronic Designers' Handbook*, Copyright © 1977 by McGraw-Hill, pp. 1–8.

51.4 UNITS AND STANDARDS

51.4.1 Physical Quantities and Their Relations

Mathematics is concerned with relations between numerical quantities, either constant or varying in a specified manner over a specified range of values. The numerical values are unique, absolute, and the same all over the world, being the expression of a fundamental perception of the mind. Any *mathematical equation* defines the values of one numerical quantity, known as the dependent, in terms of constants and one or more other numerical quantities, known as the independent variables, as, for example,

$$z = r^2 + 3x + 4 \quad y = c \cdot \int_0^x \frac{x^2}{\cos x} dx \quad (51.1)$$

where z and y are dependent variables, r and x independent variables, and c a constant.

Physics, comprising the knowledge of inanimate nature and its laws, is concerned fundamentally with the measuring of the various quantities founded or created by definition, as, for example, *length*, *mass*, and *electric charge*. In order to specify a *physical quantity* it is not sufficient to state merely a number. The value of a physical quantity can be determined only by comparison of the sample with a known amount of the same quantity by the process of *measuring*. The reference amount is called a unit, and the result of any measurement must be a statement of “how many times the sample was found to contain the reference amount.” Thus a physical quantity Q naturally appears to be the product of a numerical value N and a unit U ,

$$Q = N \cdot U \quad (51.2)$$

as, for example: The length of a particular rod is 3.5 ft, or the rod is $3\frac{1}{2}$ times the length of 1 ft. Obviously, the reproduction of a unit must be possible at any time in order to facilitate correct measurements. This is being done by means of the “standards,” which are simply a set of fundamental unit quantities kept under normalized conditions in order to preserve their values as accurately as facilities permit.

Any physical relation must be the result of a more or less obvious measurement, so that equations in physics are not merely numerical relations but express dependencies between physical quantities. Mathematics does not know “standards”; physics cannot be without “standards.” The fact that physics often uses the methods of mathematics must not lead to the identification of the two sciences; it is merely an overlapping in the border regions.

51.4.1.1 Relations Between Units A unit is a particular amount of the physical quantity to be measured defined in terms of a standard. The choice of a unit depends on convenience, facility of reproduction, and easy subdivision so as to obtain smaller units if desired. The value of a physical quantity Q must be independent of the units used, so that for two different units of the same type

$$Q = N_1 \cdot U_1 = N_2 \cdot U_2 \quad (51.3)$$

The size of the unit and the numerical value of the quantity are inversely related: the larger the unit the smaller the number of units.

A unit relation is an equation between two different units of the same type,

$$U_1 = N_{12} \cdot U_2 \quad (51.4)$$

and serves to convert from one unit U_1 to a different one U_2 . The conversion is achieved by replacing U_1 , taken as a factor, by its equivalent according to Eq. (51.4) so that

$$Q = N_1 \cdot U_1 = N_1 \cdot (N_{12} \cdot U_2) = (N_1 \cdot N_{12}) \cdot U_2 \quad (51.5)$$

As an example, express the length 3.5 ft in centimeters. The unit relation is 1 ft = 30.5 cm, and therefore $l = 3.5 \text{ ft} = 3.5 \times (30.5 \text{ cm}) = 106.75 \text{ cm}$. No error is possible if this rule is followed properly.

51.4.1.2 Physical Equations Relations between physical quantities are usually given in the form of equations. It is always possible, by the proper use of unit relations (see previous paragraph), to express each side in the same units. Since units are to be considered as factors, they may be canceled and a numerical identity must result. This fact always can be used to check the proper numerical relations and the consistency of the units used.

There are two fundamental types of physical equations:

1. The **mathematical definition** of a physical quantity determines a new quantity uniquely in terms of known quantities. An example is Newton's definition of mass by $f = m \cdot a$, where f is the force and a the acceleration of a moving body. If f and a are measured, m can be computed as a physical quantity with numerical value $N(f)/N(a)$ and unit $U(f)/U(a) = U(m)$. A definition should be in agreement with all the other known relations in a particular field of science; it can only be of restricted value if it contradicts other relations (see later the "absolute" electric systems).
2. The **statement of proportionality** defines one physical quantity as linearly depending on a combination of other, known quantities. It is always the result of an experimental investigation. An example is Newton's law of the gravitational force $f = k(m_1 m_2 / r^2)$, where m_1 and m_2 are the two masses, r their center distance, and k the proportionality factor. In the case of a proportionality it is permissible to choose arbitrary units for all measurable physical quantities involved and to use the equation as a definition of the proportionality constant that, in general, will be a physical constant with numerical value and unit. In the example the value of k would be

$$\frac{N(f) \cdot N(r^2)}{N(m_1) \cdot N(m_2)} \times \frac{U(f) \cdot U(r^2)}{U(m_1) \cdot U(m_2)} = N(k) \cdot U(k)$$

Most of the fundamental laws of physics are statements of proportionalities, leading to universal physical constants, as, for instance, the gravitational constant k , the Planck constant h , the gas constant R , the absolute permeability of free space μ_v , and the absolute dielectric constant of free space ϵ_v . It may be observed that each branch of physics is represented by at least one fundamental proportionality constant.

Derived physical quantities are, in general, the result of mathematical definitions. The units of derived quantities are expressed from the combinations of the units used

in the definition. All proportionality constants are ordinarily considered as derived physical quantities.

51.4.1.3 Fundamental Physical Quantities The physical quantities, arbitrarily chosen to define new quantities or derived quantities, are called fundamental physical quantities. Their number may vary according to needs and convenience. There is no possibility to designate any physical quantity as absolutely fundamental, or a priori fundamental. Quantities that appear to be fundamental in one special field may be derived quantities in some other field.

51.4.2 Dimensions and Dimension Systems

51.4.2.1 Definition of Dimension To choose a unit for a physical quantity one has an infinity of possibilities. The numerous units of length that were in use about 100 years ago present a good practical illustration. Yet all these units have in common the quality of being a distinct length and not, for example, a volume. It is convenient to state this fact by representing with the notation $[L]$ any unit of length whatsoever. The measurement of a physical quantity Q , therefore, leads to the statement

$$Q = N \cdot [Q] \quad (51.6)$$

where N is a numeric denoting the number of general units $[Q]$ that constitute the total quantity Q . According to Fourier, who first introduced this concept into the literature, $[Q]$ is called the “dimension” of the quantity Q . Be it clearly understood that dimension is simply the expression of a general unit and therefore a characteristic peculiarity of physical quantities not occurring in mathematics. Each new physical quantity gives rise to a new “dimension”, as, for instance, time $[T]$, force $[F]$, mass $[M]$, and so on. There are as many dimensions, or general units, as there are kinds of physical quantities.

51.4.2.2 Derived Dimensions Many physical quantities have been introduced by mathematical definition. Velocity, for example, is defined as $v = ds/dt$, where s is the length of the path measured from a definite origin and t is the time. A possible expression for the dimension of velocity would be $[V]$. It is customary and convenient, however, to make use of the mathematical definition that is but the rule for the measurement of velocity and to express the dimension in terms of the more familiar dimensions of length and time as a derived dimension $[V] = [L]/[T] = [L][T]^{-1}$. [Read: velocity is of +1 dimension in length and -1 dimension in time.] The use of mathematical definitions, leading to derived dimensions of a composite nature, reduces the number of symbols. Thus the measurement of volume, if scientifically conducted, gives $[\text{Vol}] = [L]^3$, or in words, “volume is of +3 dimensions in length $[L]$.”

Proportionality constants of physics have, in general, *derived dimensions*, as they are defined by the corresponding physical equations.

51.4.2.3 Fundamental Dimensions The more familiar dimensions used to express derived dimensions are referred to as fundamental dimensions. It is advantageous to use as few of these fundamental dimensions as possible, not because the physical relations become simpler or clearer, but merely as a matter of economy in symbols. In fact, any dimension can be chosen to be a fundamental dimension in a particular field and a derived

dimension in some other field of physics. No fundamental dimension can be made a starting point of natural philosophy.

51.4.2.4 Dimensional Equations Since a physical equation constitutes in fact two equations, one for the units and one for the numerics, one can disregard the numerical factors entirely and write the general units or dimensions only, arriving thus at a dimensional equation. For instance, the law of gravitation would read $[F] = [k][M]^2[L]^{-2}$ using $[F]$, $[k]$, $[M]$, and $[L]$ as dimensions for force, gravitation constant, mass, and length, respectively. From this dimensional equation a derived dimension can be obtained for any quantity involved. Conversely, dimensional equations are used to check the correctness of physical relations if all dimensions can be made to cancel. Finally, the validity of dimensional equations leads to the method of dimensional analysis.

A *set of fundamental dimensions* is any group of fundamental dimensions convenient and useful to express all the physical quantities of a particular field in terms of derived dimensions. The number of fundamental dimensions to make a set may vary according to the field of application. Whether or not a set of fundamental dimensions can be used beyond the field for which it was originally intended will depend upon its suitability as a dimension system. (See next paragraph.) In no case should it be used where it can lead to confusion.

A set of fundamental dimensions is *incomplete* when the number of fundamental dimensions composing it is less than the number required for a dimension system. Incomplete sets of fundamental dimensions should not be used outside the very restricted field for which they are defined; they necessarily would lead to confusing relations.

A *dimension system* is composed of the smallest number of fundamental dimensions that will form a consistent and complete set for a field of science. Since each relation between physical quantities can be split up into one relation of numerics and another one of dimensions (as general units), it is possible to combine all known relations of dimensions. In setting up these relations, all proportionality factors must be taken as physical quantities. If there are m independent relations known, $m + p$ dimensions may be involved, of which m dimensions can be expressed by any p “fundamental” dimensions chosen arbitrarily.

This set of p “fundamental” dimensions is then called a dimension system. From the theory of numbers, therefore, it is known that one generally has a choice of $\binom{m+p}{p}$ possible dimension systems. Thus, if $p = 3$, $m = 3$, then one has $\binom{6}{3} = 20$ different possibilities. A necessary condition, however, is that *each* independent relation involve at least $p + 1$ dimensions. If this is not the case, then the number of possible dimension systems is less, so that $\binom{m+p}{p}$ indicates the *upper* limit.

Any dimension system chosen in the described manner is consistent, as well as correct, and never leads to ambiguity with respect to the expression of physical quantities. Complete dimension systems in mechanics must have three, in thermodynamics four, and in electromagnetism four fundamental dimensions. It seems, according to present knowledge, that five fundamental dimensions suffice for the entire range of physics, namely, the three fundamental dimensions of mechanics, an additional one for thermodynamics, and another additional one for electromagnetism.

All the known dimension systems use length $[L]$ and time $[T]$ as primary fundamental dimensions, adding various fundamental dimensions from the available physical quantities of the fields of physics. The choice of $[L]$ and $[T]$ reduces at once the maximum number of possible dimension systems to $\binom{m+p-2}{p-2}$.

51.4.2.5 Why Dimension Systems? Since the proper choice of units is the ultimate goal of any critical analysis of physical quantities, the question may be asked: Why is it necessary to discuss dimension of systems? The answer is that each physical quantity may be measured by an infinite variety of units but has only one dimension within a given dimension system. The process of deciding upon the fundamental dimensions before fixing the units within the scope of the fundamental dimensions is, therefore, essentially a matter of economy and logic.

51.4.3 Dimension and Unit Systems

In the past different dimension systems were introduced for various fields of technology (mechanics, heat, electromagnetism) and based on different choices of fundamental dimensions, for example, for mechanics, length and time plus mass or force or energy or gravitational constant gave potentially four different dimension system classes. In turn, for a given dimensional system, a unit system could be developed, choosing for each fundamental dimension a unit desirably related to a fundamental standard or standards. In seeking to define units with appropriate size values, relationships, and so on, many different unit systems, for example, centimeter–gram–second (cgs), meter–kilogram–second (mks), “absolute,” “technical,” and so on, have been introduced over the years. (See O. W. Eshbach and M. Souders, *Handbook of Engineering Fundamentals*, 3rd ed., Wiley, New York, 1975, for a detailed exposition of the subject.)

In recent years a major step toward simplification and standardization has been taken by the increasing adoption of the International System of Units (SI).

51.4.4 The International System of Units

The SI system, composed of six fundamental units, has been adopted by the Conference Générale (BIPM Sevres, Paris, 1954 and 1960) to cover the whole range of physics and one in which all international reports are to be expressed.

Quantity	Unit	Symbol
<i>Fundamental Units</i>		
Length	Meter	m
Mass	Kilogram	kg
Time	Second	s
Intensity of electric current	Ampere	A
Thermodynamic temperature	Degree kelvin	K
Luminous intensity	Candela	cd
Amount of substance	Mole	mol
<i>Derived Units with Special Names</i>		
Area	Square meter	m ²
Volume	Cubic meter	m ³
Frequency	Hertz	Hz
Density (mass density)	Kilogram per cubic meter	kg/m ³
Velocity	Meter per second	m/s
Angular velocity	Radian per second	rad/s

(continued)

Quantity	Unit	Symbol
Acceleration	Meter per square second	m/s^2
Angular acceleration	Radian per square second	rad/s^2
Force	Newton	N, $\text{kg}\cdot\text{m/s}^2$
Pressure, stress	Newton per square meter	N/m^2
Kinematic viscosity	Square meter per second	m^2/s
Dynamic viscosity	Newton-second per square meter	$\text{N}\cdot\text{s/m}^2$
Work, energy, heat (quantity of heat)	Joule	J, $\text{N}\cdot\text{m}$
Power, radiant flux	Watt	W, J/s
Plane angle	Radian	rad
Solid angle	Steradian	sr
Electric charge	Coulomb	C, $\text{A}\cdot\text{s}$
Electric potential, potential difference, electromotive force	Volt	V, W/A
Electric field strength	Volt per meter	V/m
Resistance (to direct current)	Ohm	Ω , V/A
Electric conductance	Siemens	S, A/V
Capacitance	Farad	F, $\text{A}\cdot\text{s/V}$
Magnetic flux	Weber	Wb, $\text{V}\cdot\text{s}$
Inductance	Henry	H, $\text{V}\cdot\text{s/A}$
Magnetic flux density (magnetic induction)	Tesla	T, Wb/m^2
Magnetic field strength	Ampere per meter	A/m
Magnetomotive force	Ampere	A
Luminous flux	Lumen	lm, $\text{cd}\cdot\text{sr}$
Luminance	Candela per square meter	cd/m^2
Illumination	Lux	lx, lm/m^2
Activity (of a radionuclide)	Becquerel	Bq, 1/S
Absorbed dose	Gray	Gy, J/kg
Dose equivalent	Sievert	Sv, J/kg
<i>Other Common Derived Units</i>		
Absorbed dose rate	Gray per second	Gy/s
Acceleration	Meter per second squared	m/s^2
Angular acceleration	Radian per second squared	rad/s^2
Angular velocity	Radian per second	rad/s
Area	Square meter	m^2
Concentration (of amount of substance)	Mole per cubic meter	mol/m^3
Current density	Ampere per square meter	A/m^2
Density, mass	Kilogram per cubic meter	kg/m^3
Electric charge density	Coulomb per cubic meter	C/m^3
Electric field strength	Volt per meter	V/m
Electric flux density	Coulomb per square meter	C/m^2
Energy density	Joule per cubic meter	J/m^3
Entropy	Joule per kelvin	J/K
Exposure (X and gamma rays)	Coulomb per kilogram	C/kg
Heat capacity	Joule per kelvin	J/K
Heat flux density	Watt per square meter	W/m^2
Irradiance	Watt per square meter	W/m^2
Luminance	Candela per square meter	cd/m^2
Magnetic field strength	Ampere per meter	A/m
Molar energy	Joule per mole	J/mol
Molar entropy	Joule per mole kelvin	J/(mol · K)

Quantity	Unit	Symbol
Molar heat capacity	Joule per mole kelvin	J/(mol · K)
Moment of force	Newton meter	N · m
Permeability (magnetic)	Henry per meter	H/m
Permittivity	Farad per meter	F/m
Power density	Watt per square meter	W/m ²
Radiance	Watt per square meter steradian	W/(m ² · sr)
Radiant intensity	Watt per steradian	W/sr
Specific heat capacity	Joule per kilogram kelvin	J/(kg · K)
Specific energy	Joule per kilogram	J/kg
Specific entropy	Joule per kilogram kelvin	J/(kg · K)
Specific volume	Cubic meter per kilogram	m ³ /kg
Surface tension	Newton per meter	N/m
Thermal conductivity	Watt per meter kelvin	W/(m · K)
Velocity	Meter per second	m/s
Viscosity, dynamic	Pascal second	Pa · s
Viscosity, kinematic	Square meter per second	m ² /s
Volume	Cubic meter	m ³
Wave number	1 per meter	1/m

Definitions of Derived Units of the International System Having Special Names

Quantity	Unit and Definition
1. Absorbed dose	The <i>gray</i> is the absorbed dose when the energy per unit mass imparted to matter by ionizing radiation is one joule per kilogram. <i>Note:</i> The gray is also used for the ionizing radiation quantities: specific energy imparted, kerma, and absorbed dose index, which have the SI unit joule per kilogram.
2. Activity	The <i>becquerel</i> is the activity of a radionuclide decaying at the rate of one spontaneous nuclear transition per second.
3. Celsius temperature	The <i>degree Celsius</i> is equal to the kelvin and is used in place of the kelvin for expressing Celsius temperature (symbol <i>t</i>) defined by the equation $t = T - T_0$, where <i>T</i> is the thermodynamic temperature and $T_0 = 273.15$ K by definition.
4. Dose equivalent	The <i>sievert</i> is the dose equivalent when the absorbed dose of ionizing radiation multiplied by the dimensionless factors <i>Q</i> (quality factor) and <i>N</i> (product of any other multiplying factors) stipulated by the International Commission on Radiological Protection is one joule per kilogram.
5. Electric capacitance	The <i>farad</i> is the capacitance of a capacitor between the plates of which there appears a difference of potential of one volt when it is charged by a quantity of electricity equal to one coulomb.
6. Electric conductance	The <i>siemens</i> is the electric conductance of a conductor in which a current of one ampere is produced by an electric potential difference of one volt.
7. Electric inductance	The <i>henry</i> is the inductance of a closed circuit in which an electromotive force of one volt is produced when the electric current in the circuit varies uniformly at a rate of one ampere per second.

(continued)

Quantity	Unit and Definition
8. Electric potential difference, electromotive force	The <i>volt</i> (unit of electric potential difference and electromotive force) is the difference of electric potential between two points of a conductor carrying a constant current of one ampere when the power dissipated between these points is equal to one watt.
9. Electric resistance	The <i>ohm</i> is the electric resistance between two points of a conductor when a constant difference of potential of one volt, applied between these two points, produces in this conductor a current of one ampere, this conductor not being the source of any electromotive force.
10. Energy	The <i>joule</i> is the work done when the point of application of a force of one newton is displaced a distance of one meter in the direction of the force.
11. Force	The <i>newton</i> is that force that, when applied to a body having a mass of one kilogram, gives it an acceleration of one meter per second squared.
12. Frequency	The <i>hertz</i> is the frequency of a periodic phenomenon of which the period is one second.
13. Illuminance	The <i>lux</i> is the illuminance produced by a luminous flux of one lumen uniformly distributed over a surface of one square meter.
14. Luminous flux	The <i>lumen</i> is the luminous flux emitted in a solid angle of one steradian by a point source having a uniform intensity of one candela.
15. Magnetic flux	The <i>weber</i> is the magnetic flux that, linking a circuit of one turn, produces in it an electromotive force of one volt as it is reduced to zero at a uniform rate in one second.
16. Magnetic flux density	The <i>tesla</i> is the magnetic flux density given by a magnetic flux of one weber per square meter.
17. Power	The <i>watt</i> is the power that gives rise to the production of energy at the rate of one joule per second.
18. Pressure or stress	The <i>pascal</i> is the pressure or stress of one newton per square meter.
19. Quantity of electricity	The <i>coulomb</i> is the quantity of electricity transported in one second by a current of one ampere.

Prefixes. The SI system has adopted the following standard set of prefixes:

Multiplication Factor	Prefix	Symbol
$1\ 000\ 000\ 000\ 000\ 000\ 000 = 10^{18}$	Exa	E
$1\ 000\ 000\ 000\ 000\ 000 = 10^{15}$	Peta	P
$1\ 000\ 000\ 000\ 000 = 10^{12}$	Tera	T
$1\ 000\ 000\ 000 = 10^9$	Giga	G
$1\ 000\ 000 = 10^6$	Mega	M
$1\ 000 = 10^3$	Kilo	k
$100 = 10^2$	Hecto ¹⁴	h
$10 = 10^1$	Deka ¹⁴	da
$0.1 = 10^{-1}$	Deci ¹⁴	d
$0.01 = 10^{-2}$	Centi ¹⁴	c
$0.001 = 10^{-3}$	Milli	m
$0.000\ 001 = 10^{-6}$	Micro	μ
$0.000\ 000\ 001 = 10^{-9}$	Nano	n
$0.000\ 000\ 000\ 001 = 10^{-12}$	Pico	p
$0.000\ 000\ 000\ 000\ 001 = 10^{-15}$	Femto	f
$0.000\ 000\ 000\ 000\ 000\ 001 = 10^{-18}$	Atto	a

51.4.5 Application of SI Prefixes

General In general the SI prefixes should be used to indicate orders of magnitude, thus eliminating nonsignificant digits and leading zeros in decimal fractions and providing a convenient alternative to the powers-of-10 notation preferred in computation. For example,

$$\begin{aligned} 12,300 \text{ mm} &\text{ becomes } 12.3 \text{ m} \\ 12.3 \times 10^3 \text{ m} &\text{ becomes } 12.3 \text{ km} \\ 0.00123 \text{ }\mu\text{A} &\text{ becomes } 1.23 \text{ nA} \end{aligned}$$

Selection When expressing a quantity by a numerical value and a unit, a prefix should preferably be chosen so that the numerical value lies between 0.1 and 1000. To minimize variety, it is recommended that prefixes representing 1000 raised to an integral power be used. However, three factors may justify deviation:

1. In expressing area and volume, the prefixes hecto-, deka-, deci-, and centi- may be required, for example, square hectometer, cubic centimeter.
2. In tables of values of the same quantity or in a discussion of such values within a given context, it is generally preferable to use the same unit multiple throughout.
3. For certain quantities in particular applications, one particular multiple is customarily used. For example, the millimeter is used for linear dimensions in mechanical engineering drawings even when the values lie far outside the range 0.1–1000 mm; the centimeter is often used for body measurements and clothing sizes.

Prefixes in Compound Units² It is recommended that only one prefix be used in forming a multiple of a compound unit. Normally the prefix should be attached to a unit in the numerator. One exception to this is when the kilogram occurs in the denominator. For example,

$$\text{V/m, not mV/mm, and MJ/kg, not kJ/g}$$

Compound Prefixes Compound prefixes formed by the juxtaposition of two or more SI prefixes are not to be used. For example, use

$$\begin{aligned} 1 \text{ nm, not } 1 \text{ m}\mu\text{m} \\ 1 \text{ pF, not } 1 \mu\mu\text{F} \end{aligned}$$

If values are required outside the range covered by the prefixes, they should be expressed by using powers of 10 applied to the base unit.

Powers of Units An exponent attached to a symbol containing a prefix indicates that the multiple or sub-multiple of the unit (the unit with its prefix) is raised to the power expressed by the exponent. For example,

$$\begin{aligned} 1 \text{ cm}^3 &= (10^{-2}\text{m})^3 = 10^{-6}\text{m}^3 \\ 1 \text{ ns}^{-1} &= (10^{-9}\text{s})^{-1} = 10^9\text{s}^{-1} \\ 1 \text{ mm}^2/\text{s} &= (10^{-3}\text{m})^2/\text{s} = 10^{-6}\text{m}^2/\text{s} \end{aligned}$$

² A compound unit is a derived unit that is expressed in terms of two or more units rather than by a single special name.

Calculations Errors in calculations can be minimized if the base and the coherent derived SI units are used and the resulting numerical values are expressed in powers-of-10 notation instead of using prefixes.

51.4.6 Other Units

Units from Different Systems To assist in preserving the advantage of SI as a coherent system, it is advisable to minimize the use with it of units from other systems. Such use should be limited to units listed in this section.

A following section presents conversion factors to and from SI units.

51.4.7 Length, Mass, and Time (English Units and Standards)

51.4.7.1 Units of Length The *foot* (ft) is the *fundamental* unit of length in the foot-pound-second (fps) system. It equals, by definition, one-third of a *yard* (yd), which is the English legalized *standard* unit of length. The *U.S. yard* was defined by Act of Congress, July 28, 1866, as 3600/3937 the length of the *meter*. (See discussion of metric system for definitions of metric length.)

In Great Britain, the *Imperial yard* is measured by a bronze bar preserved in the Standards Office, Westminster. Its length, in terms of the *international prototype meter*, is 3600/3937.0113 m. For engineering purposes, the U.S. and British *yards* may be considered identical.

As subunits, the *inch* (in.) is defined as one-twelfth of one standard foot, and the *mil* as the one-thousandth part of one inch. The *nautical mile* (mi) is defined as one minute of arc on the earth's surface at the equator, whereas the U.S. mile (U.S. mi statute) is exactly 5280 ft and practically identical with the British mile.

51.4.7.2 Unit of Capacity (Dry) The *bushel* (bu) is the *standard* unit of *dry* capacity. The *Winchester bushel* (U.S. standard) has a volume of 2150.42 in.³

In Great Britain, the *Imperial bushel* (bu) is defined as the volume of 80 lb of pure water at 62°F weighed against brass weights in air at the same temperature as the water and with the barometer at 30 in. Its volume is approximately 2219.36 in.³

51.4.7.3 Unit of Capacity (Liquid) The *gallon* (gal) is the *standard* unit of *liquid* capacity. The *U.S. gallon* has a volume of 231 in.³

In Great Britain, the *Imperial gallon* is defined as the volume of 10 lb of pure water at 62°F weighed against brass weights in air at the same temperature as the water and with the barometer at 30 in. Its volume is approximately 277.420 in.³ The Imperial gallon (liquid measure) equals exactly one-eighth of the Imperial bushel (dry measure). Subunits are the quart (qt), which is one-fourth of the standard gallon, and the pint (pt), which is $\frac{1}{2}$ qt.

51.4.7.4 Units of Mass The *pound (avoirdupois)* (lb avdp) is the *fundamental* unit of mass in the fps system.³ It is also the English legalized *standard* unit of mass. The *U.S.*

³ The slug of mass, which is extensively used by engineers and physicists, is (in the English system) the mass to which an acceleration of one foot per second per second would be given by the application of a one-pound force. Under any gravity conditions, 1 slug of mass = 32.1739 lb of mass.

pound (avoirdupois) was defined by Act of Congress, 1866, as $1/2.2046$ kg, but since 1895 there has been used, for greater accuracy, a value that agrees with that given by law as far as the latter is given, namely, 453.5924277 g. This value is now used by the Bureau of Standards as an exact definition and is the basis of the customary U.S. weights (Circular 47, Bureau of Standards).

In Great Britain, the *Imperial pound (avoirdupois)* is the mass of a *platinum cylinder* preserved in the Standards Office, Westminster. Its legal equivalent is 453.59243 g. For engineering purposes, the U.S. and British pounds (avoirdupois) may be considered as identical.

Subunits of mass are the grain (gr), defined as $\frac{1}{7000}$ of the standard pound (avoirdupois) and the ounce (avoirdupois) (oz-avdp), which is $\frac{1}{16}$ of the standard pound (avoirdupois). The grain was used as a fundamental unit in the so-called foot–gram–second (fgs) system of units prior to 1873.

51.4.7.5 Weight versus Mass Unfortunately, the word “weight” is used in two different senses, namely, (1) by the layman (as well as loosely by the scientist) to designate a given *mass* or quantity of matter and (2) by the scientist to designate the *pull* in standard gravitational force units that is exerted by the earth upon a piece of matter. The result of the commercial act of “weighing” a specific quantity is independent of the local gravitational pull of the earth, since both spring scales and balances are calibrated locally by comparison with standard masses.

51.4.7.6 Auxiliary Fundamental Units Auxiliary Fundamental Units and their principal derived units are defined and discussed under the sections of this handbook pertaining to the topics to which they apply. In general, however, conversion factors are included in the tables of Section 51.5.

For an interesting and rather complete history see *British Weights and Measures*, London, 1910, by Sir C. M. Watson.

Metric (or SI) Units and Standards

The development of the SI system and the operations of the international bodies (BIPM, CIPM, and CGPM) having cognizance over weights and measures are described in appendices to ASTM’s *Standard for Metric Practice* (ASTM E380-82, American Society for Testing and Materials, Philadelphia, 1982).

Units of Length The *centimeter* (cm) is the *fundamental* unit of length in the cgs system. It equals, by definition, $\frac{1}{100}$ of a *meter* (m). The meter has been standardized by international agreement as 1,650,763.73 times the wavelength in vacuum of the unperturbed transition ($2p_{10}-5d_5$) of krypton 86. The *basic* meter for international comparisons is the *international prototype meter*, which is the distance, at zero degrees Centigrade, between two lines on a platinum–iridium bar located at the International Bureau of Weights and Measures at Sevres, France. This meter is the nearest to a duplicate ever constructed of the *original* meter, which was constructed and deposited in the Archives of the French Republic in 1799. The meter is very nearly equal to one ten-millionth of the distance, measured at sea level, from the equator to either pole.

An interesting history of the development of the *international prototype meter* (as well as the *international prototype kilogram*—see discussion on unit of mass) is given

by W. Parry, National Bureau of Standards, in *Merriman's Civil Engineers' Handbook*, as follows:

The use of the meter as the basis of geodetic surveys had become so general throughout Europe that a conference was called in Paris, France, in 1870, for the purpose of establishing a central bureau where the standards of the different countries could be compared. As a result of this conference an International Bureau of Weights and Measures was established near Paris in 1875, by the concurrent action of the principal nations of the world. One of the first tasks undertaken by the Bureau was the construction of exact copies of the meter and kilogram deposited in the Archives. Thirty-one standard meters of iridio-platinum and forty kilograms of the same alloy were constructed and carefully compared with the standards of the Archives and with one another. This great work was completed in 1889, and the meter and kilogram which agreed most nearly with the original standards were called international prototypes, and were deposited at the International Bureau, where they are maintained today subject to the authority of the International Committee on Weights and Measures. The remaining meters and kilograms were distributed by lot to the different nations which contributed to the support of the Bureau. The United States secured two copies of the meter and two copies of the kilogram, which are in the custody of the Bureau of Standards at Washington. One of the meters, known as No. 27, and one kilogram, No. 20, were selected as the United States standards, while the other meter and kilogram are used as secondary standards. It was the declared intention of the International Committee that the various national prototypes should be returned to the International Bureau at regular intervals for the purpose of recomparing them with the international standards and with one another. In this way all measurements based upon metric standards throughout the world are ultimately referred to the international meter and kilogram.

Unit of Capacity The *liter* (L) is the *standard* unit of capacity. It is defined as the volume of one kilogram of pure water at the temperature of maximum density (4°C) under a pressure of 76 cm of mercury. For all practical purposes, the liter may be regarded as the equivalent of the cubic decimeter, although the former is actually slightly greater, in the amount of less than three parts in one hundred thousand.

Unit of Mass The *gram* (g) is the *fundamental* unit of mass in the cgs system.⁴ It equals, by definition, $\frac{1}{1000}$ of a *kilogram* (kg), which is the *standard* unit of mass. The *basic* kilogram for international comparisons is the *international prototype kilogram*, which is a cylinder of platinum-iridium located at the International Bureau of Weights and Measures at Sèvres, France. This mass is the nearest to a duplicate ever constructed of the *original* kilogram, which was constructed and deposited in the Archives of the French Republic in 1799. The latter was made as nearly as possible equal to the mass of a cube of pure water at 4°C, the sides of the cube being one-tenth the length of the original meter.

An interesting history of the development of the *international prototype kilogram* was given under the discussion on units of length.

Weight versus Mass See discussion under this same subheading of the English units and standards.

⁴The slug of mass, which is extensively used by engineers and physicists, is (in the metric system) the mass to which an acceleration of one meter per second per second would be given by the application of a one-kilogram force. Under any gravity conditions, 1 slug of mass = 9.80665 kg of mass.

Auxiliary Fundamental Units Auxiliary fundamental units and their principal derived units are defined and discussed under the sections of this handbook pertaining to the topics to which they apply. In general, however, conversion factors are included in Tables 51.27–51.64.

51.4.8 Standard of Time

Unit of Time The *second* has been standardized by international agreement as $1/31,556,925.9747$ of the tropical year at 12 hr, ephemeris time, January 0 for the year 1900.0. (This definition has been retained for the time being as an astronomical time standard—the following atomic standard of time interval is 100 times more precise.) The second has been standardized by international agreement as the time taken for $9,192,631,770.0$ vibrations of the unperturbed hyperfine transition $4,0-3,0$ for the $^2S_{1/2}$ fundamental state of the cesium 133 atom. The ^{133}Cs standard has been adopted provisionally (see resolution 5 of the 12th General Conference of Weights and Measures, BIPM, Sèvres, Paris, Oct. 1964). A more accurate hydrogen maser standard may be available in the near future that is 100 times more accurate than the ^{133}Cs standard.

Measures of Time A *solar day* is measured by the rotation of the earth about its axis with respect to the sun. In *astronomical computations* and in *nautical time* the day commences at noon, and in the former it is counted throughout the 24 hr. In *civil computations* the day commences at midnight and is divided into two parts of 12 hr each.

A *solar year* is the time in which the earth makes one revolution around the sun. Its average time, called the *mean solar year*, is 365 days, 5 hr, 48 min, and 45.9747 sec, or nearly $365\frac{1}{4}$ days.

51.4.9 Force, Energy, and Power

51.4.9.1 Dynamical and Gravitational Units According to the use of two different dimension and unit systems, the dynamical (or physical, or “absolute”) system and the gravitational (or technical) system, two different sets of units of force, energy, power, and derived quantities are defined in both the English and the metric systems. *One dynamical unit of force* produces an acceleration of unity on unit standard mass. The *gravitational unit of force* is defined as that force required to give a unit standard mass an acceleration equal to that produced by the gravitational pull of the earth. As the acceleration due to gravity, g , varies with location and altitude,⁵ the gravitational unit of force is not constant, and, therefore, its relation to the dynamical unit of force will vary. By international agreement, the value $g_0 = 980.665 \text{ cm/sec}^2 = 32.1739 \text{ ft/sec}^2$ (British) has been chosen as the standard acceleration of gravity to make invariant the gravitational unit of force.

English Units

Units of Force The dynamical or physical unit of force is the *poundal*, defined as the force required to give a mass of one pound an acceleration of one foot per second per second.

⁵The variation of g with latitude ϕ and altitude H is given approximately by (ϕ in degrees, H in meters) $g = 978.039 (1 + 0.005295 \sin^2 \phi) - 0.000307H$. See *International Critical Tables*, Vol. 1, p. 395.

The *pound-force* (or weight of the pound mass) is the gravitational or technical unit of force. It is, by definition, the force required to give a mass of one pound an acceleration of 32.1739 ft/sec². If a force is measured by “weighing,” the result in pounds weight must be multiplied by g/g_0 , the ratio of local to standard acceleration of gravity, in order to obtain the absolute value in pound-force units. For engineering purposes this correction can usually be neglected.

Unit of Pressure This is defined as the unit of force acting upon a unit area. The most commonly employed unit is the *pound (force) per square inch*.

Standard atmospheric pressure is defined to be the force exerted by a column of mercury 760 mm (29.92 in.) high at 0°. This corresponds to 0.101325 MPa or 14.695 psi. Reference or fixed points for pressure calibration exist and are analogous to the phase changes used for temperature standards. These pressure references are based on phase changes or resistance jumps in selected materials.

Units of Work or Energy The foot-poundal is the physical unit of work or energy and is defined as the work done by a force of one poundal in moving a body through the distance of one foot in the direction of the force.

The *foot-pound (force)* is the technical unit of work or energy and is defined as the work required to raise a mass (or weight) of one pound through a vertical distance of one foot at standard acceleration of gravity g_0 . If measurements are made in places where the local value of the acceleration of gravity g is different from g_0 , a correction factor g/g_0 must be applied if the exact value of work or energy is desired.

The *British thermal unit* (Btu) is the quantity of heat required to raise the temperature of a one-pound mass of water either at 39°F (at its maximum density) or at 60°F and standard pressure through 1°F. The mean British thermal unit is defined as the $\frac{1}{180}$ part of the heat required to raise the temperature of a one-pound mass of water from 32 to 212°F at standard pressure. It is obvious that the reference temperature must be indicated with the unit used.

Units of Power Power is the time rate at which work is done. Its physical unit is the *foot-poundal per second*, its technical units are the *foot-pound (force) per second*, or the *British thermal unit per second*. The *horsepower* (hp or Hp) is defined as 33,000 ft-lb (force) per minute or 550 ft-lb (force) per second.

Units of Torque Torque is the effectiveness of a force to produce rotation. It is defined as the product of the force and the perpendicular distance from its line of action to the instantaneous center of rotation. Its physical unit is the poundal-foot, and its technical unit the pound (force)-foot. (Note the reversal of force and length units in the designation of the units of torque as compared with the units of energy or work.)

Metric Units

Units of Force The dynamical, or physical, unit of force is the *dyne*, defined as the force required to give a mass of one gram an acceleration of one centimeter per second per second.

The *newton* is the SI unit of force. It is the force required to give a mass of one kilogram an acceleration of one meter per second per second.

The *kilogram force* (or weight of the kilogram mass) is the gravitational or technical unit of force. It is, by definition, the force required to give a mass of one kilogram an acceleration of 980.665 cm/sec². If a force is measured by “weighing,” the result in kilograms weight must be multiplied by g/g_0 , the ratio of local to standard acceleration of gravity, in order to obtain the absolute value in kilogram-force units. For engineering purposes this correction can usually be neglected.

In the electrotechnical system of units the systematic unit of force is defined as the *joule per meter*, based on the fundamental definition of the joule. (See discussion on metric units of energy.)

Unit of Pressure This is defined as the unit of force acting upon a unit of area.

The *newton per square meter* is the SI unit of pressure and is called the *pascal*.

The *kilogram force per square meter* is the technical unit of pressure. With respect to correction for local gravity, see discussion on force versus weight.

Pressure is measured also by the height in centimeters of the column of water at 4°C, or of the column of mercury at 0°C, which it supports. (See conversion Table 51.41.)

The *normal atmosphere* (*at*), or the standard atmospheric pressure, is defined as the pressure exerted by a column of 76 cm of mercury at sea level and 0°C at standard acceleration of gravity g_0 . It is equal to 1.01321 bars or 1.0332 kg/cm² force and is used extensively in the engineering literature. Some confusion exists since the unit of 1 kg/cm² is occasionally called 1 practical atmosphere.

Units of Work or Energy The *joule* is the physical or so-called absolute unit of work or energy. It is defined as the work done by a force of one newton acting through the distance of one meter. A larger unit is the *theoretical* or “absolute” *joule* defined as 10⁷ ergs; it is a systematic unit in the practical electrical unit systems that is based on the theoretical unit systems. (See discussion on electrical units.)

The *international joule* is defined as the energy expended during one second by an electric current of one international ampere flowing through a resistance of one international ohm. (See discussion on electrical units.) The latest value of the international joule is equal to 1.000165 theoretical joules.⁶

The *kilowatt-hour* is the practical unit of energy in electrical metering. It is defined as a theoretical or an international unit (see definition of joule already given) and is equal to 3.6 megajoules.

The *meter-kilogram force* (commonly referred to as the kilogram-meter) is the technical unit of work or energy. It is defined as the work required to raise the mass (or weight) of one kilogram through a vertical distance of one meter at standard acceleration of gravity g_0 . If measurements are made in places where the local value of the acceleration of gravity g is different from g_0 , a correction factor g/g_0 has to be applied, if the exact value of work or energy is desired. (See discussion on force versus weight.)

The *gram-calorie* or small calorie is the physical unit of heat energy. It is defined as the quantity of heat required to raise the temperature of one gram mass of water either from 14.5 to 15.5°C or from 19.5 to 20.5°C at standard pressure. The two values are designated as 15 and 20°C cal, respectively. The mean gram-calorie is defined as $\frac{1}{100}$ part of the quantity of heat required to raise the temperature of one gram mass of water from 0 to 100°C at

⁶ *Mechanical Engineering*, Feb., 1930, pp. 122, 139.

standard pressure. The same definitions apply to kilogram-calorie, or large calorie, if the kilogram mass is used as reference standard mass.

The *Ostwald calorie* is the quantity of heat required to raise the temperature of one gram mass from 0 to 100°C. This unit is frequently used by electrochemists and is equal to 100 mean gram-calories.

The *international kilocalorie* or international steamtable calorie (IT cal) is defined as the $\frac{1}{860}$ part of the international kilowatt-hour. This new unit avoids any reference to the thermal properties of water and was recommended for international adoption at the first International Steam Table Conference (1929).⁷ Its value is very nearly equal to the mean kilocalorie, 1 IT cal = 1.00037 kilogram-calories (mean).

Units of Power Power is the time rate at which work is done. Its physical unit is the watt, defined as the power which gives rise to the production of energy at the rate of one joule per second.

The *international watt* is defined as the power expended by an electric current of one international ampere flowing through a resistance of one international ohm. (See discussion on electrical units.) The latest value of the international watt is equal to 1.000165 theoretical watts.⁸

The *electrical horsepower* is defined as 746 absolute watts and is commonly used in the United States and in England in rating electrical machinery.

The *meter-kilogram force per second* (commonly referred to as the kilogram-meter per second) is the technical unit of power. The *metric horsepower* is defined as 75 kg-m/sec and is the most common mechanical unit of power.

Units of Torque Torque is the effectiveness of a force to produce rotation. It is defined as the product of the force and the perpendicular distance from its line of action to the instantaneous center of rotation. Its physical unit is the dyne-centimeter, and its technical unit the kilogram force meter. (Note the reversal of force and length units in the designation of the units of torque as compared with the units of energy and work.)

51.4.10 Thermal Units and Standards Temperature

51.4.10.1 Definition of Temperature The *temperature* of a body may be defined as its thermal state considered from the standpoint of its ability to communicate heat to other bodies. When two bodies are placed in thermal communication, the one that loses heat to the other is said to be at the higher temperature.

51.4.10.2 Standard Temperatures Certain thermal states or “temperatures” may be reproduced and recognized by the fact that definite physical phenomena occur at these temperatures. Such thermal states are called “fixed points,” and they may, quite apart from any temperature scale, be specified by the physical phenomena characteristic of those temperatures. The two fundamental fixed points are the ice point and the steam point.

⁷ *Mechanical Engineering*, Nov., 1935, p. 710.

⁸ *Announcement of Changes in Electrical and Photometric Units*, Circular of National Bureau of Standards C459, Washington, DC, 1947.

The *ice point* is defined as the temperature of melting ice, which is realized experimentally as the temperature at which pure finely divided ice is in equilibrium with pure, air-saturated water under standard atmospheric pressure. The effect of increased pressure is to lower the freezing point to the extent of 0.007°C per atmosphere.

The *steam point* is defined as the temperature of condensing water vapor at standard atmospheric pressure, and it is realized experimentally by the use of a hypsometer so constructed as to avoid superheat of the vapor around the thermometer or contamination with air or other impurities. If the desired conditions have been attained, the observed temperature should be independent of the rate of heat supply to the boiler, except as this may affect the pressure within the hypsometer, and of the length of time the hypsometer has been in operation.

51.4.10.3 Definition of Temperature Scale The purpose of establishing a temperature scale is to assign a number to every thermal state or temperature and to provide a means for determining the temperature of any particular body.

A *temperature scale* may be defined by (1) selecting definite numbers for certain fixed points, (2) selecting some physical property of a definite substance that varies with temperature, and (3) selecting a mathematical law expressing temperatures on the scale in question in terms of the selected property of the thermometric substance. For example, on the Centigrade mercury-in-glass scale, the ice and steam points are numbered 0 and 100, respectively, the relative or “apparent” expansion of a volume of mercury enclosed in glass of a definite kind is the property used, and the mathematical relation used to express temperature on this scale is that equal increments of apparent volume of the mercury in this glass correspond to equal increments of temperature. If some other substance is substituted for mercury, or if glass of a different kind is used, another scale is obtained that agrees with it at 0 and 100 but not at other temperatures.

Although, in general, a temperature scale depends on the thermometric substance as well as on the expression for the temperature in terms of some property of this substance, Lord Kelvin has shown that, if the property selected is the availability of energy, the scale so defined is wholly independent of the substance and depends only on the mathematical relation chosen. Any scale so defined is known as a thermodynamic scale.

51.4.10.4 Kelvin Temperature Scale The temperature scale finally chosen by Lord Kelvin is the one on which the temperature interval from the ice point to the steam point is 100° and the ratio of the values of any two temperatures is equal to the ratio of the heat taken in to the heat rejected by a reversible thermodynamic engine working with a source and refrigerator at the higher and lower temperatures, respectively. On this scale, which is also known as the absolute thermodynamic scale, the lowest attainable temperature is 0 and the ice point is found experimentally to be 273.16° . The steam point therefore is 373.16° or 100° higher.

The *degree Kelvin* ($^{\circ}\text{K}$) or degree of absolute temperature is the absolute unit of temperature and is, for practical purposes, identical with the degree Centigrade ($^{\circ}\text{C}$) of the international temperature scale.

51.4.10.5 Thermodynamic Centigrade Scale This is derived by subtracting from the Kelvin scale a constant number of the proper magnitude to make the ice point 0° . On this

scale, therefore, the ice and steam points are 0 and 100°, respectively, and the so-called absolute zero is −273.16°.

51.4.10.6 International Centigrade Scale This is a practical representation of the thermodynamic Centigrade scale to such a degree of accuracy as is possible with present-day apparatus and methods. It was adopted at the General Conference on Weights and Measures at Sevres, France, in 1927 and is subject to revision and amendment as improved and more accurate methods of measurement are evolved.

The unit of temperature on the international scale is the *degree Centigrade* (°, or °C int) and is very nearly equal to $\frac{1}{100}$ the difference between the temperature of melting ice and the temperature of condensing water vapor under standard atmospheric pressure. (See discussion on metric units for pressure.)

The standard of the international temperature scale between −190 and +660°C is deduced from the electrical resistance of a standard platinum resistance thermometer by means of a formula connecting the resistance R_t at any temperature $t^\circ\text{C}$ within the above range with the resistance R_0 at 0°C. The purity of the platinum of which the thermometer is made should be such that the ratio R_t/R_0 for certain fixed temperatures is within specified limits. See also U.S. Bureau of Standards, *Journal of Research*, Vol. 1, p. 636, 1928.

The degree Centigrade is most widely used in scientific publications and increasingly also in the engineering literature. In many countries in Europe it is the common everyday temperature unit. The subdivision into a hundred degrees of the temperature interval between the ice point and the steam point was first used by Celsius, a German, in 1742; therefore, in the European literature “°C” is read “degree Celsius.”

51.4.10.7 Fahrenheit Temperature Scale This scale subdivides the temperature interval between the ice point and the steam point into 180 parts, one part of which is chosen as the unit of temperature and named *degree Fahrenheit* (°F). The ice point is assigned the value 32° F, so that the steam point has a temperature of 212°F.

The Fahrenheit unit of temperature is in common everyday use in the English-speaking countries. It was first introduced in England about 1665 by the physicist Fahrenheit; the choice of 32°F for the ice point has its explanation in the fact that Fahrenheit chose as zero the lowest temperature attainable by means of a salt–ice mixture.

51.4.10.8 Rankine Absolute Temperature Scale (°R) This is the thermodynamic Fahrenheit scale where absolute zero is 0°R (−459.69°F). The ice point is assigned the value 491.69°R and the steam point 671.69°R.

51.4.10.9 Relations Between Temperature Scales The following table shows the interrelations between the various temperature scales in the form of equations.

Temperature Interrelationships		
$x^\circ\text{F} =$	$9/5(t^\circ\text{K} - 273.16) + 32$	$9/5(t^\circ\text{C}) + 32$
$x^\circ\text{K} =$	$5/9(t^\circ\text{F} - 32) + 273.16$	$(t^\circ) + 273.16$
$x^\circ\text{C} =$	$5/9(t^\circ\text{F} - 32)$	$(t^\circ\text{K}) - 273.16$
$x^\circ\text{R} =$	$(t^\circ\text{F}) + 459.699$	$9/5(t^\circ\text{C}) + 491.69$

Here, X indicates the unknown number of chosen temperature units and t the known number of given temperature units:

51.4.11 Quantity of Heat and Some Derived Quantities

Units of Quantity of Heat Quantity of heat is defined as the energy transferred from one body to another by a thermal process, that is, by radiation or conduction. The units for the quantity of heat are the *British thermal unit* and the *calorie* as specific thermal units and the *erg* and *joule* as general physical units (see discussion on units of energy, metric and English system of units).

Thermal Capacity or Specific Heat of a Substance This is the quantity of heat required to produce a unit change in temperature in a unit of mass of the substance. The common English unit is the British thermal unit per degree Fahrenheit per pound mass (Btu per °F per lb); the usual metric unit is the gram-calorie per degree Centigrade per gram mass (cal per °C per g); and the general physical unit used in the scientific literature is the erg per degree Centigrade per gram mass (erg per °C per g). In the technical literature thermal capacity of a substance is often expressed in watt-seconds (or joules) per degree Centigrade per kilogram mass (W-sec per °C per kg) on account of the easy comparison with other technical units.

Calorimetric or Water Equivalent This is the quantity of heat required to produce a unit change in temperature of a body or system. It is numerically equivalent to the mass of water (in units as involved in the definition of the unit of quantity of heat used) that could be raised a unit temperature by the same total quantity of heat. The thermal capacity is expressed in British thermal units per degree Fahrenheit (Btu per °F), calories per degree Centigrade (cal per °C), or watt-seconds per degree Centigrade (W-sec per °C).

Thermal Conductivity This is the time rate of heat transfer through a unit area across a unit thickness per unit difference in temperature between the end surfaces. It is measured in British thermal units per second per degree Fahrenheit per inch thickness per square inch cross section (Btu per sec per °F per in. per in.²), in calories per second per degree Centigrade per centimeter thickness per square centimeter cross section (cal per sec per °C per cm per cm²), or in watts per degree Centigrade per meter thickness per square meter cross section (W per °C per m per m²).

Thermal Transmittance The surface coefficient of transfer is the time rate of heat emitted by a unit area for a unit difference in temperature between the surface in question and the surroundings. It is measured in British thermal units per second per degree Fahrenheit per square inch (Btu per sec per °F per in.²), in calories per second per degree Centigrade per square centimeter (cal per sec per °C per cm²), or in watts per degree Centigrade per square meter (W per °C per m²).

Joule Equivalent

The **Joule equivalent** is defined as the number of foot-pounds of energy per Btu. The numerical values for the various energy units used in the English and metric systems are shown in the table below.

	Joules “Absolute”	Foot- pounds (force)	Foot- poundals	Meter- kilogram (force)	Kilowatt- hour “international”
1 British thermal unit (Btu) (mean) =	1055.18	778.26	25.040	107.599	2.93019×10^{-4}
1 gram-calorie (cal) (mean) =	4.1873	3.0884	99.366	0.42699	1.16279×10^{-6}
1 international kilocalorie (IT cal) =	4187.3	3088.4	99.366	426.99	1.16279×10^{-3}
1 Ostwald calorie =	418.73	308.84	9936.6	42.699	1.16279×10^{-4}

51.4.12 Chemical Units and Standards

51.4.12.1 Atomic Weight The present definition of atomic weights (1961) is based on ^{12}C , which is the most abundant isotope of carbon and whose atomic weight is defined as exactly 12.

51.4.12.2 Standard Cell Potential A very large class of chemical reactions are characterized by the transfer of protons or electrons. Substances losing electrons in a reaction are said to be oxidized, those gaining electrons are said to be reduced. Many such reactions can be carried out in a galvanic cell that forms a natural basis for the concept of the half-cell, that is, the overall cell is conceptually the sum of two half-cells, one corresponding to each electrode. The half-cell potential measures the tendency of one reaction, for example, oxidation, to proceed at its electrode; the other half-cell of the pair measures the corresponding tendency for reduction to proceed at the other electrode. Measurable cell potentials are the sum of the two half-cell potentials. Standard cell potentials refer to the tendency of reactants in their standard state to form products in their standard states. The standard conditions are 1 *M* concentration for solutions, 101.325 kPa (1 atm) for gases, and for solids, their most stable form at 25°C.

Since half-cell potentials cannot be measured directly, numerical values are obtained by assigning the hydrogen gas–hydrogen ion half reaction the half-cell potential of 0 V. Thus, by a series of comparisons referred directly or indirectly to the standard hydrogen electrode, values for the strength of a number of oxidants or reductants can be obtained, and standard reduction potentials can be calculated from established values.

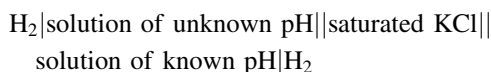
Standard cell potentials are meaningful only when they are calibrated against an electromotive force (emf) scale. To achieve an absolute value of emf, electrical quantities must be referred to the basic metric system of mechanical units. If the current unit ampere and the resistance unit ohm can be defined, then the volt may be defined by Ohm's law as the voltage drop across a resistor of one standard ohm (Ω) when passing one standard ampere (A) of current. In the ohm measurement, a resistance is compared to the reactance of an inductor or capacitor at a known frequency. This reactance is calculated from the measured dimensions and can be expressed in terms of the meter and second. The ampere determination measures the force between two interacting coils while they carry the test current. The force between the coils is opposed by the force of gravity acting on a known mass; hence, the ampere can be defined in terms of the meter, kilogram, and second. Such a means of establishing a reference voltage is inconvenient for frequent use and reference is made to a previously calibrated standard cell.

Ideally, a standard cell is constructed simply and is characterized by a high constancy of emf, a low temperature coefficient of emf, and an emf close to 1 v. The Weston cell,

which uses a standard cadmium sulfate electrolyte and electrodes of cadmium amalgam and a paste of mercury and mercurous sulfate, essentially meets these conditions. The voltage of the cell is 1.0183 V at 20°C. The alternating current (ac) Josephson effect, which relates the frequency of a superconducting oscillator to the potential difference between two superconducting components, is used by the National Bureau of Standards to maintain the unit of emf, but the definition of the volt remains the Ω/A derivation described.

51.4.12.3 Concentration The basic unit of concentration in chemistry is the mole, which is the amount of substance that contains as many entities, for example, atoms, molecules, ions, electrons, protons, and so on, as there are atoms in 12 g of ^{12}C , that is, Avogadro's number $N_A = 6.022045 \times 10^{23}$. Solution concentrations are expressed on either a weight or volume basis. *Molality* is the concentration of a solution in terms of the number of moles of solute per kilogram of solvent. *Molarity* is the concentration of a solution in terms of the number of moles of solute per liter of solution.

A particular concentration measure of acidity of aqueous solutions is pH, which, usually, is regarded as the common logarithm of the reciprocal of the hydrogen ion concentration (qv). More precisely, the potential difference of the hydrogen electrode in normal acid and in normal alkali solution (-0.828 V at 25°C) is divided into 14 equal parts or pH units; each pH unit is 0.0591 V. Operationally, pH is defined by $\text{pH} = \text{pH}(\text{soln}) + E/K$, where E is the emf of the cell:



and $K = 2.303 RT/F$, where R is the gas constant, 8.314 J/(mol/K) [1.987 cal/(mol · K)], T is the absolute temperature, and F is the value of the Faraday, $9.64845 \times 10^4 \text{C/mol}$. pH usually is equated to the negative logarithm of the hydrogen ion activity, although there are differences between these two quantities outside the pH range 4.0–9.2:

$$-\log q_{\text{H}^+} m_{\text{H}^+} = \begin{cases} \text{pH} + 0.014 (\text{pH} - 9.2) & \text{for pH} > 9.2 \\ \text{pH} + 0.009 (4.0 - \text{pH}) & \text{for pH} < 4.0 \end{cases}$$

51.4.13 Theoretical, or Absolute, Electrical Units

With the general adoption of SI as the form of metric system that is preferred for all applications, further use of cgs units of electricity and magnetism is deprecated. Nonetheless, for historical reasons as well as for comprehensiveness, a brief review is included in this section and section 4.10.

The definitions of the *theoretical*, or “*absolute*,” **units** are based on a particular choice of the numerical value of either k_e , the constant in Coulomb's, electrostatic force law, or k_m , the constant in Ampere's electrodynamic force law. The designation absolute units is generally used because of historical tradition; an interesting account of the history can be found in Glazebrook's *Handbook for Applied Physics*, Vol. II, “Electricity,” pp. 211 ff., 1922. Because of the theoretical background of the unit definitions, they have also been designated as “theoretical” units, which is in good contradistinction to practical units based on physical standards.

Theoretical Electrostatic Units

The *theoretical electrostatic units* are based on the cgs system of mechanical units and the choice of the numerical value unity for k_{ev} in Coulomb's law. They are frequently referred to as the *cgs electrostatic units*, but no specific unit names are available. In order to avoid the cumbersome writing, for example, one "theoretical electrostatic unit of charge," it had been proposed to use the theoretical "practical" unit names and prefix them with either stat or E.S. as, for example, statcoulomb, or E.S. coulomb. The first alternative will be used here.

The *absolute dielectric constant (permittivity)* of free space is the reciprocal of the Coulomb constant k_{ev} and is chosen as the fourth fundamental quantity in the theoretical electrostatic system of units. Its numerical value is defined as unity, and it is identical with one statfarad per centimeter if use is made of prefixing the corresponding unit of the "practical" series.

The theoretical electrostatic unit of *charge*, or the *statcoulomb*, is defined as the quantity of electricity that, when concentrated at a point and placed at one centimeter distance from an equal quantity of electricity similarly concentrated, will experience a mechanical force of one dyne in free space. An alternative definition, based on the concept of field lines, gives the theoretical electrostatic unit of charge as a positive charge from which in free space exactly 4π displacement lines emerge.

The theoretical electrostatic unit of *displacement flux (dielectric flux)* is the "line of displacement flux," or $\frac{1}{4}\pi$ of the theoretical electrostatic unit of charge. This definition provides the basis for graphical field mapping insofar as it gives a definite rule for the selection of displacement lines to represent the distribution of the field quantitatively.

The theoretical electrostatic unit of *displacement*, or *dielectric flux density*, is chosen as one displacement line per square centimeter area perpendicular to the direction of the displacement lines. It can be given also as $\frac{1}{4}\pi$ statcoulomb per square centimeter (according to Gauss's law). In isotropic media the displacement has the same direction as the potential gradient, and the surfaces perpendicular to the field lines become the equipotential surfaces; the theoretical electrostatic unit of displacement can then be defined as one displacement line per square centimeter of equipotential surface.

The theoretical electrostatic unit of *electrostatic potential*, or the *statvolt*, is defined as existing at a point in an electrostatic field, if the work done to bring the theoretical electrostatic unit of charge, or the statcoulomb, from infinity to this point equals one erg. This customary definition implies, however, that the potential vanishes at infinite distances and has, therefore, only restricted validity. As it is fundamentally impossible to give absolute values of potential, the use of potential difference and its unit (see below) should be preferred.

The theoretical electrostatic unit of *electrical potential difference* or *voltage*, is the *statvolt* and is defined as existing between two points in space if the work done to bring the theoretical electrostatic unit of charge, or the statcoulomb, from one of these points to the other equals one erg. Potential difference is counted positive in the direction in which a negative quantity of electricity would be moved by the electrostatic field.

The theoretical electrostatic unit of *capacitance*, or the *statfarad*, is defined as the capacitance that maintains an electrical potential difference of one statvolt between two conductors charged with equal and opposite electrical charges of one statcoulomb. In the older literature, the cgs electrostatic unit of capacitance is identified with the "centimeter"; this was replaced by statfarad to avoid confusion.

The theoretical electrostatic unit of *electric potential gradient*, or *field strength* (field intensity), is defined to exist at a point in an electric field if the mechanical force exerted

upon the theoretical electrostatic unit of charge concentrated at this point is equal to one dyne. It is expressed as one statvolt per centimeter.

The theoretical electrostatic unit of *current*, or the *statampere*, is defined as the time rate of transfer of the theoretical electrostatic unit of charge and is identical with the statcoulomb per second.

The theoretical electrostatic unit of *electrical resistance*, or the *statohm*, is defined as the resistance of a conductor in which a current of one statampere is produced if a potential difference of one statvolt is applied at its ends.

The theoretical electrostatic unit of *electromotive force (emf)* is defined as equivalent to the theoretical electrostatic unit of potential difference if it produces a current of one statampere in a conductor of one statohm resistance. It is identical with the statvolt but, according to its concept, requires an independent definition.

The theoretical electrostatic unit of *magnetic intensity* is defined as the magnetic intensity at the center of a circle of 4π centimeters diameter in which a current of one statampere is flowing. This unit is equal to 4π statamperes per centimeter but has no name as the factor 4π excludes the possibility of using the prefixed “practical” unit name.

The theoretical electrostatic unit of *magnetic flux*, or the *statweber*, is defined as the magnetic flux whose time rate of change through a linear conductor loop (linear conductor is used to designate a conductor of infinitely small cross section) produces in this loop an emf of one statvolt.

The theoretical electrostatic unit of *magnetic flux density*, or *induction*, is defined as the electrostatic unit of magnetic flux per square centimeter area, or the statweber per square centimeter.

The *absolute magnetic permeability of free space* is defined as the ratio of magnetic induction to the magnetic intensity. Its unit is the stathenry per centimeter as a derived unit.

The theoretical electrostatic unit of *inductance*, or the *stathenry*, is defined as connected with a conductor loop carrying a steady current of one statampere that produces a magnetic flux of one statweber. A more general definition, applicable to varying fields with nonlinear relation between magnetic flux and current, gives the stathenry as connected with a conductor loop in which a time rate of change in the current of one statcoulomb produces a time rate of change in the magnetic flux of one statweber per second.

Theoretical Electromagnetic Units

The *theoretical electromagnetic units* are based on the cgs system of mechanical units and Coulomb’s law of mechanical force action between two isolated magnetic quantities m_1 and m_2 (approximately true for very long bar magnets) that must be written as

$$F_m = \frac{k_m m_1 m_2}{2 r^2} \quad (51.7)$$

where k_m is the proportionality constant of Ampere’s law for force action between parallel currents that is more basic, and amenable to much more accurate measurement, than (51.7). The factor $\frac{1}{2}$ appears here because of the three-dimensional character of the field distribution around point magnets as compared with the two-dimensional field of two parallel currents.

The theoretical electromagnetic units are obtained by defining the numerical value of $k_m/2$ (for vacuum) as unity; they are frequently referred to as the cgs electromagnetic units. Only a few specific unit names are available. In order to avoid cumbersome writing,

for example, one “theoretical electromagnetic unit of charge,” it had been proposed to use the theoretical “practical” unit names and prefix them with either ab- or E.M. as, for example, abcoulomb, or E.M. coulomb. The first alternative will be used here.

The *absolute magnetic permeability* of free space is the value $k_{mv}/2$ in (51.7) and is chosen as the fourth fundamental quantity in the theoretical electromagnetic system of units. Its numerical value is assumed as unity, and it is identical with one abhenry per centimeter if use is made of prefixing the corresponding unit of the “practical” series.

The theoretical electromagnetic unit of *magnetic quantity* is defined as the magnetic quantity that, when concentrated at a point and placed at one centimeter distance from an equal magnetic quantity similarly concentrated, will experience a mechanical force of one dyne in free space. An alternative definition, based on the concept of magnetic intensity lines, gives the theoretical electromagnetic unit of magnetic quantity as a positive magnetic quantity from which, in free space, exactly 4π magnetic intensity lines emerge.

The theoretical electromagnetic unit of *magnetic moment* is defined as the magnetic moment possessed by a magnet formed by two theoretical electromagnetic units of magnetic quantity of opposite sign, concentrated at two points one centimeter apart. As a vector, its positive direction is defined from the negative to the positive magnetic quantity along the center line.

The theoretical electromagnetic unit of *magnetic induction (magnetic flux density)*, or the *gauss*, is defined to exist at a point in a magnetic field, if the mechanical torque exerted upon a magnet with theoretical electromagnetic unit of magnetic moment and directed perpendicular to the magnetic field is equal to one dyne-centimeter. The lines to which the vector of magnetic induction is tangent at every point are called induction lines or magnetic flux lines; on the basis of this flux concept, magnetic induction is identical with magnetic flux density.

The theoretical electromagnetic unit of *magnetic flux*, or the *maxwell*, is the “field line” or line of magnetic induction. In free space, the theoretical electromagnetic unit of magnetic quantity issues 4π induction lines; the unit of magnetic flux, or the maxwell, is then $1/4\pi$ of the theoretical electromagnetic unit of magnetic quantity times the absolute permeability of free space.

The theoretical electromagnetic unit of *magnetic intensity (magnetizing force)*, or the *oersted*, is defined to exist at a point in a magnetic field in free space where one measures a magnetic induction of one gauss.

The theoretical electromagnetic unit of *current*, or the *abampere*, is defined as the current that flows in a circle of one centimeter diameter and produces at the center of this circle a magnetic intensity of one oersted.

The theoretical electromagnetic unit of *inductance*, or the *abhenry*, is defined as connected with a conductor loop in which a time rate of change of one maxwell per second in the magnetic flux produces a time rate of change in the current of one abampere per second. In the older literature, the cgs electromagnetic unit of inductance is identified with the “centimeter”; this should be replaced by a henry to avoid confusion.

The theoretical electromagnetic unit of *magnetomotive force (mmf)* is defined as the magnetic driving force produced by a conductor loop carrying a steady current of $\frac{1}{4}\pi$ abamperes; it has the name one gilbert. The concept of magnetomotive force as the driving force in a “magnetic circuit” permits an alternative definition of the gilbert as the magnetomotive force that produces a uniform magnetic intensity of one oersted over a length of one centimeter in the magnetic circuit. Obviously, one gilbert equals one oersted-centimeter.

The theoretical electromagnetic unit of *magneto-static potential* is defined as the potential existing at a point in a magnetic field if the work done to bring the theoretical

electromagnetic unit of magnetic quantity from infinity to this point equals one erg. This customary definition implies, however, that the potential vanishes at infinite distances, and the definition has therefore only restricted validity. The unit, thus defined, is identical with one gilbert. The difference in magnetostatic potential between any two points is usually called magnetomotive force (mmf).

The theoretical electromagnetic unit of *reluctance* is defined as the reluctance of a magnetic circuit in which a magnetomotive force of one gilbert produces a magnetic flux of one maxwell.

The theoretical electromagnetic unit of *electric charge*, or the *abcoulomb*, is defined as the quantity of electricity that passes through any section of an electric circuit in one second if the current is one abampere.

The theoretical electromagnetic unit of *displacement flux* (*dielectric flux*) is the "line of displacement flux," or $\frac{1}{4}\pi$ of the theoretical electromagnetic unit of electric charge. This definition provides the basis for graphical field mapping insofar as it gives a definite rule for the selection of displacement lines to represent the character of the field.

The theoretical electromagnetic unit of *displacement*, or *dielectric flux density*, is chosen as one displacement line per square centimeter area perpendicular to the direction of the displacement lines. It can also be given as $\frac{1}{4}\pi$ abcoulombs per square centimeter (according to Gauss's law). In isotropic media the theoretical electromagnetic unit of displacement can be defined as one displacement line per square centimeter of equipotential surface. (See discussion on theoretical electrostatic unit of displacement.)

The theoretical electromagnetic unit of *electrical potential difference*, or *voltage*, is the *abvolt* and is defined as the potential difference existing between two points in space if the work done in bringing the theoretical electromagnetic unit of charge, or the abcoulomb, from one of these points to the other equals one erg. Potential difference is counted positive in the direction in which a negative quantity of electricity would be moved by the electrostatic field.

The theoretical electromagnetic unit of *capacitance*, or the *abfarad*, is defined as the capacitance that maintains an electrical potential difference of one abvolt between two conductors charged with equal and opposite electrical quantities of one abcoulomb.

The theoretical electromagnetic unit of *potential gradient*, or *field strength* (field intensity), is defined to exist at a point in an electric field if the mechanical force exerted upon the theoretical electromagnetic unit of charge concentrated at this point is equal to one dyne. It is expressed as one abvolt per centimeter.

The theoretical electromagnetic unit of *resistance*, or the *abohm*, is defined as the resistance of a conductor in which a current of one abampere is produced if a potential difference of one abvolt is applied at its ends.

The theoretical electromagnetic unit of *electromotive force* (*emf*) is defined as the electromotive force acting in an electric circuit in which a current of one abampere is flowing and electrical energy is converted into other kinds of energy at the rate of one erg per second. This unit is identical with the abvolt.

The *absolute dielectric constant of free space* is defined as the ratio of displacement to the electric field intensity. Its unit is the abfarad per centimeter, a derived unit.

Theoretical Electrodynamic Units

The theoretical electrodynamic units are based on the cgs system of mechanical units and are therefore frequently referred to as the *cgs electrodynamic units*. In contradistinction to the theoretical electromagnetic units, these units are derived from a significant experimental law, Ampere's experiment on the mechanical force between two parallel currents.

The units as proposed by Ampere and used by W. Weber differ from the electromagnetic units by factors of 2 and multiples thereof. They can be made to coincide with the theoretical electromagnetic units by proper definition of the fundamental unit of current. Some of the important definitions will be given for this latter case only.

For the *absolute magnetic permeability* of free space, see discussion on theoretical electromagnetic units.

The theoretical electrodynamic unit of *current*, or the *abampere*, is defined as the current flowing in a circuit consisting of two infinitely long parallel wires one centimeter apart when the electrodynamic force of repulsion between the two wires is *two* dynes per centimeter length in free space. If the more natural choice of *one* dyne per centimeter length is made, the original proposal of Ampere is obtained and the unit of current becomes $1/\sqrt{2}$ abampere.

The theoretical electrodynamic unit of *magnetic induction* is defined as the magnetic induction inducing an electromotive force of one abvolt in a conductor of one centimeter length and moving with a velocity of one centimeter per second if the conductor, its velocity, and the magnetic induction are mutually perpendicular. The unit thus defined is called one gauss.

The theoretical electrodynamic unit of *magnetic flux*, or the *maxwell*, is defined as the magnetic flux represented by a uniform magnetic induction of one gauss over an area of one square centimeter perpendicular to the direction of the magnetic induction.

The theoretical electrodynamic unit of *magnetic intensity*, or the *oersted*, is defined as the magnetic intensity at the center of a circle of 4π centimeters diameter in which a current of one abampere is flowing.

All the other unit definitions, which do not pertain to magnetic quantities, are identical with the definitions for the theoretical electromagnetic units.

51.4.14 Internationally Adopted Electrical Units and Standards

In October 1946, at Paris, the International Committee on Weights and Measures decided to abandon the so-called international practical units based on physical standards (see below) and to adopt effective January 1, 1948, the so-called absolute practical units for international use.

Adopted Absolute Practical Units

By a series of international actions, the “absolute” practical electrical units are defined as exact powers of 10 of corresponding theoretical electrodynamic and electromagnetic units because they are based on the choice of the proportionality constant in Ampere’s law for free space as $k_{mv} = 2 \times 10^{-7}$ H/m.

The *absolute practical unit of current*, or the absolute is defined as the current flowing in a circuit consisting of two very long parallel thin wires spaced 1 m apart in free space if the electrodynamic force action between the wires is 2×10^{-7} N = 0.02 dyne per meter length. It is 10^{-1} of the theoretical or absolute electrodynamic or electromagnetic unit of current and was adopted internationally in 1881.

The absolute practical unit of *electric charge*, or the *absolute coulomb*, is defined as the quantity of electricity that passes through a cross-sectional surface in one second if the current is one absolute ampere. It is 10^{-1} of the theoretical or absolute electromagnetic unit of electric charge and was adopted internationally in 1881.

The absolute practical unit of *electric potential difference*, or the *absolute volt*, is defined as the potential difference existing between two points in space if the work done in bringing an electric charge of one absolute coulomb from one of these points to another is equal to one absolute joule = 10^7 ergs. It is 10^8 of the theoretical or absolute electromagnetic unit of potential difference and was adopted internationally in 1881.

The absolute practical unit of *resistance*, or the *absolute ohm*, is defined as the resistance of a conductor in which a current of one absolute ampere is produced if a potential difference of one absolute volt is applied at its ends. It is 10^9 of the theoretical or absolute electromagnetic unit of resistance and was adopted internationally in 1881.

The absolute practical unit of *magnetic flux*, or the *absolute weber*, is defined to be linked with a closed loop of thin wire of total resistance one absolute ohm if upon removing the wire loop from the magnetic field a total charge of one absolute coulomb is passed through any cross section of the wire. It is 10^8 of the theoretical or absolute electromagnetic unit of magnetic flux, the maxwell, and was adopted internationally in 1933.

The absolute practical unit of *inductance*, or the *absolute henry*, is defined as connected with a closed loop of thin wire in which a time rate of change of one absolute weber per second in the magnetic flux produces a time rate of change in the current of one absolute ampere. It is 10^9 of the theoretical or absolute electromagnetic unit of inductance and was adopted internationally in 1893.

The absolute practical unit of *capacitance*, or the *absolute farad*, is defined as the capacitance that maintains an electric potential difference of one absolute volt between two conductors charged with equal and opposite electrical quantities of one coulomb. It is 10^{-9} of the theoretical or absolute electromagnetic unit of capacitance and was adopted internationally in 1881.

Abandoned International Practical Units

The International System of electrical and magnetic units is a system for electrical and magnetic quantities that takes as the four fundamental quantities resistance, current, length, and time. The units of resistance and current are defined by physical standards that were originally aimed to be exact replicas of the “absolute” practical units, namely the absolute ampere and the absolute ohm. On account of long-range variations in the physical standards, it proved impossible to rely upon them for international use and they recently have been replaced by the absolute practical units.

The international practical standards are defined as follows:

The *international ohm* is the resistance at 0°C of a column of mercury of uniform cross section having a length of 106.300 cm and a mass of 14.4521 g.

The *international ampere* is defined as the current that will deposit silver at the rate of 0.00111800 g/sec.

From these fundamental units, all other electrical and magnetic units can be defined in a manner similar to the absolute practical units. Because of the inconvenience of the silver voltameter as a standard, the various national laboratories actually used a volt, defining its value in terms of the other two standards.

At its conference in October 1946 in Paris, the International Committee on Weights and Measures accepted as the best relations between the international and the absolute practical units the following:

$$1 \text{ mean international ohm} = 1.00049 \text{ absolute ohms}$$

$$1 \text{ mean international volt} = 1.00034 \text{ absolute volts}$$

These mean values are the averages of values measured in six different national laboratories. On the basis of these mean values, the specific unit relation for converting international units appearing on certificates of the National Bureau of Standards, Washington, DC, into absolute practical units are as follows:

1 international ampere	= 0.999835 absolute ampere
1 international coulomb	= 0.999835 absolute coulomb
1 international henry	= 1.000495 absolute henries
1 international farad	= 0.999505 absolute farad
1 international watt	= 1.000165 absolute watts
1 international joule	= 1.000165 absolute joules

BIBLIOGRAPHY FOR UNITS AND MEASUREMENTS

Cohen, E. R., and Taylor, B. N., “The 1986 Adjustment of the Fundamental Physical Constants,” *Report of the CODATA Task Group on Fundamental Constants, November 1986*, CODATA Bulletin No. 63, International Council of Scientific Unions, Committee on Data for Science and Technology, Pergamon, 1986.

Hvistendahl, H. S., *Engineering Units and Physical Quantities*, Macmillan, London, 1964.

Jerrard, H. G., and McNeill, D. B., *A Dictionary of Scientific Units*, 2nd ed., Chapman & Hall, London, 1964.

Letter Symbols for Units of Measurement, ANSI/IEEE Std. 260-1978, Institute of Electrical and Electronic Engineers, New York, 1978.

Quantities, Units, Symbols, Conversion Factors, and Conversion Tables, ISO Reference 31, 15 sections, International Organization for Standardization Geneva, 1973–1979.

Standard for Metric Practice, ASTM E 380-82, American Society for Testing and Materials, Philadelphia, 1982.

Young, L., *System of Units in Electricity and Magnetism*, Oliver and Boyd, Edinburgh, 1969.

Young, L., *Research Concerning Metrology and Fundamental Constants*, National Academy Press, Washington, DC, 1983.

51.5 TABLES OF CONVERSION FACTORS⁹

J. G. Brainerd
(revised and extended by J. H. Westbrook)

TABLE 51.27 Temperature Conversion

$^{\circ}\text{F} = (^{\circ}\text{C} \times \frac{9}{5}) + 32 = (^{\circ}\text{C} + 40) \times \frac{9}{5} - 40$
$^{\circ}\text{C} = (^{\circ}\text{F} - 32) \times \frac{5}{9} = (^{\circ}\text{F} + 40) \times \frac{5}{9} - 40$
$^{\circ}\text{R} = ^{\circ}\text{F} + 459.69$
$^{\circ}\text{K} = \text{C} + 273.16$

⁹ Boldface units in Tables 51.28–51.63 are SI.

TABLE 51.28 Length [L]

<div>Multiply Number of → by → to Obtain ↓</div>										
	Centimeters	Feet	Inches	Kilometers	Nautical Miles	Meters	Mils	Miles	Millimeters	Yards
	1	30.48	2.540	10 ⁵	1.853 × 10 ⁵	100	2.540 × 10 ⁻³	1.609 × 10 ⁵	0.1	91.44
	3.281 × 10 ⁻²	1	8.333 × 10 ⁻²	3281	6080.27	3.281	8.333 × 10 ⁻⁵	5280	3.281 × 10 ⁻³	3
	0.3937	12	1	3.937 × 10 ⁴	7.296 × 10 ⁴	39.37	0.001	6.336 × 10 ⁴	3.937 × 10 ⁻²	36
	10 ⁻⁵	3.048 × 10 ⁻⁴	2.540 × 10 ⁻⁵	1	1.853	0.001	2.540 × 10 ⁻⁸	1.609	10 ⁻⁶	9.144 × 10 ⁻⁴
		1.645 × 10 ⁻⁴	—	0.5396	1	5.396 × 10 ⁻⁴		0.8684	—	4.934 × 10 ⁻⁴
	0.01	0.3048	2.540 × 10 ⁻²	1000	1853	1		1609	0.001	0.9144
	393.7	1.2 × 10 ⁴	1000	3.937 × 10 ⁷	—	3.937 × 10 ⁴	1	—	39.37	3.6 × 10 ⁴
	6.214 × 10 ⁻⁶	1.894 × 10 ⁻⁴	1.578 × 10 ⁻⁵	0.6214	1.1516	6.214 × 10 ⁻⁴	—	1	6.214 × 10 ⁻⁷	5.682 × 10 ⁻⁴
10	304.8	25.40	10 ⁶	—	1000	2.540 × 10 ⁻²	—	1	914.4	
1.094 × 10 ⁻²	0.3333	2.778 × 10 ⁻²	1094	2027	1.094	2.778 × 10 ⁻⁵	1760	1.094 × 10 ⁻³	1	

Length

Land Measure

- 7.92 inches = 1 link
- 25 links = 1 rod = 16.5 feet = 5.5 yards (1 rod = 1 pole = 1 perch)
- 4 rods = 1 chain (Gunther's) = 66 feet = 22 yards = 100 links
- 10 chains = 1 furlong = 660 feet = 220 yards = 1000 links = 40 rods
- 8 furlongs = 1 mile = 5280 feet = 1760 yards = 8000 links = 320 rods = 80 chains

Ropes and Cables

- 2 yards = 1 fathom 120 fathoms = 1 cable length

Nautical Measure

- 6080.27 feet = 1 nautical mile = 1.15156 statute miles
- 3 nautical miles = 1 league (U.S.) 3 statute miles = 1 league (Gr. Britain)

(*Note:* A nautical mile is the length of a minute of longitude of the earth at the equator at sea level. The British Admiralty uses the round figure of 6080 feet. The word “knot” is used to denote “nautical miles per hour.”)

Miscellaneous

- 3 inches = 1 palm 9 inches = 1 span
- 4 inches = 1 hand 2½ feet = 1 military pace

TABLE 51.29 Area [L^2]

<div>Multiply Number of → to Obtain ↓ by →</div>										
	Acres	Circular Mils	Square Centimeters	Square Feet	Square Inches	Square Kilometers	Square Meters	Square Miles	Square Millimeters	Square Yards
Acres	1	—	—	2.296×10^{-5}	—	247.1	2.471×10^{-4}	640	—	2.066×10^{-4}
Circular Mils	—	1	1.973×10^5	1.833×10^8	1.273×10^6	—	1.973×10^9	—	1973	—
Square Centimeters	—	5.067×10^{-6}	1	929.0	6.452	10^{10}	10^4	2.590×10^{10}	0.01	8361
Square Feet	4.356×10^4	—	1.076×10^{-3}	1	6.944×10^{-3}	1.076×10^7	10.76	2.788×10^7	1.076×10^{-5}	9
Square Inches	6,272,640	7.854×10^{-7}	0.1550	144	1	1.550×10^9	1550	4.015×10^9	1.550×10^{-3}	1296
Square Kilometers	4.047×10^{-3}	—	10^{-10}	9.290×10^{-8}	6.452×10^{-10}	1	10^{-6}	2,590	10^{-12}	8.361×10^{-7}
Square Meters	4047	—	0.0001	9.290×10^{-2}	6.452×10^{-4}	10^6	1	2.590×10^6	10^{-6}	0.8361
Square Miles	1.562×10^{-3}	—	3.861×10^{-11}	3.587×10^{-8}	—	0.3861	3.861×10^{-7}	1	3.861×10^{-13}	3.228×10^{-7}
Square Millimeters	—	5.067×10^{-4}	100	9.290×10^4	645.2	10^{12}	10^6	—	1	8.361×10^5
Square Yards	4840	—	1.196×10^{-4}	0.1111	7.716×10^{-4}	1.196×10^6	1.196	3.098×10^6	1.196×10^{-6}	1

Area

Land Measure

- $30\frac{1}{4}$ square yards = 1 square rod = $272\frac{1}{4}$ square feet
- 16 square rods = 1 square chain = 484 square yards = 4356 square feet
- $2\frac{1}{2}$ square chains = 1 rod = 40 square rods = 1210 square yards
 - 4 rods = 1 acre = 10 square chains = 160 square rods
- 640 acres = 1 square mile = 2560 rods = 102, 400 square rods
- 1 section of land = 1 square mile; 1 quarter section = 160 acres

Architect' s Measure

100 square feet = 1 square

Circular Inch and Circular Mil A circular inch is the area of a circle 1 inch in diameter = 0.7854 square inch

1square inch = 1.2732circular inches

A circular mil is the area of a circle 1 mil (or 0.001 inch) in diameter = 0.7854 square mil

- 1 square mil = 1.2732 circular mils
- 1 circular inch = 10^6 circular mils = 0.7854×10^6 square mils
- 1 square inch = 1.2732×10^6 circular mils = 10^6 square mils

TABLE 51.30 Volume [L^3]

<div>Multiply Number of \rightarrow by \rightarrow to Obtain \downarrow</div>	Bushels (Dry)	Cubic Centimeters	Cubic Feet	Cubic Inches	Cubic Meters	Cubic Yards	Gallons (Liquid)	Liters	Pints (Liquid)	Quarts (Liquid)
Bushels (Dry)	1	—	0.8036	4.651×10^{-4}	28.38	—	—	2.838×10^{-2}	—	—
Cubic Centimeters	3.524×10^4	1	2.832×10^4	16.39	10^6	7.646×10^5	3785	1000	473.2	946.4
Cubic Feet	1,2445	3.531×10^{-5}	1	5.787×10^{-4}	35.31	27	0.1337	3.531×10^{-2}	1.671×10^{-2}	3.342×10^{-2}
Cubic Inches	2150.4	6.102×10^{-2}	1728	1	6.102×10^4	46.656	231	61.02	28.87	57.75
Cubic Meters	3.524×10^{-2}	10^{-6}	2.832×10^{-2}	1.639×10^{-5}	1	0.7646	3.785×10^{-3}	0.001	4.732×10^{-4}	9.464×10^{-4}
Cubic Yards	—	1.308×10^{-6}	3.704×10^{-2}	2.143×10^{-5}	1,308	1	4.951×10^{-3}	1.308×10^{-3}	6.189×10^{-4}	1.238×10^{-3}
Gallons (Liquid)	—	2.642×10^{-4}	7.481	4.329×10^{-3}	264.2	202.0	1	0.2642	0.125	0.25
Liters	35.24	0.001	28.32	1.639×10^{-2}	1000	764.6	3.785	1	0.4732	0.9464
Pints (Liquid)	—	2.113×10^{-3}	59.84	3.463×10^{-2}	2113	1616	8	2.113	1	2
Quarts (Liquid)	—	1.057×10^{-3}	29.92	1.732×10^{-2}	1057	807.9	4	1.057	0.5	1

Volume**Cubic Measure**

$$\begin{aligned}
 1 \text{ cord of wood} &= \text{pile cut 4 feet long piled 4 feet} \\
 &\quad \text{high and 8 feet on the ground} \\
 &= 128 \text{ cubic feet} \\
 1 \text{ perch of stone} &= \text{quantity } 1\frac{1}{2} \text{ feet thick,} \\
 &\quad 1 \text{ foot high, and } 16\frac{1}{2} \text{ feet long} \\
 &= 24\frac{3}{4} \text{ cubic feet}
 \end{aligned}$$

(*Note:* A perch of stone is, however, often computed differently in different localities; thus, in most if not all of the states west of the Mississippi, stonemasons figure rubble by the perch of $16\frac{1}{2}$ cubic feet. In Philadelphia, 22 cubic feet is called a perch. In Chicago, stone is measured by the cord of 100 cubic feet. Check should be made against local practice.)

Board Measure In board measure, boards are assumed to be one inch in thickness. Therefore, feet board measure of a stick of square timber = length in feet \times breadth in feet \times thickness in inches.

Shipping Measure For register tonnage or measurement of the entire internal capacity of a vessel, it is arbitrarily assumed, to facilitate computation, that

$$100 \text{ cubic feet} = 1 \text{ register ton}$$

For the measurement of cargo:

$$\begin{aligned}
 40 \text{ cubic feet} &= 1 \text{ U.S. shipping ton} \\
 &= 32.143 \text{ U.S. bushels} \\
 42 \text{ cubic feet} &= 1 \text{ British shipping ton} \\
 &= 32.703 \text{ Imperial bushels}
 \end{aligned}$$

Dry Measure One U.S. Winchester bushel contains 1.2445 cubic feet or 2150.42 cubic inches. It holds 77.601 pounds distilled water at 62°F.

(*Note:* This is a *struck* bushel. A *heaped* bushel in general equals $1\frac{1}{4}$ struck bushels, although for apples and pears it contains 1.2731 struck bushels = 2737.72 cubic inches.)

One U. S. gallon (dry measure) = $\frac{1}{8}$ bushel and contains 268.8 cubic inches.

(*Note:* This is not a legal U.S. *dry measure* and therefore is given for comparison only.)

One British Imperial bushel contains 1.2843 cubic feet or 2219.36 cubic inches. It holds 80 pounds distilled water at 62°F.

$$1 \text{ British Imperial gallon} = \frac{1}{8} \text{ Imperial bushel and contains } 277.42 \text{ cubic inches.}$$

$$1 \text{ Winchester bushel} = 0.9694 \text{ Imperial bushel}$$

$$1 \text{ Imperial bushel} = 1.032 \text{ Winchester bushels}$$

Same relations as before maintain for gallons (dry measure).

[*Note:* 1 U.S. gallon (dry) = 1.164 U. S. gallons (liquid).]

U.S. UNITS¹⁰

2 pints = 1 quart = 67.2 cubic inches

4 quarts = 1 gallon = 8 pints = 268.8 cubic inches

2 gallons = 1 peck = 16 pints = 8 quarts = 537.6 cubic inches

4 pecks = 1 bushel = 64 pints = 32 quarts = 8 gallons = 2150.42 cubic inches

1 cubic foot contains 6.428 gallons (dry measure)

Liquid Measure One U.S. gallon (liquid measure) contains 231 cubic inches. It holds 8.336 pounds distilled water at 62°F.

One British Imperial gallon contains 277.42 cubic inches. It holds 10 pounds distilled water at 62°F.

1 U.S. gallon (liquid) = 0.8327 Imperial gallon

1 Imperial gallon = 1.201 U.S. gallons (liquid)

[*Note:* 1 U.S. gallon (liquid) = 0.8594 U.S. gallon (dry).]

U.S. UNITS

4 gills = 1 pint = 16 fluid ounces

2 pints = 1 quart = 8 gills = 32 fluid ounces

4 quarts = 1 gallon = 32 gills = 8 pints = 128 fluid ounces

1 cubic foot contains 7.4805 gallons (liquid measure)

Apothecaries' Fluid Measure

60 minims = 1 fluid drachm

8 drachms = 1 fluid ounce

In the United States a fluid ounce is the 128th part of a U.S. gallon, or 1.805 cubic inches or 29.58 cubic centimeters. It contains 455.8 grains of water at 62°F. In Great Britain the fluid ounce is 1.732 cubic inches and contains 1 ounce avoirdupois (or 437.5 grains) of water at 62°F.

¹⁰The *gallon* is not a U.S. legal *dry measure*.

TABLE 51.31 Plane Angle (No Dimensions)

<div>Multiply Number of → to by → Obtain ↓</div>	Degrees	Minutes	Quadrants	Radians ^a	Revolutions ^a (Circumferences)	Seconds
Degrees	1	1.667×10^{-2}	90	57.30	360	2.778×10^{-4}
Minutes	60	1	5400	3438	2.16×10^4	1.667×10^{-2}
Quadrants	1.111×10^{-2}	1.852×10^{-4}	1	0.6366	4	3.087×10^{-6}
Radians^a	1.745×10^{-2}	2.909×10^{-4}	1.571	1	6.283	4.848×10^{-6}
Revolutions ^a (Circumferences)	2.778×10^{-3}	4.630×10^{-5}	0.25	0.1591	1	7.716×10^{-7}
Seconds	3600	60	3.24×10^5	2.063×10^5	1.296×10^6	1

^a 2π rad = 1 circumference = 360° by definition.

TABLE 51.32 Solid Angle (No Dimensions)

<div>Multiply Number of → to Obtain ↓ by →</div>	Hemispheres	Spheres ^a	Spherical Right Angles	Steradians ^b
Hemispheres	1	2	0.25	0.1592
Spheres ^a	0.5	1	0.125	7.958×10^{-2}
Spherical Right Angles	4	8	1	0.6366
Steradians^b	6.283	12.57	1.571	1

^aA sphere is the total solid angle about a point.

^b 4π steradians = 1 sphere by definition.

TABLE 51.33 Time [T]

<div>Multiply Number of → to by → Obtain ↓</div>	Days	Hours	Minutes	Months (Average) ^a	Seconds	Weeks
Days	1	4.167×10^{-2}	6.944×10^{-4}	30.42	1.157×10^{-5}	7
Hours	24	1	1.667×10^{-2}	730.0	2.778×10^{-4}	168
Minutes	1440	60	1	4.380×10^{-4}	1.667×10^{-2}	1.008×10^4
Months (Average) ^a	3.288×10^{-2}	1.370×10^{-3}	2.283×10^{-5}	1	3.806×10^{-7}	0.2302
Seconds	8.64×10^4	3600	60	2.628×10^6	1	6.048×10^5
Weeks	0.1429	5.952×10^{-3}	9.921×10^{-5}	4.344	1.654×10^{-6}	1

^aOne common year = 365 days; one leap year = 366 days; one average month = $\frac{1}{12}$ of a common year.

TABLE 51.34 Linear Velocity [LT^{-1}]

<div>Multiply Number of → by → to Obtain ↓</div>										
	Centimeters per Second	Feet per Minute	Feet per Second	Kilometers per Hour	Kilometers per Minute	Knots ^a	Meters per Minute	Meters per Second	Miles per Hour	Miles per Minute
Centimeters per Second	1	0.5080	30.48	27.78	1667	51.48	1.667	100	44.70	2682
Feet per Minute	1.969	1	60	54.68	3281	101.3	3.281	196.8	88	5280
Feet per Second	3.281×10^{-2}	1.667×10^{-2}	1	0.9113	54.68	1.689	5.468×10^{-2}	3.281	1.467	88
Kilometers per Hour	0.036	1.829×10^{-2}	1.097	1	60	1.853	0.06	3.6	1.609	96.54
Kilometers per Minute	0.0006	3.048×10^{-4}	1.829×10^{-2}	1.667×10^{-2}	1	3.088×10^{-2}	0.001	0.06	2.682×10^{-2}	1.609
Knots ^a	1.943×10^{-2}	9.868×10^{-3}	0.5921	0.5396	32.38	1	3.238×10^{-2}	1.943	0.8684	52.10
Meters per Minute	0.6	0.3048	18.29	16.67	1000	30.88	1	60	26.82	1609
Meters per Second	0.01	5.080×10^{-3}	0.3048	0.2778	16.67	0.5148	1.667×10^{-2}	1	0.4470	26.82
Miles per Hour	2.237×10^{-2}	1.136×10^{-2}	0.6818	0.6214	37.28	1.152	3.728×10^{-2}	2.237	1	60
Miles per Minute	3.728×10^{-4}	1.892×10^{-4}	1.136×10^{-2}	1.036×10^{-2}	0.6214	1.919×10^{-2}	6.214×10^{-4}	3.728×10^{-2}	1.667×10^{-2}	1

^aNautical miles per hour.

Linear Velocity

The Miner's Inch. The miner's inch is used in measuring flow of water. An act of the California legislature, May 23, 1901, makes the standard miner's inch $1.5 \text{ ft}^3/\text{min}$, measured through any aperture or orifice.

The term miner's inch is more or less indefinite, for the reason that California water companies do not all use the same head above the center of the aperture, and the inch varies from 1.36 to $1.73 \text{ ft}^3/\text{min}$, but the most common measurement is through an aperture 2 in. high and whatever length is required and through a plank $1\frac{1}{4}$ in. thick. The lower edge of the aperture should be 2 in. above the bottom of the measuring box and the plank 5 in. high above the aperture, thus making a 6-in. head above the center of the stream. Each square inch of this opening represents a miner's inch, which is equal to a flow of $1.5 \text{ ft}^3/\text{min}$.

Avoirdupois Weight. Used Commercially.

$$\begin{aligned} 27.343 \text{ grains} &= 1 \text{ drachm} \\ 16 \text{ drachms} &= 1 \text{ ounce(oz)} = 437.5 \text{ grains} \\ 16 \text{ ounces} &= 1 \text{ pound(lb)} = 7000 \text{ grains} \\ 28 \text{ pounds} &= 1 \text{ quarter(qr)} \\ 4 \text{ quarters} &= 1 \text{ hundredweight(cwt)} \\ &= 112 \text{ pounds} \end{aligned}$$

$$20 \text{ hundredweight} = 1 \text{ gross or long ton}^{11}$$

$$\begin{aligned} 200 \text{ pounds} &= 1 \text{ net or short ton} \\ 14 \text{ pounds} &= 1 \text{ stone} \\ 100 \text{ pounds} &= 1 \text{ quintal} \end{aligned}$$

Troy Weight. Used in weighing gold or silver.

$$\begin{aligned} 24 \text{ grains} &= 1 \text{ pennyweight(dwt)} \\ 20 \text{ pennyweights} &= 1 \text{ ounce(oz)} = 480 \text{ grains} \\ 12 \text{ ounces} &= 1 \text{ pound(lb)} = 5760 \text{ grains} \end{aligned}$$

The grain is the same in avoirdupois, troy, and apothecaries' weights. A carat, for weighing diamonds, $= 3.086 \text{ grains} = 0.200 \text{ gram}$ (International Standard, 1913.)

$$\begin{aligned} 1 \text{ pound troy} &= 0.8229 \text{ pound avoirdupois} \\ 1 \text{ pound avoirdupois} &= 1.2153 \text{ pounds troy} \end{aligned}$$

Apothecaries' Weight. Used in compounding medicines.

$$\begin{aligned} 20 \text{ grains} &= 1 \text{ scruple()} \\ 3 \text{ scruples} &= 1 \text{ drachm()} = 60 \text{ grains} \\ 8 \text{ drachms} &= 1 \text{ ounce()} = 480 \text{ grains} \\ 12 \text{ ounces} &= 1 \text{ pound(lb)} = 5760 \text{ grains} \end{aligned}$$

The grain is the same in avoirdupois, troy, and apothecaries' weights.

$$\begin{aligned} 1 \text{ pound apothecaries} &= 0.82286 \text{ pound avoirdupois} \\ 1 \text{ pound avoirdupois} &= 1.2153 \text{ pounds apothecaries} \end{aligned}$$

¹¹ The long ton is used by the U.S. custom houses in collecting duties upon foreign goods. It is also used in freighting coal and selling it wholesale.

TABLE 51.35 Angular Velocity [T^{-1}]

<div><div>Multiply Number of →</div><div>to Obtain ↓</div></div>	Degrees per Second	Radians per Second	Revolutions per Minute	Revolutions per Second
Degrees per Second	1	57.30	6	360
Radians per Second	1.745×10^{-2}	1	0.1047	6.283
Revolutions per Minute	0.1667	9.549	1	60
Revolutions per Second	2.778×10^{-3}	0.1592	1.667×10^{-2}	1

TABLE 51.36 Linear Acceleration^a [LT^{-2}]

<div><div>Multiply Number of →</div><div>to Obtain ↓</div></div>	Centimeters per Second per Second	Feet per Second per Second	Kilometers per Hour per Second	Meters per Second per Second	Miles per Hour per Second
Centimeters per Second per Second	1	30.48	27.78	100	44.70
Feet per Second per Second	3.281×10^{-2}	1	0.9113	3.281	1.467
Kilometers per Hour per Second	0.036	1.097	1	3.6	1.609
Meters per Second per Second	0.01	0.3048	0.2778	1	0.4470
Miles per Hour per Second	2.237×10^{-2}	0.6818	0.6214	2.237	1

^aThe (standard) acceleration due to gravity (g_0) = 980.7 cm/sec sec, = 32.17 ft/sec sec = 35.30 km/hr sec = 9.807 m/sec sec = 21.94 mph/sec.

TABLE 51.37 Angular Acceleration [T^{-2}]

<div><div>Multiply Number of →</div><div>to Obtain ↓</div></div>	Radians per Second per Second	Revolutions per Minute per Minute	Revolutions per Minute per Second	Revolutions per Second per Second
Radians per Second per Second	1	1.745×10^{-3}	0.1047	6.283
Revolutions per Minute per Minute	573.0	1	60	3600
Revolutions per Minute per Second	9.549	1.667×10^{-2}	1	60
Revolutions per Second per Second	0.1592	2.778×10^{-4}	1.667×10^{-2}	1

TABLE 51.38 Mass [*M*] and Weight^a

Multiply Number of → to Obtain ↓											
	Grains	Grams	Kilograms	Milligrams	Ounces ^b	Pounds ^b	Tons (Long)	Tons (Metric)	Tons (Short)		
Grains	1	15.43	1.543×10^4	1.543×10^{-2}	437.5	7000					
Grams	6.481×10^{-2}	1	1000	0.001	28.35	453.6	1.016×10^6	$\times 10^6$	9.072×10^5		
Kilograms	6.481×10^{-5}	0.001	1	10^{-6}	2.835×10^{-2}	0.4536	1016	1000	907.2		
Milligrams	64.81	1000	10^6	1	2.835×10^4	4.536×10^5	1.016×10^9	10^9	9.072×10^8		
Ounces ^b	2.286×10^{-3}	3.527×10^{-2}	35.27	3.527×10^{-5}	1	16	3.584×10^4	3.527×10^4	3.2×10^4		
Pounds ^b	1.429×10^{-4}	2.205×10^{-3}	2.205	2.205×10^{-6}	6.250×10^{-2}	1	2240	2205	2000		
Tons (Long)	—	9.842×10^{-7}	9.842×10^{-4}	9.842×10^{-10}	2.790×10^{-5}	4.464×10^{-4}	1	0.9842	0.8929		
Tons (Metric)	—	10^{-6}	0.001	10^{-9}	2.835×10^{-5}	4.536×10^{-4}	1.016	1	0.9072		
Tons (Short)	—	1.102×10^{-6}	1.102×10^{-3}	1.102×10^{-9}	3.125×10^{-5}	0.0005	1.120	1.102	1		

^aThese same conversion factors apply to the *gravitational* units of force having the corresponding names. The dimensions of these units when used as gravitational units of force are MLT^{-2} ; see Table 51.40.

^bAvoirdupois pounds and ounces.

TABLE 51.39 Density or Mass per Unit Volume [ML^{-3}]

<div>to Obtain ↓</div> <div>Multiply Number of →</div> <div>$by \rightarrow$</div>	Grams per Cubic Centimeter	Kilograms per Cubic Meter	Pounds per Cubic Foot	Pounds per Cubic Inch
Grams per Cubic Centimeter	1	0.001	1.602×10^{-2}	27.68
Kilograms per Cubic Meter	1000	1	16.02	2.768×10^4
Pounds per Cubic Foot	62.43	6.243×10^{-2}	1	1728
Pounds per Cubic Inch	3.613×10^{-2}	3.613×10^{-5}	5.787×10^{-4}	1
Pounds per Mil Foot ^a	3.405×10^{-7}	3.405×10^{-10}	5.456×10^{-9}	9.425×10^{-6}

^aUnit of volume is a volume one foot long and one circular mil in cross-sectional area.

TABLE 51.40 Force^a [MLT^{-2}] or [F]

Multiply Number of → by → to Obtain ↓								
	Dynes	Grams	Joules per Centimeter	Newtons, or Joules per Meter	Kilograms	Pounds	Poundals	
Dynes	1	980.7	10 ⁷	10 ⁵	9.807 × 10 ⁵	4.448 × 10 ⁵	1.383 × 10 ⁴	
Grams	1.020 × 10 ⁻³	1	1.020 × 10 ⁴	102.0	1000	453.6	14.10	
Joules per Centimeter	10 ⁻⁷	9.807 × 10 ⁻⁵	1	0.01	9.807 × 10 ⁻²	4.448 × 10 ⁻²	1.383 × 10 ⁻³	
Newtons, or Joules per Meter	10 ⁻⁵	9.807 × 10 ⁻³	100	1	9.807	4.448	0.1383	
Kilograms	1.020 × 10 ⁻⁶	0.001	10.20	0.1020	1	0.4536	1.410 × 10 ⁻²	
Pounds	2.248 × 10 ⁻⁶	2.205 × 10 ⁻³	22.48	0.2248	2.205	1	3.108 × 10 ⁻²	
Poundals	7.233 × 10 ⁻⁵	7.093 × 10 ⁻²	723.3	7.233	70.93	32.17	1	

^aConversion factors between absolute and gravitational units apply only under standard acceleration due to gravity conditions. (See Section 51.4.)

TABLE 51.41 Pressure or Force per Unit Area [$ML^{-1}T^{-2}$] or [FL^{-2}]

<div> <div>Multiply Number of \rightarrow</div> <div>by \rightarrow</div> <div>to Obtain \downarrow</div> </div>										
	Atmospheres ^a	Baryes or Dynes per Square Centimeter	Centimeters of Mercury at 0°C ^b	Inches of Mercury at 0°C ^b	Inches of Water at 4°C	Kilograms per Square Meter ^c	Pounds per Square Foot	Pounds per Square Inch	Tons (Short) per Square Foot	Pascal
Atmospheres ^a	1	9.869×10^{-7}	1.316×10^{-2}	3.342×10^{-2}	2.458×10^{-3}	9.678×10^{-5}	4.725×10^{-4}	6.804×10^{-2}	0.9450	9.869×10^{-6}
Baryes or Dynes per Square Centimeter	1.013×10^6	1	1.333×10^4	3.386×10^4	2.491×10^{-3}	98.07	478.8	6.895×10^4	9.576×10^5	10
Centimeters of Mercury at 0°C ^b	76.00	7.501×10^{-5}	1	2.540	0.1868	7.356×10^{-3}	3.591×10^{-2}	5.171	71.83	7.501×10^{-4}
Inches of Mercury at 0°C ^b	29.92	2.953×10^{-5}	0.3937	1	7.355×10^{-2}	2.896×10^{-3}	1.414×10^{-2}	2.036	28.28	2.953×10^{-4}
Inches of Water at 4°C	406.8	4.015×10^{-4}	5.354	13.60	1	3.937×10^{-2}	0.1922	27.68	384.5	4.015×10^{-8}
Kilograms per Square Meter ^c	1.033×10^4	1.020×10^{-2}	136.0	345.3	25.40	1	4.882	703.1	9765	0.1020
Pounds per Square Foot	2117	2.089×10^{-3}	27.85	70.73	5.204	0.2048	1	144	2000	2.089×10^{-2}
Pounds per Square Inch	14.70	1.450×10^{-5}	0.1934	0.4912	3.613×10^{-2}	1.422×10^{-3}	6.944×10^{-3}	1	13.89	1.450×10^{-4}
Tons (Short) per Square Foot	1.058	1.044×10^{-6}	1.392×10^{-2}	3.536×10^{-2}	2.601×10^{-3}	1.024×10^{-4}	0.0005	0.072	1	1.044×10^{-5}
Pascal	1.013×10^5	10^{-1}	1.333×10^3	3.386×10^3	2.491×10^{-4}	9.807	47.88	6.895×10^3	9.576×10^4	1

^aDefinition: One atmosphere (standard) = 76 cm of mercury at 0°C.

^bTo convert height h of a column of mercury at t degrees Centigrade to the equivalent height h_0 at 0°C use $h_0 = h\{1 - (m - b)t/(1 + mt)\}$, where $m = 0.0001818$ and $b = 18.4 \times 10^{-6}$ if the scale is engraved on brass; $b = 8.5 \times 10^{-6}$ if on glass. This assumes the scale is correct at 0°C; for other cases (any liquid) see *International Critical Tables*, Vol. 1, p. 68

^c1 g/cm² = 10 kg/m².

TABLE 51.42 Torque or Moment of Force [ML^2T^{-2}] or $[FL]^a$

<div><div>Multiply Number of →</div><div>to Obtain ↓</div></div>	Dyne- Centimeters	Gram- Centimeters	Kilogram- Meters	Pound-Feet	Newton- Meter
Dyne-Centimeters	1	980.7	9.807×10^7	1.356×10^7	10^7
Gram-Centimeters	1.020×10^{-3}	1	10^5	1.383×10^4	1.020×10^4
Kilogram-Meters	1.020×10^{-8}	10^{-5}	1	0.1383	0.1020
Pound-Feet	7.376×10^{-8}	7.233×10^{-5}	7.233	1	0.7376
Newton-Meter	10^{-7}	9.807×10^{-4}	9.807	1.356	1

^aSame dimensions as energy; more properly torque should be expressed as newton-meters per radian to avoid this confusion.

TABLE 51.43 Moment of Inertia [ML^2]

<div><div>Multiply Number of →</div><div>to Obtain ↓</div></div>	Gram- Centimeters Squared	Kilogram- Meters Squared	Pound-Inches Squared	Pound-Feet Squared	Slug-Feet Squared
Gram-Centimeters Squared	1	10^7	2.9266×10^3	4.21434×10^5	1.3559×10^7
Kilogram-Meters Squared	10^{-7}	1	2.9266×10^{-4}	4.21434×10^{-2}	1.3559
Pound-Inches Squared	3.4169×10^{-4}	3.4169×10^3	1	144	4.63304×10^3
Pound-Feet Squared	2.37285×10^{-6}	23.7285	6.944×10^{-3}	1	32.1739
Slug-Feet Squared	7.37507×10^{-8}	0.737507	2.15841×10^{-4}	3.10811×10^{-2}	1

TABLE 51.44 Energy, Work and Heat^a [ML^2T^{-2}] or [FL]

<div> <div>Multiply Number of →</div> <div>by →</div> <div>to Obtain ↓</div> </div>	British Thermal Units ^b	Centimeter- Grams	Ergs or Centimeter- Dynes	Foot-Pounds	Horsepower- Hours	Joules, ^c or Watt-Seconds	Kilogram- Calories ^b	Kilowatt- Hours	Meter- Kilograms	Watt-Hours
	British Thermal Units ^b	Centimeter-Grams	Ergs or Centimeter-Dynes	Foot-Pounds	Horsepower-Hours	Joules, ^c or Watt-Seconds	Kilogram-Calories ^b	Kilowatt-Hours	Meter-Kilograms	Watt-Hours
British Thermal Units ^b	1	9.297×10^{-8}	9.480×10^{-11}	1.285×10^{-3}	2545	9.480×10^{-4}	3.969	3413	9.297×10^{-3}	3.413
Centimeter-Grams	1.076×10^7	1	1.020×10^{-3}	1.383×10^4	2.737×10^{10}	1.020×10^4	4.269×10^7	3.671×10^{10}	10^5	3.671×10^7
Ergs or Centimeter-Dynes	1.055×10^{10}	980.7	1	1.356×10^7	2.684×10^{12}	10^7	4.186×10^{10}	3.6×10^{13}	9.807×10^7	3.6×10^{10}
Foot-Pounds	778.0	7.233×10^{-5}	7.367×10^{-8}	1	1.98×10^6	0.7376	3087	2.655×10^6	7.233	2655
Horsepower-Hours	3.929×10^{-4}	3.654×10^{-11}	3.722×10^{-14}	5.050×10^{-7}	1	3.722×10^{-7}	1.559×10^{-3}	1.341	3.653×10^{-6}	1.341×10^{-3}
Joules, ^c or Watt-Seconds	1054.8	9.807×10^{-5}	10^{-7}	1.356	2.684×10^6	1	4186	3.6×10^6	9.807	3600
Kilogram-Calories ^b	0.2520	2.343×10^{-8}	2.389×10^{-11}	3.239×10^{-4}	641.3	2.389×10^{-4}	1	860.0	2.343×10^{-3}	0.8600
Kilowatt-Hours	2.930×10^{-4}	2.724×10^{-11}	2.778×10^{-14}	3.766×10^{-7}	0.7457	2.778×10^{-7}	1.163×10^{-3}	1	2.724×10^{-6}	0.001
Meter-Kilograms	107.6	10^{-5}	1.020×10^{-8}	0.1383	2.737×10^5	0.1020	426.9	3.671×10^5	1	367.1
Watt-Hours	0.2930	2.724×10^{-8}	2.778×10^{-11}	3.766×10^{-4}	745.7	2.778×10^{-4}	1.163	1000	2.724×10^{-3}	1

^aSee note at the bottom of Table 51.45.

^bMean calorie and Btu used throughout. One gram-calorie = 0.001 kilogram-calorie; one Ostwald calorie = 0.1 kilogram-calorie.

The IT cal, 1000 international steam table calories, has been defined as the 1/860th part of the international kilowatthour (see *Mechanical Engineering*, Nov. 1935, p. 710). Its value is very nearly equal to the mean kilogram-calorie, 1 IT cal = 1.00037 kilogram-calories (mean). 1 Btu = 251.996 IT cal.

^cAbsolute joule, defined as 10^7 ergs. The international joule, based on the international ohm and ampere, equals 1.0003 absolute joules.

TABLE 51.45 Power or Rate of Doing Work^a [ML^2T^{-3}] or [FL^{-1}]

<div>Multiply Number of → by → to Obtain ↓</div>	British Thermal Units per Minute									
	Ergs per Second	Foot-Pounds per Minute	Foot-Pounds per Second	Horsepower ^a	Kilogram-Calories per Minute	Kilowatts	Metric Horsepower	Watts		
British Thermal Units per Minute	1	5.689×10^{-9}	1.285×10^{-3}	7.712×10^{-2}	42.41	3.969	56.89	41.83×10^{-2}		
Ergs per Second	1.758×10^8	1	2.259×10^5	1.356×10^7	7.457×10^9	6.977×10^8	10^{10}	7.355×10^9		
Foot-Pounds per Minute	778.0	4.426×10^{-6}	1	60	3.3×10^4	3087	4.426×10^4	3.255×10^4		
Foot-Pounds per Second	12.97	7.376×10^{-8}	1.667×10^{-2}	1	550	51.44	737.6	542.5		
Horsepower ^a	2.357×10^{-2}	1.341×10^{-10}	3.030×10^{-5}	1.818×10^{-3}	1	9.355×10^{-2}	1.341	0.9863		
Kilogram-Calories per Minute	0.2520	1.433×10^{-9}	3.239×10^{-4}	1.943×10^{-2}	10.69	1	14.33	10.54		
Kilowatts	1.758×10^{-2}	10^{-10}	2.260×10^{-5}	1.356×10^{-3}	0.7457	6.977×10^{-2}	1	0.7355		
Metric Horsepower	2.390×10^{-2}	1.360×10^{-10}	3.072×10^{-5}	1.843×10^{-3}	1.014	9.485×10^{-2}	1.360	1		
Watts	17.58	10^{-7}	2.260×10^{-2}	1.356	745.7	69.77	1000	735.5		

Note:

1 Cheval-vapeur = 75 kilogram-meters per second

1 Poncelet = 100 kilogram-meters per second

^aThe “horsepower” used in these tables is equal to 550 foot-pounds per second by definition. Other definitions are one horsepower equals 746 watts (U.S. and Great Britain) and one horsepower equals 736 watts (continental Europe). Neither of these latter definitions is equivalent to the first; the “horsepowers” defined in these latter definitions are widely used in the rating of electrical machinery.

TABLE 51.46 Quantity of Electricity and Dielectric Flux [*Q*]

<div> <div>to Obtain</div> <div>↓</div> </div> <div> <div>Multiply</div> <div>Number</div> <div>of →</div> </div>	Abcoulombs	Ampere-Hours	Coulombs	Faradays	Stat coulombs
Abcoulombs	1	360	0.1	9649	3.335×10^{-11}
Ampere-Hours	2.778×10^{-3}	1	2.778×10^{-4}	26.80	9.259×10^{-14}
Coulombs	10	3600	1	9.649×10^4	3.335×10^{-10}
Faradays	1.036×10^{-4}	3.731×10^{-2}	1.036×10^{-5}	1	3.457×10^{-15}
Statcoulombs	2.998×10^{10}	1.080×10^{13}	2.998×10^9	2.893×10^{14}	1

TABLE 51.47 Charge per Unit Area and Electric Flux Density [*QL*⁻²]

<div> <div>to Obtain</div> <div>↓</div> </div> <div> <div>Multiply</div> <div>Number</div> <div>of →</div> </div>	Abcoulombs per Square Centimeter	Coulombs per Square Centimeter	Coulombs per Square Inch	Statcoulombs per Square Centimeter	Coulombs per Square Meter
Abcoulombs per Square Centimeter	1	0.1	1.550×10^{-2}	3.335×10^{-11}	10^{-5}
Coulombs per Square Centimeter	10	1	0.1550	3.335×10^{-10}	10^{-4}
Coulombs per Square Inch	64.52	6.452	1	2.151×10^{-9}	6.452×10^{-4}
Statcoulombs per Square Centimeter	2.998×10^{10}	2.998×10^9	4.647×10^8	1	2.998×10^5
Coulombs per Square Meter	10^5	10^4	1550	3.335×10^{-6}	1

TABLE 51.48 Electric Current [*QT*⁻¹]

<div> <div>to Obtain</div> <div>↓</div> </div> <div> <div>Multiply</div> <div>Number</div> <div>of →</div> </div>	Abamperes	Amperes	Statamperes
Abamperes	1	0.1	3.335×10^{-11}
Amperes	10	1	3.335×10^{-10}
Statamperes	2.998×10^{10}	2.998×10^9	1

TABLE 51.49 Current Density [$QT^{-1}L^{-2}$]

<div><div>to Obtain ↓</div><div>Multiply Number of →</div><div>\swarrow \searrow \downarrow</div></div>	Abamperes per Square Centimeter	Amperes per Square Centimeter	Amperes per Square Inch	Statamperes per Square Centimeter	Amperes per Square Meter
Abamperes per Square Centimeter	1	0.1	1.550×10^{-2}	3.335×10^{-11}	10^{-5}
Amperes per Square Centimeter	10	1	0.1550	3.335×10^{-10}	10^{-4}
Amperes per Square Inch	64.52	6.452	1	2.151×10^{-9}	6.452×10^{-4}
Statamperes per Square Centimeter	2.998×10^{10}	2.998×10^9	4.647×10^8	1	2.998×10^5
Amperes per Square Meter	10^5	10^4	1550	3.335×10^{-6}	1

TABLE 51.50 Electric Potential and Electromotive Force [$MQ^{-1}L^2T^{-2}$] or [$FQ^{-1}L$]

<div><div>to Obtain ↓</div><div>Multiply Number of →</div><div>\swarrow \searrow \downarrow</div></div>	Abvolts	Microvolts	Millivolts	Statvolts	Volts
Abvolts	1	100	10^5	2.998×10^{10}	10^8
Microvolts	0.01	1	1000	2.998×10^8	10^6
Millivolts	10^{-5}	0.001	1	2.998×10^5	1000
Statvolts	3.335×10^{-11}	3.335×10^{-9}	3.335×10^{-6}	1	3.335×10^{-3}
Volts	10^{-8}	10^{-6}	0.001	299.8	1

TABLE 51.51 Electric Field Intensity and Potential Gradient [$MQ^{-1}LT^{-2}$] or [FQ^{-1}]

to Obtain ↓	Multiply Number of → by →	Abvolts per Centimeter	Microvolts per Meter	Millivolts per Meter	Statvolts per Centimeter	Volts per Centimeter	Kilovolts per Centimeter	Volts per Inch	Volts per Mil	Volts per Meter
Abvolts per Centimeter		1	1	1000	2.998×10^{10}	10^8	10^{11}	3.937×10^7	3.937×10^{10}	10^6
Microvolts per Meter		1	1	1000	2.998×10^{10}	10^8	10^{11}	3.937×10^7	3.937×10^{10}	10^6
Millivolts per Meter		0.001	0.001	1	2.998×10^7	10^5	10^8	3.937×10^4	3.937×10^7	1000
Statvolts per Centimeter		3.335×10^{-11}	3.335×10^{-11}	3.335×10^{-8}	1	3.335×10^{-3}	3.335	1.313×10^{-3}	1.313	3.335×10^{-5}
Volts per Centimeter		10^{-8}	10^{-8}	10^{-5}	299.8	1	1000	0.3937	393.7	10^{-2}
Kilovolts per Centimeter		10^{-11}	10^{-11}	10^{-8}	0.2998	0.001	1	3.937×10^{-4}	0.3937	10^{-5}
Volts per Inch		2.540×10^{-8}	2.540×10^{-8}	2.540×10^{-5}	761.6	2.540	2.540	1	1000	2.540×10^{-2}
Volts per Mil		2.540×10^{-11}	2.540×10^{-11}	2.540×10^{-8}	0.7616	2.540×10^{-3}	2.540	0.001	1	2.540×10^{-5}
Volts per Meter		10^{-6}	10^{-6}	10^{-3}	2.998×10^4	100	10^5	39.37	3.937×10^4	1

TABLE 51.52 Electric Resistance $[MQ^{-2}L^2T^{-1}]$ or $[FQ^{-2}LT]$

<div> <div>to Obtain ↓</div> <div>Multiply Number of → by ↘</div> </div>	Abohms	Megohms	Microhms	Ohms	Statohms
Abohms	1	10^{15}	1000	10^9	8.988×10^{20}
Megohms	10^{-15}	1	10^{-12}	10^{-6}	8.988×10^5
Microhms	0.001	10^{12}	1	10^6	8.988×10^{17}
Ohms	10^{-9}	10^6	10^{-6}	1	8.988×10^{11}
Statohms	1.112×10^{-21}	1.112×10^{-6}	1.112×10^{-18}	1.112×10^{-12}	1

Note: Electric Conductance $[F^{-1}Q^2L^{-1}T^{-1}]$. 1 Siemens = 1 mho = 1 ohm⁻¹ = 10^{-6} megmho = 10^6 micromho.

TABLE 51.53 Electric Resistivity^a $[MQ^{-2}L^3T^{-1}]$ or $[FQ^{-2}L^2T]$

<div> <div>to Obtain ↓</div> <div>Multiply Number of → by ↘</div> </div>	Abohm- Centimeters	Microhm- Centimeters	Microhm- Inches	Ohms (Mil, Foot)	Ohms (Meter, Gram) ^b	Ohm- Meters
Abohm-Centimeters	1	1000	2540	166.2	$10^5/\delta$	10^{11}
Microhm-Centimeters	0.001	1	2.540	0.1662	100/ δ	10^8
Microhm-Inches	3.937×10^{-4}	0.3937	1	6.545×10^{-2}	39.37/ δ	3.937×10^7
Ohms (Mil, Foot)	6.015×10^{-3}	6.015	15.28	1	601.5/ δ	6.015×10^8
Ohms (Meter, Gram) ^b	$10^{-5}\delta$	0.01 δ	$2.540 \times 10^{-2}\delta$	$1.662 \times 10^{-3}\delta$	1	$10^{-6}\delta$
Ohm-Meters	10^{-11}	10^{-8}	2.540×10^{-8}	1.662×10^{-9}	$10^{-6}/\delta$	1

^aIn this table δ is density in grams per cubic-centimeters. The following names, corresponding respectively to those at the tops of columns, are sometimes used: abohms per centimeter cube; microhms per centimeter cube; microhms per inch cube; ohms per milfoot; ohms per meter-gram. The first four columns are headed by units of *volume* resistivity, the last by a unit of *mass* resistivity. The dimensions of the latter are $Q^{-2}L^6T^{-1}$, not those given in the heading of the table.

^bOne ohm (meter, gram) = 5710 ohms (mile, pound).

TABLE 51.54 Electric Conductivity^a $[M^{-1}Q^2L^{-3}T]$ or $[F^{-1}Q^2L^{-2}T^{-1}]$

<div> <div>to Obtain ↓</div> <div>Multiply Number of → by ↘</div> </div>	Abmhos per Centimeter	Mhos (Mil, Foot)	Mhos (Meter, Gram)	Micromhos per Centimeter	Micromhos per Inch	Siemens per Meter
Abmhos per Centimeter	1	6.015×10^{-3}	$10^{-5}\delta$	0.001	3.937×10^{-4}	10^{-11}
Mhos (Mil, Foot)	166.2	1	$1.662 \times 10^{-3}\delta$	0.1662	6.524×10^{-2}	1.662×10^{-9}
Mhos (Meter, Gram)	$10^5/\delta$	601.5/ δ	1	100/ δ	39.37/ δ	$10^{-6}/\delta$
Micromhos per Centimeter	1000	6.015	0.01 δ	1	0.3937	10^{-8}
Micromhos per Inch	2540	15.28	$2.540 \times 10^{-2}\delta$	2.540	1	2.54×10^{-8}
Siemens per Meter	10^{11}	6.015×10^8	$10^6\delta$	10^8	3.937×10^7	1

^aSee footnote of Table 51.53. Names sometimes used are abmho per centimeter cube, mho per mil-foot, etc. Dimensions of mass conductivity are $Q^2L^{-6}T$.

TABLE 51.55 Capacitance [$M^{-1}Q^2L^{-2}T^2$] or [$F^{-1}Q^2L^{-1}$]

<div><div>to Obtain</div><div>↓</div><div><div>Multiply Number of →</div><div>\swarrow $by \rightarrow$</div></div></div>	Abfarads	Farads	Microfarads	Statfarads
Abfarads	1	10^{-9}	10^{-15}	1.112×10^{-21}
Farads	10^9	1	10^{-6}	1.112×10^{-12}
Microfarads	10^{15}	10^6	1	1.112×10^{-6}
Statfarads	8.988×10^{20}	8.988×10^{11}	8.988×10^5	1

TABLE 51.56 Inductance [$MQ^{-2}L^2$] or [$FQ^{-2}LT^2$]

<div><div>to Obtain</div><div>↓</div><div><div>Multiply Number of →</div><div>\swarrow $by \rightarrow$</div></div></div>	Abhenries ^a	Henries	Microhenries	Millihenries	Stathenries
Abhenries ^a	1	10^9	1000	10^6	8.988×10^{20}
Henries	10^{-9}	1	10^{-6}	0.001	8.988×10^{11}
Microhenries	0.001	10^6	1	1000	8.988×10^{17}
Millihenries	10^{-6}	1000	0.001	1	8.988×10^{14}
Stathenries	1.112×10^{-21}	1.112×10^{-12}	1.112×10^{-18}	1.112×10^{-15}	1
					1

^aAn abhenry is sometimes called a “centimeter.”

TABLE 51.57 Magnetic Flux [$MQ^{-1}L^2T^{-1}$] or [$FQ^{-1}LT$]

<div><div>to Obtain</div><div>↓</div><div><div>Multiply Number of →</div><div>\swarrow $by \rightarrow$</div></div></div>	Kilolines	Maxwells (or Lines)	Webers
Kilolines	1	0.001	10^5
Maxwells (or Lines)	1000	1	10^8
Webers	10^{-5}	10^{-8}	1

TABLE 51.58 Magnetic Flux Density [$MQ^{-1}T^{-1}$] or [$FQ^{-1}L^{-1}T$]

<div><div>to Obtain</div><div>↓</div><div><div>Multiply Number of →</div><div>\swarrow $by \rightarrow$</div></div></div>	Gausses (or Lines per Square Centimeter)	Lines per Square Inch	Webers per Square Centimeter	Webers per Square Inch	Tesla (Webers per Square Meter)
Gausses (or Lines per Square Centimeter)	1	0.1550	10^8	1.550×10^7	10^4
Lines per Square Inch	6.452	1	6.452×10^8	10^8	6.452×10^4
Webers per Square Centimeter	10^{-8}	1.550×10^{-9}	1	0.1550	10^{-4}
Webers per Square Inch	6.452×10^{-8}	10^{-8}	6.452	1	6.452×10^{-4}
Tesla (Webers per Square Meter)	10^{-4}	1.550×10^{-5}	10^4	1550	1

TABLE 51.59 Magnetic Potential and Magnetomotive Force [QT^{-1}]

<div>to Obtain ↓</div> <div><div>Multiply Number of →</div><div>$b_y \rightarrow$</div></div>	Abampere-Turns	Ampere-Turns	Gilberts
	Abampere-Turns		
Abampere-Turns	1	0.1	7.958×10^{-2}
Ampere-Turns	10	1	0.7958
Gilberts	12.57	1.257	1

TABLE 51.60 Magnetic Field Intensity, Potential Gradient, and Magnetizing Force [$QL^{-1}T^{-1}$]

<div>to Obtain ↓</div> <div><div>Multiply Number of →</div><div>$b_y \rightarrow$</div></div>	Abampere-Turns per Centimeter	Ampere-Turns per Centimeter	Ampere-Turns per Inch	Oersteds (Gilberts per Centimeter)	Ampere-Turns per Meter
	Abampere-Turns per Centimeter				
Abampere-Turns per Centimeter	1	0.1	3.937×10^{-2}	7.958×10^{-2}	10^{-3}
Ampere-Turns per Centimeter	10	1	0.3937	0.7958	10^{-2}
Ampere-Turns per Inch	25.40	2.540	1	2.021	2.54×10^{-2}
Oersteds (Gilberts per Centimeter)	12.57	1.257	0.4950	1	1.257×10^{-2}
Ampere-Turns per Meter	10^3	10^2	39.37	79.58	1

TABLE 51.61 Specific Heat [$L^2T^{-2}t^{-1}$] (t = temperature)

To change specific heat in gram-calories per gram per degree Centigrade to the units given in any line of the following table, multiply by the factor in the last column.

Unit of Heat or Energy	Unit of Mass	Temperature Scale ^a	Factor
Gram-calories	Gram	Centigrade	1
Kilogram-calories	Kilogram	Centigrade	1
British thermal units	Pound	Centigrade	1.800
British thermal units	Pound	Fahrenheit	1.000
Joules	Gram	Centigrade	4.186
Joules	Pound	Fahrenheit	1055
Joules	Kilogram	Kelvin	4.187×10^3
Kilowatt-hours	Kilogram	Centigrade	1.163×10^{-3}
Kilowatt-hours	Pound	Fahrenheit	2.930×10^{-4}

^aTemperature conversion formulas:

t_c = temperature in Centigrade degrees
 t_f = temperature in Fahrenheit degrees
 t_K = temperature in Kelvin degrees
 $1\text{ }^\circ\text{F} = \frac{5}{9}\text{ }^\circ\text{C}$
 $1\text{ K} = 1\text{ }^\circ\text{C}$
 $t_c = \frac{5}{9}(t_f - 32)$
 $t_f = \frac{9}{5}t_c + 32$
 $t_K = t_c + 273$

TABLE 51.62 Thermal Conductivity^a [$LM T^{-3} t^{-1}$]

<div> <div>Multiply Number of →</div> <div>by →</div> <div>to Obtain ↓</div> </div>		Btu · ft/h · ft ² · °F	Btu · in./ h · ft ² · °F	Btu · in./ sec · ft ² · °F	J/m · s · °C	kcal/m · h · °C	erg/cm · s · °C	kcal/m · s · °C	cal/cm · s · °C	W/ft · °C	W/m · K
Btu · ft/h · ft ² · °F	1	8.333 × 10 ⁻³	2.778 × 10 ⁻²	3.0 × 10 ²	5.778 × 10 ⁻¹	6.720 × 10 ⁻¹	5.778 × 10 ⁻⁶	2.419 × 10 ³	2.419 × 10 ²	1.895	5.778 × 10 ⁻¹
Btu · in./h · ft ² · °F	12	1	1	3.6 × 10 ³	6.933	8.064	6.933 × 10 ⁻⁵	2.903 × 10 ⁴	2.903 × 10 ³	2.275 × 10 ¹	6.933
Btu · in./s · ft ² · °F	3.333 × 10 ⁻³	2.778 × 10 ⁻⁴	1	1	1.926 × 10 ⁻³	2.240 × 10 ⁻³	1.926 × 10 ⁻⁸	8.064	8.064 × 10 ⁻¹	6.319 × 10 ⁻³	1.926 × 10 ⁻³
J/m · s · °C	1.731	1.442 × 10 ⁻¹	1.442 × 10 ⁻¹	5.192 × 10 ²	1	1.163	1.000 × 10 ⁻⁵	4.187 × 10 ³	4.187 × 10 ²	3.281	1.0
kcal/m · h · °C	1.483	1.240 × 10 ⁻¹	1.240 × 10 ⁻¹	4.465 × 10 ²	8.599 × 10 ⁻¹	1	8.599 × 10 ⁻⁶	3.6 × 10 ³	3.6 × 10 ²	2.821	8.599 × 10 ⁻¹
erg/cm · s · °C	1.731 × 10 ⁵	1.442 × 10 ⁻⁴	1.442 × 10 ⁻⁴	5.192 × 10 ⁷	1.0 × 10 ⁵	1.163 × 10 ⁵	1	4.187 × 10 ⁸	4.187 × 10 ⁷	3.281 × 10 ⁵	1.0 × 10 ⁵
kcal/m · s · °C	4.134 × 10 ⁻⁴	3.445 × 10 ⁻⁵	3.445 × 10 ⁻⁵	1.240 × 10 ⁻¹	2.388 × 10 ⁻⁴	2.778 × 10 ⁻⁴	2.388 × 10 ⁻⁹	1	1.0 × 10 ⁻¹	7.835 × 10 ⁻⁴	2.388 × 10 ⁻⁴
cal/cm · s · °C	4.134 × 10 ⁻³	3.445 × 10 ⁻⁴	3.445 × 10 ⁻⁴	1.240	2.388 × 10 ⁻³	2.778 × 10 ⁻³	2.388 × 10 ⁻⁸	10	1	7.835 × 10 ⁻³	2.388 × 10 ⁻³
W/ft · °C	5.276 × 10 ⁻¹	4.395 × 10 ⁻²	4.395 × 10 ⁻²	1.582 × 10 ²	3.048 × 10 ⁻¹	3.545 × 10 ⁻¹	3.048 × 10 ⁻⁶	1.276 × 10 ³	1.276 × 10 ²	1	3.048 × 10 ⁻¹
W/m · K	1.731	1.442 × 10 ⁻¹	1.442 × 10 ⁻¹	5.192 × 10 ²	1.0	1.163	1.00 × 10 ⁻⁵	4.187 × 10 ³	4.187 × 10 ²	3.281	1

^aInternational Table Btu = 1.055056 × 10³ joules and International Table cal = 4.1868 J are used throughout.

TABLE 51.63 Photometric Units

	Common Unit	Multiply by	to Get SI Unit
Luminous intensity	International candle	9.81×10^{-1}	cd
Luminance	cd/in. ²	1.550×10^3	cd/m²
	cd/cm ²	1×10^4	cd/m²
Luminous flux	Foot · lambert	3.4263	cd/m²
	cd · sr	1.0000	lm
	Candle power (spher.)	12.566	lm
Quantity of light flux			lm·
Luminous exitance ^a			lm/m²
Illuminance ^b	lm	3.103×10^3	cd/m²
	Foot candles	1.0764×10	lm/m²
	lmft ²	1.0764×10	lm/m²
	lx	1.000	lm/m²
	Phots	1×10^4	lm/m²
Luminous efficacy			lm/W

^aLuminous emittance.

^bLuminous flux density.

TABLE 51.64 Specific Gravity Conversions

Specific Gravity 60°/60°			lb/gal 60°F, wt in Air	lb/ft ³ at 60°F, wt in Air	Specific Gravity 60°/60°			lb/gal 60°F, wt in Air	lb/ft ³ at 60°F, wt in Air
	°Be	°API				°Be	°API		
0.600	103.33	104.33	4.9929	37.350	0.730	61.78	62.34	6.0769	45.458
0.605	101.40	102.38	5.0346	37.662	0.735	60.48	61.02	6.1186	45.770
0.610	99.51	100.47	5.0763	37.973	0.740	59.19	59.72	6.1603	46.082
0.615	97.64	98.58	5.1180	38.285	0.745	57.92	58.43	6.2020	46.394
0.620	95.81	96.73	5.1597	38.597	0.750	56.67	57.17	6.2437	46.706
0.625	94.00	94.90	5.2014	39.910	0.755	55.43	55.92	6.2854	47.018
0.630	92.22	93.10	5.2431	39.222	0.760	54.21	54.68	6.3271	47.330
0.635	90.47	91.33	5.2848	39.534	0.765	53.01	53.47	6.3688	47.642
0.640	88.75	89.59	5.3265	39.845	0.770	51.82	52.27	6.4104	47.953
0.645	87.05	87.88	5.3682	40.157	0.775	50.65	51.08	6.4521	48.265
0.650	85.38	86.19	5.4098	40.468	0.780	49.49	49.91	6.4938	48.577
0.655	83.74	84.53	5.4515	40.780	0.785	48.34	48.75	6.5355	48.889
0.660	82.12	82.89	5.4932	41.092	0.790	47.22	47.61	6.5772	49.201
0.665	80.53	81.28	5.5349	41.404	0.795	46.10	46.49	6.6189	49.513
0.670	78.96	79.69	5.5766	41.716	0.800	45.00	45.38	6.6606	49.825
0.675	77.41	78.13	5.6183	42.028	0.805	43.91	44.28	6.7023	50.137
0.680	75.88	76.59	5.6600	42.340	0.810	42.84	43.19	6.7440	50.448
0.685	74.38	75.07	5.7017	42.652	0.815	41.78	42.12	6.7857	50.760
0.690	72.90	73.57	5.7434	42.963	0.820	40.73	41.06	6.8274	51.072
0.695	71.44	72.10	5.7851	43.275	0.825	39.70	40.02	6.8691	51.384
0.700	70.00	70.64	5.8268	43.587	0.830	38.67	38.98	6.9108	51.696
0.705	68.58	69.21	5.8685	43.899	0.835	37.66	37.96	6.9525	52.008
0.710	67.18	67.80	5.9101	44.211	0.840	36.67	36.95	6.9941	52.320
0.715	65.80	66.40	5.9518	44.523	0.845	35.68	35.96	7.0358	52.632
0.720	64.44	65.03	5.9935	44.834	0.850	34.71	34.97	7.0775	52.943
0.725	63.10	63.67	6.0352	45.146	0.855	33.74	34.00	7.1192	53.225

TABLE 51.64 (Continued)

Specific Gravity 60°/60°	°Be	°API	lb/gal 60°F, wt in Air	lb/ft ³ at 60°F, wt in Air	Specific Gravity 60°/60°	°Be	°TW	lb/gal 60°F, wt in Air	lb/ft ³ at 60°F, wt in Air
0.860	32.79	33.03	7.1609	53.567	1.075	10.12	15	8.9537	66.978
0.865	31.85	32.08	7.2026	53.879	1.080	10.74	16	8.9954	67.290
0.870	30.92	31.14	7.2443	54.191	1.085	11.36	17	9.0371	67.602
0.875	30.00	30.21	7.2860	54.503	1.090	11.97	18	9.0787	67.914
0.880	29.09	29.30	7.3277	54.815	1.095	12.58	19	9.1204	68.226
0.885	28.19	28.38	7.3694	55.127					
0.890	27.30	27.49	7.4111	55.438	1.100	13.18	20	9.1621	68.537
0.895	26.42	26.60	7.4528	55.750	1.105	13.78	21	9.2038	68.849
					1.110	14.37	22	9.2455	69.161
0.900	25.76	25.72	7.4944	56.062	1.115	14.96	23	9.2872	69.473
0.905	24.70	24.85	7.5361	56.374	1.120	15.54	24	9.3289	69.785
0.910	23.85	23.99	7.5777	56.685	1.125	16.11	25	9.3706	70.097
0.915	23.01	23.14	7.6194	56.997	1.130	16.68	26	9.4123	70.409
0.920	22.17	22.30	7.6612	57.410	1.135	17.25	27	9.4540	70.721
0.925	21.35	21.47	7.7029	57.622	1.140	17.81	28	9.4957	71.032
0.930	20.54	20.65	7.7446	57.934	1.145	18.36	29	9.5374	71.344
0.935	19.73	19.84	7.7863	58.246	1.150	18.91	30	9.5790	71.656
0.940	18.94	19.03	7.8280	58.557	1.155	19.46	31	9.6207	71.968
0.945	18.15	18.24	7.8697	58.869	1.160	20.00	32	9.6624	72.280
0.950	17.37	17.45	7.9114	59.181	1.165	20.54	33	9.7041	72.592
0.955	16.60	16.67	7.9531	59.493	1.170	21.07	34	9.7458	72.904
0.960	15.83	15.90	7.9947	59.805	1.175	21.60	35	9.7875	73.216
0.965	15.08	15.13	8.0364	60.117	1.180	22.12	36	9.8292	73.528
0.970	14.33	14.38	8.0780	60.428	1.185	22.64	37	9.8709	73.840
0.975	13.59	13.63	8.1197	60.740	1.190	23.15	38	9.9126	74.151
0.980	12.86	12.89	8.1615	61.052	1.195	23.66	39	9.9543	74.463
0.985	12.13	12.15	8.2032	61.364					
0.990	11.41	11.43	8.2449	61.676	1.200	24.17	40	9.9960	74.775
0.995	10.70	10.71	8.2866	61.988	1.205	24.67	41	10.0377	75.087
					1.210	25.17	42	10.0793	75.399
Specific Gravity 60°/60°	°Be	°TW	lb/gal 60°F, wt in Air	lb/ft ³ at 60°F, wt in Air	1.215	25.66	43	10.1210	75.711
1.000	10.00	10.00	8.3283	62.300	1.220	26.15	44	10.1627	76.022
1.005	0.72	1	8.3700	62.612	1.225	26.63	45	10.2044	76.334
1.010	1.44	2	8.4117	62.924	1.230	27.11	46	10.2461	76.646
1.015	2.14	3	8.4534	63.236	1.235	27.59	47	10.2878	76.958
1.020	2.84	4	8.4950	63.547	1.240	28.06	48	10.3295	77.270
1.025	3.54	5	8.5367	63.859	1.245	28.53	49	10.3712	77.582
1.030	4.22	6	8.5784	64.171	1.250	29.00	50	10.4129	77.894
1.035	4.90	7	8.6201	64.483	1.255	29.46	51	10.4546	78.206
1.040	5.58	8	8.6618	64.795	1.260	29.92	52	10.4963	78.518
1.045	6.24	9	8.7035	65.107	1.265	30.38	53	10.5380	78.830
1.050	6.91	10	8.7452	65.419	1.270	30.83	54	10.5797	79.141
1.055	7.56	11	8.7869	65.731	1.275	31.27	55	10.6214	79.453
1.060	8.21	12	8.8286	66.042	1.280	31.72	56	10.6630	79.765
1.065	8.85	13	8.8703	66.354	1.285	32.16	57	10.7047	80.077
1.070	9.49	14	8.9120	66.666	1.290	32.60	58	10.7464	80.389
					1.295	33.03	59	10.7881	80.701
					1.300	33.46	60	10.8298	81.013

TABLE 51.64 (Continued)

Specific Gravity 60°/60°	°Be	°TW	lb/gal 60°F, wt in Air	lb/ft ³ at 60°F, wt in Air	Specific Gravity 60°/60°	°Be	°TW	lb/gal 60°F, wt in Air	lb/ft ³ at 60°F, wt in Air
1.305	33.89	61	10.8715	81.325	1.56	52.05	112	12.998	97.23
1.310	34.31	62	10.9132	81.636	1.57	52.64	114	13.081	97.85
1.315	34.73	63	10.9549	81.948	1.58	53.23	116	13.165	98.48
1.320	35.15	64	10.9966	82.260	1.59	53.81	118	13.248	99.10
1.325	35.57	65	11.0383	82.572	1.60	54.38	120	13.331	99.73
1.330	35.98	66	11.0800	82.884	1.61	54.94	122	13.415	100.35
1.335	36.39	67	11.1217	83.196	1.62	55.49	124	13.498	100.97
1.340	36.79	68	11.1634	83.508	1.63	56.04	126	13.582	101.60
1.345	37.19	69	11.2051	83.820	1.64	56.59	128	13.665	102.22
1.350	37.59	70	11.2467	84.131	1.65	57.12	130	13.748	102.84
1.355	37.99	71	11.2884	84.443	1.66	57.65	132	13.832	103.47
1.360	38.38	72	11.3301	84.755	1.67	58.17	134	13.915	104.09
1.365	38.77	73	11.3718	85.067	1.68	58.69	136	13.998	104.72
1.370	39.16	74	11.4135	85.379	1.69	59.20	138	14.082	105.34
1.375	39.55	75	11.4552	85.691	1.70	59.71	140	14.165	105.96
1.380	39.93	76	11.4969	86.003	1.71	60.20	142	14.249	106.59
1.385	40.31	77	11.5386	86.315	1.72	60.70	144	14.332	107.21
1.390	40.68	78	11.5803	86.626	1.73	61.18	146	14.415	107.83
1.395	41.06	79	11.6220	86.938	1.74	61.67	148	14.499	108.46
1.400	41.43	80	11.6637	87.250	1.75	62.14	150	14.582	109.08
1.405	41.80	81	11.7054	87.562	1.76	62.61	152	14.665	109.71
1.410	42.16	82	11.7471	87.874	1.77	63.08	154	14.749	110.32
1.415	42.53	83	11.7888	88.186	1.78	63.54	156	14.832	110.95
1.420	42.89	84	11.8304	88.498	1.79	63.99	158	14.916	111.58
1.425	43.25	85	11.8721	88.810	1.80	64.44	160	14.999	112.20
1.430	43.60	86	11.9138	89.121	1.81	64.89	162	15.082	112.82
1.435	43.95	87	11.9555	89.433	1.82	65.33	164	15.166	113.45
1.440	44.31	88	11.9972	89.745	1.83	65.77	166	15.249	114.07
1.445	44.65	89	12.0389	90.057	1.84	66.20	168	15.333	114.70
1.450	45.00	90	12.0806	90.369	1.85	66.62	170	15.416	115.31
1.455	45.34	91	12.1223	90.681	1.86	67.04	172	15.499	115.94
1.460	45.68	92	12.1640	90.993	1.87	67.46	174	15.583	116.56
1.465	46.02	93	12.2057	91.305	1.88	67.87	176	15.666	117.19
1.470	46.36	94	12.2473	91.616	1.89	68.28	178	15.750	117.81
1.475	46.69	95	12.2890	91.928	1.90	68.68	180	15.832	118.43
1.480	47.03	96	12.3307	92.240	1.91	69.08	182	15.916	119.06
1.485	47.36	97	12.3724	92.552	1.92	69.48	184	16.000	119.68
1.490	47.68	98	12.4141	92.864	1.93	69.87	186	16.083	120.31
1.495	48.01	99	12.4558	93.176	1.94	70.26	188	16.166	120.93
1.500	48.33	100	12.4975	93.488	1.95	70.64	190	16.250	121.56
1.51	48.97	102	12.581	94.11	1.96	71.02	192	16.333	122.18
1.52	49.61	104	12.644	94.79	1.97	71.40	194	16.417	122.80
1.53	50.23	106	12.748	95.36	1.98	71.77	196	16.500	123.43
1.54	50.84	108	12.831	95.98	1.99	72.14	198	16.583	124.05
1.55	51.45	110	12.914	96.61	2.00	72.50	200	16.667	124.68

51.6 STANDARD SIZES

51.6.1 Preferred Numbers

Selection of standard sizes or ratings of many diverse products can be performed advantageously through the use of a geometrically based progression introduced by C. Renard. He originally adopted as a basis a rule that would yield a 10th multiple of the value a after every 5th step of the series:

$$a \times q^5 = 10a \quad \text{or} \quad q = \sqrt[5]{10}$$

where the numerical series $a, a[\sqrt[5]{10}], a[\sqrt[5]{10}]^2, a[\sqrt[5]{10}]^3, [\sqrt[5]{10}]^4, 10a$, the values of which, to five significant figures, are $a, 1.5849a, 2.5119a, 3.9811a, 6.309a, 10a$.

Renard's idea was to substitute, for these values, more rounded but more practical values. He adopted as a a power of 10, positive, nil, or negative, obtaining the series 10, 16, 25, 40, 63, 100, which may be continued in both directions.

From this series, designated by the symbol R5, the R10, R20, R40 series were formed, each adopted ratio being the square root of the preceding one: $\sqrt[10]{10}, \sqrt[20]{10}, \sqrt[40]{10}$. Thus each series provided Renard with twice as many steps in a decade as the preceding one.

Preferred numbers are immediately applicable to commercial sizes and ratings of products. It is advantageous to minimize the number of initial sizes and also to have adequate provision for logical expansion if and when additional sizes are required. By making the initial sizes correspond to a coarse series such as R5, unnecessary expense can be avoided if subsequent demand for the product is disappointing. If, on the other hand, the product is accepted, intermediate sizes may be selected in a rational manner by using the next finer series R10, and so on. Such a procedure assures a justifiable relationship between successive sizes and is a decided contrast to haphazard selection.

The application of preferred numbers to raw material sizes and to the dimensions of parts also has enormously important potentialities. Under present conditions, commercial sizes of material are the result of a great many dissimilar gauge systems. The current trend in internationally acceptable metric sizing is to use preferred numbers. Even here, though, in the midst of the greatest opportunity for worldwide standardization through the acceptance of Renard series, we have fallen prey to our individualistic nature. The preferred number 1.6 is used by most nations as a standard 1.6 mm material thickness. German manufacturers, however, like 1.5 mm of the International Organization for Standardization (ISO) 497 for a more rounded preferred number. Similarly in metric screw sizes, 6.3 mm is consistent with the preferred number series; yet, 6.0 mm (more rounded) has been adopted as a standard fastener diameter.

The International Electrochemical Commission (IEC) used preferred numbers to establish standard current ratings in amperes as follows: 1, 1.25, 1.6, 2.5, 3.15, 4.5, 6.3. Notice that R10 series is used except for 4.5, which is a third step R20 series.

The American Wire Gauge size for copper wire is based on a geometric series. However, instead of using 1.1220, the rounded value of $\sqrt[20]{10}$, in $a \times q^{20} = 10a$, the q chosen is 1.123.

A special series of preferred numbers is used for designating the characteristic values of capacitors, resistors, inductors, and other electronic products. Instead of using the Renard series R5, R10, R20, R40, R80 as derived from the geometric series of numbers $10^{N/5}, 10^{N/10}, 10^{N/20}, 10^{N/40}, 10^{N/80}$, the geometric series used is $10^{N/6}, 10^{N/12}, 10^{N/24}, 10^{N/48}, 10^{N/96}, 10^{N/192}$, which are designated respectively E6, E12, E24, E48, E96, E192.

It should be evident that any series of preferred numbers can be generated to serve any specific case. Examples taken from the American National Standards Institute (ANSI) and ISO standards are reproduced in Tables 51.65–51.68.

TABLE 51.65 Basic Series of Preferred Numbers: R5, R10, R20, and R40 Series

R5	R10	R20	R40	Theoretical Values		Differences between Basic Series and Calculated Values (%)
				Mantissas of Logarithms	Calculated Values	
1.00	1.00	1.00	1.00	000	1.0000	0
			1.06	025	1.0593	+0.07
			1.12	050	1.1220	−0.18
			1.18	075	1.1885	−0.71
		1.25	1.25	100	1.2589	−0.71
			1.32	125	1.3335	−1.01
			1.40	150	1.4125	−0.88
			1.50	175	1.4962	+0.25
		1.60	1.60	200	1.5849	+0.95
			1.70	225	1.6788	+1.26
			1.80	250	1.7783	+1.22
			1.90	275	1.8836	+0.87
		2.00	2.00	300	1.9953	+0.24
			2.12	325	2.1135	+0.31
			2.24	350	2.2387	+0.06
			2.36	375	2.3714	−0.48
	2.50	2.50	2.50	400	2.5119	−0.47
			2.65	425	2.6607	−0.40
			2.80	450	2.8184	−0.65
			3.00	475	2.9854	+0.49
		3.15	3.15	500	3.1623	−0.39
			3.35	525	3.3497	+0.01
			3.55	550	3.5481	+0.05
			3.75	575	3.7584	−0.22
	4.00	4.00	4.00	600	3.9811	+0.47
			4.25	625	4.2170	+0.78
			4.50	650	4.4668	+0.74
			4.75	675	4.7315	+0.39
		5.00	5.00	700	5.0119	−0.24
			5.30	725	5.3088	−0.17
			5.60	750	5.6234	−0.42
			6.00	775	5.9566	+0.73
	6.30	6.30	6.30	800	6.3096	−0.15
			6.70	825	6.6834	+0.25
			7.10	850	7.0795	+0.29
			7.50	875	7.4989	+0.01
		8.00	8.00	900	7.9433	+0.71
			8.50	925	8.4140	+1.02
			9.00	950	8.9125	+0.98
			9.50	975	9.4406	+0.63
	10.00	10.00	10.00	000	10.0000	0
			10.00	000	10.0000	0

TABLE 51.66 Basic Series of Preferred Numbers: R80 Series

1.00	1.80	3.15	5.60
1.03	1.85	3.25	5.80
1.06	1.90	3.35	6.00
1.09	1.95	3.45	6.15
1.12	2.00	3.55	6.30
1.15	2.06	3.65	6.50
1.18	2.12	3.75	6.70
1.22	2.18	3.87	6.90
1.25	2.24	4.00	7.10
1.28	2.30	4.12	7.30
1.32	2.36	4.25	7.50
1.36	2.43	4.37	7.75
1.40	2.50	4.50	8.00
1.45	2.58	4.62	8.25
1.50	2.65	4.75	8.50
1.55	2.72	4.87	8.75
1.60	2.80	5.00	9.00
1.65	2.90	5.15	9.25
1.70	3.00	5.20	9.50
1.75	3.07	5.45	9.75

TABLE 51.67 Expansion of R5 Series

Preferred Number	Divided by 10	Multiplied by 10	Multiplied by 100	Multiplied by 1000
1.0	0.10	10	100	1000
1.6	0.16	16	160	1600
2.5	0.25	25	250	2500
4.0	0.40	40	400	4000
6.3	0.63	63	630	6300

TABLE 51.68 Rounding of Preferred Numbers^a

Preferred Number	First Rounding	Second Rounding
1.12	1.1	1.1
1.25	1.25	1.2
1.60	1.6	1.5 ^a
2.24	2.2	2.2
3.15	3.2	3.0
3.55	3.6	3.5
5.60	5.6	5.5
6.30	6.3	6.0
7.10	7.1	7.0

^aRounded only when using the R5 or R10 series.

Applicable Documents

Adoption of Renard’s preferred number system by international standardization bodies resulted in a host of national standards being generated for particular applications. The current organization in the United States that is charged with generating American national standards is the ANSI. Accordingly, the following national and international standards are in use in the United States.

ANSI Z17.1-1973	American National Standard for Preferred Numbers
ANSI C83.2-1971	American National Standard Preferred Values for Components for Electronic Equipment
EIA Standard RS-385	Preferred Values for Components for Electronic Equipment (issued by the Electronics Industries Association; Same as ANSI C83.2-1971)
ISO 3-1973	Preferred numbers—series of preferred numbers
ISO 17-1973	Guide to the use of preferred numbers and of series of preferred numbers
ISO 497-1973	Guide to the choice of series of preferred numbers and of series containing more rounded values of preferred numbers

Table 51.67 shows the expansibility of preferred numbers in the positive direction. The same expansibility can be made in the negative direction. Table 51.68 shows a deviation by roundings for cases where adhering to a basic preferred number would be absurd as in 31.5 teeth in a gear when clearly 32 makes sense.

51.6.2 Gages

51.6.2.1 Wire Gages The sizes of wires having a diameter less than $\frac{1}{2}$ in. are usually stated in terms of certain arbitrary scales called “gages.” The size or gage number of a solid wire refers to the cross section of the wire perpendicular to its length; the size or gage number of a stranded wire refers to the total cross section of the constituent wires, irrespective of the pitch of the spiraling. Larger wires are usually described in terms of their area expressed in circular mils. A circular mil is the area of a circle 1 mil in diameter, and the area of any circle in circular mils is equal to the square of its diameter in mils.

TABLE 51.69 U.S. Standard Gage^a for Sheet and Plate Iron and Steel and Its Extension^b

Gage Number	Weight per Square Foot		Weight per Square Meter	Approximate Thickness			
				Wrought Iron, 480 lb/ft ³		Steel and open- hearth Iron, 489.6 lb/ft ³	
	oz.	lb	kg	in.	mm	in.	mm
0000000	320	20.00	97.65	0.500	12.70	0.490	12.45
000000	300	18.75	91.55	0.469	11.91	0.460	11.67
00000	280	17.50	85.44	0.438	11.11	0.429	10.90
0000	260	16.25	79.34	0.406	10.32	0.398	10.12
000	240	15.00	73.24	0.375	9.52	0.368	9.34
00	220	13.75	67.13	0.344	8.73	0.337	8.56
0	200	12.50	61.03	0.312	7.94	0.306	7.78
1	180	11.25	54.93	0.2812	7.14	0.2757	7.00
2	170	10.62	51.88	0.2656	6.75	0.2604	6.62
3	160	10.00	48.82	0.2500	6.35	0.2451	6.23
4	150	9.375	45.77	0.2344	5.95	0.2298	5.84

TABLE 51.69 (Continued)

Gage Number	Weight per Square Foot		Weight per Square Meter	Approximate Thickness			
				Wrought Iron, 480 lb/ft ³		Steel and open- hearth Iron, 489.6 lb/ft ³	
	oz.	lb	kg	in.	mm	in.	mm
5	140	8.750	42.72	0.2188	5.56	0.2145	5.45
6	130	8.125	39.67	0.2031	5.16	0.1991	5.06
7	120	7.500	36.62	0.1875	4.76	0.1838	4.67
8	110	6.875	33.57	0.1719	4.37	0.1685	4.28
9	100	6.250	30.52	0.1562	3.97	0.1532	3.89
10	90	5.625	27.46	0.1406	3.57	0.1379	3.50
11	80	5.000	24.41	0.1250	3.18	0.1225	3.11
12	70	4.375	21.36	0.1094	2.778	0.1072	2.724
13	60	3.750	18.31	0.0938	2.381	0.0919	2.335
14	50	3.125	15.26	0.0781	1.984	0.0766	1.946
15	45	2.812	13.73	0.0703	1.786	0.0689	1.751
16	40	2.500	12.21	0.0625	1.588	0.0613	1.557
17	36	2.250	10.99	0.0562	1.429	0.0551	1.400
18	32	2.000	9.765	0.0500	1.270	0.0490	1.245
19	28	1.750	8.544	0.0438	1.111	0.0429	1.090
20	24	1.500	7.324	0.0375	0.952	0.0368	0.934
21	22	1.375	6.713	0.0344	0.873	0.0337	0.856
22	20	1.250	6.103	0.0312	0.794	0.0306	0.778
23	18	1.125	5.493	0.0281	0.714	0.0276	0.700
24	16	1.000	4.882	0.0250	0.635	0.0245	0.623
25	14	0.8750	4.272	0.0219	0.556	0.0214	0.545
26	12	0.7500	3.662	0.0188	0.476	0.0184	0.467
27	11	0.6875	3.357	0.0172	0.437	0.0169	0.428
28	10	0.6250	3.052	0.0156	0.397	0.0153	0.389
29	9	0.5625	2.746	0.0141	0.357	0.0138	0.350
30	8	0.5000	2.441	0.0125	0.318	0.0123	0.311
31	7	0.4375	2.136	0.0109	0.278	0.0107	0.272
32	6½	0.4062	1.983	0.0102	0.258	0.0100	0.253
33	6	0.3750	1.831	0.0094	0.238	0.0092	0.233
34	5½	0.3438	1.678	0.0086	0.218	0.0084	0.214
35	5	0.3125	1.526	0.0078	0.198	0.0077	0.195
36	4½	0.2812	1.373	0.0070	0.179	0.0069	0.175
37	4¼	0.2656	1.297	0.0066	0.169	0.0065	0.165
38	4	0.2500	1.221	0.0062	0.159	0.0061	0.156
39	3¾	0.2344	1.144	0.0059	0.149	0.0057	0.146
40	3½	0.2188	1.068	0.0055	0.139	0.0054	0.136
41	3⅜	0.2109	1.030	0.0053	0.134	0.0052	0.131
42	3¼	0.2031	0.9917	0.0051	0.129	0.0050	0.126
43	3⅛	0.1953	0.9536	0.0049	0.124	0.0048	0.122
44	3	0.1875	0.9155	0.0047	0.119	0.0046	0.117

^aFor the Galvanized Sheet Gage, add 2.5 oz to the weight per square foot as given in the table. Gage numbers below 8 and above 34 are not used in the Galvanized Sheet Gage.

^bGage numbers greater than 38 were not in the standard as set up by law but are in general use.

TABLE 51.70 American Wire Gage: Weights of Copper, Aluminum, and Brass Sheets and Plates

Gage Number	Thickness		Approximate Weight, ^a lb/ft ²		
	in.	mm	Copper	Aluminum	Commercial (High) Brass
0000	0.4600	11.68	21.27	6.49	20.27
000	0.4096	10.40	18.94	5.78	18.05
00	0.3648	9.266	16.87	5.14	16.07
0	0.3249	8.252	15.03	4.58	14.32
1	0.2893	7.348	13.38	4.08	12.75
2	0.2576	6.544	11.91	3.632	11.35
3	0.2294	5.827	10.61	3.234	10.11
4	0.2043	5.189	9.45	2.880	9.00
5	0.1819	4.621	8.41	2.565	8.01
6	0.1620	4.115	7.49	2.284	7.14
7	0.1443	3.665	6.67	2.034	6.36
8	0.1285	3.264	5.94	1.812	5.66
9	0.1144	2.906	5.29	1.613	5.04
10	0.1019	2.588	4.713	1.437	4.490
11	0.0907	2.305	4.195	1.279	3.996
12	0.0808	2.053	3.737	1.139	3.560
13	0.0720	1.828	3.330	1.015	3.172
14	0.0641	1.628	2.965	0.904	2.824
15	0.0571	1.450	2.641	0.805	2.516
16	0.0508	1.291	2.349	0.716	2.238
17	0.0453	1.150	2.095	0.639	1.996
18	0.0403	1.024	1.864	0.568	1.776
19	0.0359	0.9116	1.660	0.506	1.582
20	0.0320	0.8118	1.480	0.451	1.410
21	0.0285	0.7230	1.318	0.402	1.256
22	0.0253	0.6438	1.170	0.3567	1.115
23	0.0226	0.5733	1.045	0.3186	0.996
24	0.0201	0.5106	0.930	0.2834	0.886
25	0.0179	0.4547	0.828	0.2524	0.789
26	0.0159	0.4049	0.735	0.2242	0.701
27	0.0142	0.3606	0.657	0.2002	0.626
28	0.0126	0.3211	0.583	0.1776	0.555
29	0.0113	0.2859	0.523	0.1593	0.498
30	0.0100	0.2546	0.4625	0.1410	0.4406
31	0.00893	0.2268	0.4130	0.1259	0.3935
32	0.00795	0.2019	0.3677	0.1121	0.3503
33	0.00708	0.1798	0.3274	0.0998	0.3119
34	0.00630	0.1601	0.2914	0.0888	0.2776
35	0.00561	0.1426	0.2595	0.0791	0.2472
36	0.00500	0.1270	0.2312	0.0705	0.2203
37	0.00445	0.1131	0.2058	0.0627	0.1961
38	0.00397	0.1007	0.1836	0.0560	0.1749
39	0.00353	0.0897	0.1633	0.0498	0.1555
40	0.00314	0.0799	0.1452	0.0443	0.1383

^aAssumed specific gravities or densities in grams per cubic centimeter; copper, 8.89; aluminum, 2.71; brass, 8.47.

TABLE 51.71 Comparison of Wire Gage Diameters in Mils^a

Gage No.	American Wire Gage (Brown & Sharpe)	Steel Wire Gage	Birmingham Wire Gage (Stubs ^c)	Old English Wire Gage (London)	Stubs' Steel Wire Gage	(British) Standard Wire Gage	Metric Gage ^b
7-0	—	490.0	—	—	—	500	—
6-0	—	461.5	—	—	—	464	—
5-0	—	430.5	—	—	—	432	—
4-0	460	393.8	454	454	—	400	—
3-0	410	362.5	425	425	—	372	—
2-0	365	331.0	380	380	—	348	—
0	325	306.5	340	340	—	324	—
1	289	283.0	300	300	227	300	3.94
2	258	262.5	284	284	219	276	7.87
3	229	243.7	259	259	212	252	11.8
4	204	225.3	238	238	207	232	15.7
5	182	207.0	220	220	204	212	19.7
6	162	192.0	203	203	201	192	23.6
7	144	177.0	180	180	199	176	27.6
8	128	162.0	165	165	197	160	31.5
9	114	148.3	148	148	194	144	35.4
10	102	135.0	134	134	191	128	39.4
11	91	120.5	120	120	188	116	—
12	81	105.5	109	109	185	104	47.2
13	72	91.5	95	95	182	92	—
14	64	80.0	83	83	180	80	55.1
15	57	72.0	72	72	178	72	—
16	51	62.5	65	65	175	64	63.0
17	45	54.0	58	58	172	56	—
18	40	47.5	49	49	168	48	70.9
19	36	41.0	42	42	164	40	—
20	32	34.8	35	35	161	36	78.7
21	28.5	31.7	32	31.5	157	32	—
22	25.3	28.6	28	29.5	155	28	—
23	22.6	25.8	25	27.0	153	24	—
24	20.1	23.0	22	25.0	151	22	—
25	17.9	20.4	20	23.0	148	20	98.4
26	15.9	18.1	18	20.5	146	18	—
27	14.2	17.3	16	18.75	143	16.4	—
28	12.6	16.2	14	16.50	139	14.8	—
29	11.3	15.0	13	15.50	134	13.6	—
30	10.0	14.0	12	13.75	127	12.4	118
31	8.9	13.2	10	12.25	120	11.6	—
32	8.0	12.8	9	11.25	115	10.8	—
33	7.1	11.8	8	10.25	112	10.0	—
34	6.3	10.4	7	9.50	110	9.2	—
35	5.6	9.5	5	9.00	108	8.4	138
36	5.0	9.0	4	7.50	106	7.6	—
37	4.5	8.5	—	6.50	103	6.8	—

(continued)

TABLE 51.71 (Continued)

Gage No.	American Wire Gage (Brown & Sharpe)	Steel Wire Gage	Birmingham Wire Gage (Stubs')	Old English Wire Gage (London)	Stubs' Steel Wire Gage	(British) Standard Wire Gage	Metric Gage ^b
38	4.0	8.0	—	5.75	101	6.0	—
39	3.5	7.5	—	5.00	99	5.2	—
40	3.1	7.0	—	4.50	97	4.8	157
41	—	6.6	—	—	95	4.4	—
42	—	6.2	—	—	92	4.0	—
43	—	6.0	—	—	88	3.6	—
44	—	5.8	—	—	85	3.2	—
45	—	5.5	—	—	81	2.8	177
46	—	5.2	—	—	79	2.4	—
47	—	5.0	—	—	77	2.0	—
48	—	4.8	—	—	75	1.6	—
49	—	4.6	—	—	72	1.2	—
50	—	4.4	—	—	69	1.0	197

^aBureau of Standards, Circulars No. 31 and No. 67.^bFor diameters corresponding to metric gage numbers, 1.2, 1.4, 1.6, 1.8, 2.5, 3.5, and 4.5, divide those of 12, 14, etc., by 10.**TABLE 51.72** Standard Engineering Drawing Sizes^a

<i>Flat Sizes^b</i>					
Size Designation	Width ^c (Vertical)	Length (Horizontal)	Margin		
			Horizontal	Vertical	
A (horizontal)	8.5	11.0	0.38	0.25	
A (vertical)	11.0	8.5	0.25	0.38	
B	11.0	17.0	0.38	0.62	
C	17.0	22.0	0.75	0.50	
D	22.0	34.0	0.50	1.00	
E	34.0	44.0	1.00	0.50	
F	28.0	40.0	0.50	0.50	
<i>Roll Sizes</i>					
Size Designation	Width ^b (Vertical)	Length ^c (Horizontal)		Margin ^c	
		Min	Max	Horizontal	Vertical
G	11.0	22.5	90.0	0.38	0.50
H	28.0	44.0	143.0	0.50	0.50
J	34.0	55.0	176.0	0.50	0.50
K	40.0	55.0	143.0	0.50	0.50

^aSee ANSI Y14.1-1980.^bAll dimensions are in inches.^cNot including added protective margins.

TABLE 51.73 Eleven International Paper Sizes

International Paper Size	Millimeters	Inches, Approximate
A-0	841 × 1189	33 ¹ / ₈ × 46 ³ / ₄
A-1	594 × 841	23 ³ / ₈ × 33 ¹ / ₈
A-2	420 × 594	16 ¹ / ₂ × 23 ³ / ₈
A-3	297 × 420	11 ³ / ₄ × 16 ¹ / ₂
A-4	210 × 297	8 ¹ / ₄ × 11 ³ / ₄
A-5	148 × 210	5 ⁷ / ₈ × 8 ¹ / ₄
A-6	105 × 148	4 ¹ / ₈ × 5 ⁷ / ₈
A-7	74 × 105	2 ⁷ / ₈ × 4 ¹ / ₈
A-8	52 × 74	2 × 2 ⁷ / ₈
A-9	37 × 52	1 ¹ / ₂ × 2
A-10	26 × 37	1 × 1 ¹ / ₂

51.6.3 Paper Sizes

51.6.3.1 International Paper Sizes Countries that are committed to the International System of Units (SI) have a standard series of paper sizes for printing, writing, and drafting. These paper sizes are called the “international paper sizes.”

The advantages of the international paper sizes are as follows:

- 1. The ratio of width to length remains constant for every size, namely:

$$\frac{\text{Width}}{\text{Length}} = \frac{1}{\sqrt{2}} \quad \text{or} \quad \frac{1}{1.414} \text{ approximately}$$

Since this is the same ratio as the D aperture in the unitized 35-mm microfilm frame, the advantages are apparent.

- 2. If a sheet is cut in half, that is, if the $\sqrt{2}$ length is cut in half, the two halves retain the constant width-to-length ratio of $1/\sqrt{2}$. No other ratio could do this.
- 3. All international sizes are created from the A-0 size by single cuts without waste. In storing or stacking they fit together like parts of a jigsaw puzzle—without waste.

51.6.4 Sieve Sizes

TABLE 51.74 Tyler Standard Screen Scale Sieves

This screen scale has as its base an opening of 0.0029 in., which is the opening in 200-mesh 0.0021-in. wire, the standard sieve, as adopted by the Bureau of Standards of the U.S. government, the openings increasing in the ratio of the square root of 2 or 1.414.

Where a closer sizing is required, column 5 shows the Tyler Standard Screen Scale with intermediate sieves. In this series the sieve openings increase in the ratio of the fourth root of 2, or 1.189.

Tyler Stan- dard Screen Scale $\sqrt{2}$ or 1.414 Open- ings (in.) (1)	Every Other Sieve from 0.0029 to 0.742 in., Ratio of 2 to 1 (2)	Every Sieve from 0.0041 to 1.050 in., Ratio of 2 to 1 (3)	Every Fourth Sieve from 0.0029 to 0.742 in., Ratio of 4 to 1 (4)	For Closer Sizing Sieves from 0.0029 to 1.050 in., Ratio $\sqrt[4]{2}$ or 1.189 (5)	openings (mm) (6)	Openings in Frac- tions of inch (approx.) (7)	Mesh (8)	Diameter of Wire (9)
1.050	—	1.050	—	1.050	26.67	1	—	0.148
—	—	—	—	0.883	22.43	$\frac{7}{8}$	—	0.135
0.742	0.742	—	0.742	0.742	18.85	$\frac{3}{4}$	—	0.135
—	—	—	—	0.624	15.85	$\frac{5}{8}$	—	0.120
0.525	—	0.525	—	0.525	13.33	$\frac{1}{2}$	—	0.105
—	—	—	—	0.441	11.20	$\frac{7}{16}$	—	0.105
0.371	0.371	—	—	0.371	9.423	$\frac{3}{8}$	—	0.092
—	—	—	—	0.312	7.925	$\frac{5}{16}$	$2\frac{1}{2}$	0.088
0.263	—	0.263	—	0.263	6.680	$\frac{1}{4}$	3	0.070
—	—	—	—	0.221	5.613	$\frac{7}{32}$	$3\frac{1}{2}$	0.065
0.185	0.185	—	0.185	0.185	4.699	$\frac{3}{16}$	4	0.065
—	—	—	—	0.156	3.962	$\frac{5}{32}$	5	0.044
0.131	—	0.131	—	0.131	3.327	$\frac{1}{8}$	6	0.036
—	—	—	—	0.110	2.794	$\frac{7}{64}$	7	0.0328
0.093	0.093	—	—	0.093	2.362	$\frac{3}{32}$	8	0.032
—	—	—	—	0.078	1.981	$\frac{5}{84}$	9	0.033
0.065	—	0.065	—	0.065	1.651	$\frac{1}{16}$	10	0.035
—	—	—	—	0.055	1.397	—	12	0.028
0.046	0.046	—	0.046	0.046	1.168	$\frac{3}{64}$	14	0.025
—	—	—	—	0.0390	0.991	—	16	0.0235
0.0328	—	0.0328	—	0.0328	0.833	$\frac{1}{32}$	20	0.0172
—	—	—	—	0.0276	0.701	—	24	0.0141
0.0232	0.0232	—	—	0.0232	0.589	—	28	0.0125
—	—	—	—	0.0195	0.495	—	32	0.0118
0.0164	—	0.0164	—	0.0164	0.417	$\frac{1}{64}$	35	0.0122
—	—	—	—	0.0138	0.351	—	42	0.0100
0.0116	0.0116	—	0.0116	0.0116	0.295	—	48	0.0092
—	—	—	—	0.0097	0.246	—	60	0.0070
0.0082	—	0.0082	—	0.0082	0.208	—	65	0.0072
—	—	—	—	0.0069	0.175	—	80	0.0056
0.0058	0.0058	—	—	0.0058	0.147	—	100	0.0042
—	—	—	—	0.0049	0.124	—	115	0.0038
0.0041	—	0.0041	—	0.0041	0.104	—	150	0.0026
—	—	—	—	0.0035	0.088	—	170	0.0024
0.0029	0.0029	—	0.0029	0.0029	0.074	—	200	0.0021

TABLE 51.75 Nominal Dimensions, Permissible Variations, and Limits for Woven Wire Cloth of Standard Sieves, U.S. Series, ASTM Standard^a

Size or Sieve Designation		Sieve Opening		Permissible Variations in Average Opening (±%)	Permissible Variations in Maximum Opening (±%)	Wire Diameter	
μm	No.	mm	in. (approx. equivalents)			mm	in. (approx. equivalents)
5660	3½	5.66	0.233	3	10	1.28–1.90	0.050–0.075
4760	4	4.76	0.187	3	10	1.14–1.68	0.045–0.066
4000	5	4.00	0.157	3	10	1.00–1.47	0.039–0.058
3360	6	3.36	0.132	3	10	0.87–1.32	0.034–0.052
2830	7	2.83	0.111	3	10	0.80–1.20	0.031–0.047
2380	8	2.38	0.0937	3	10	0.74–1.10	0.0291–0.0433
2000	10	2.00	0.0787	3	10	0.68–1.00	0.0268–0.0394
1680	12	1.68	0.0661	3	10	0.62–0.90	0.0244–0.0354
1410	14	1.41	0.0555	3	10	0.56–0.80	0.0220–0.0315
1190	16	1.19	0.0469	3	10	0.50–0.70	0.0197–0.0276
1000	18	1.00	0.0394	5	15	0.43–0.62	0.0169–0.0244
840	20	0.84	0.0331	5	15	0.38–0.55	0.0150–0.0217
710	25	0.71	0.0280	5	15	0.33–0.48	0.0130–0.0189
590	30	0.59	0.0232	5	15	0.29–0.42	0.0114–0.0165
500	35	0.50	0.0197	5	15	0.26–0.37	0.0102–0.0146
420	40	0.42	0.0165	5	25	0.23–0.33	0.0091–0.0130
350	45	0.35	0.0138	5	25	0.20–0.29	0.0079–0.0114
297	50	0.297	0.0117	5	25	0.170–0.253	0.0067–0.0100
250	60	0.250	0.0098	5	25	0.149–0.220	0.0059–0.0087
210	70	0.210	0.0083	5	25	0.130–0.187	0.0051–0.0074
177	80	0.177	0.0070	6	40	0.114–0.154	0.0045–0.0061
149	100	0.149	0.0059	6	40	0.096–0.125	0.0038–0.0049
125	120	0.125	0.0049	6	40	0.079–0.103	0.0031–0.0041
105	140	0.105	0.0041	6	40	0.063–0.087	0.0025–0.0034
88	170	0.088	0.0035	6	40	0.054–0.073	0.0021–0.0029
74	200	0.074	0.0029	7	60	0.045–0.061	0.0018–0.0024
62	230	0.062	0.0024	7	90	0.039–0.052	0.0015–0.0020
53	270	0.053	0.0021	7	90	0.035–0.046	0.0014–0.0018
44	325	0.044	0.0017	7	90	0.031–0.040	0.0012–0.0016
37	400	0.037	0.0015	7	90	0.023–0.035	0.0009–0.0014

^aFor sieves from the 1000-μm (No. 18) to the 37-μm (No. 400) size, inclusive, not more than 5% of the openings shall exceed the nominal opening by more than one-half of the permissible variation in the maximum opening.

51.6.5 Standard Structural Sizes — Steel

Steel Sections. Tables 76–83 give the dimensions, weights, and properties of *rolled steel* structural sections, including wide-flange sections, American standard beams, channels, angles, tees, and zebs. The values for the various structural forms, taken from the eighth edition, 1980, of *Steel Construction*, by the kind permission of the publisher, the American Institute of Steel Construction, give the section specifications required in designing steel structures. The theory of design is covered in Section 4—Mechanics of Deformable Bodies.

TABLE 51.76 Properties of Wide-Flange Sections


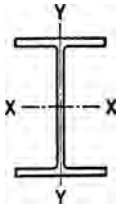
<div><div></div><div></div></div>												
Nominal Size (in.)	Weight per Foot (lb)	Area (in. ²)	Depth (in.)	Flange		Web Thickness (in.)	Axis X-X			Axis Y-Y		
				Width (in.)	Thickness (in.)		<i>I</i> (in. ⁴)	<i>S</i> (in. ³)	<i>r</i> (in.)	<i>I</i> (in. ⁴)	<i>S</i> (in. ³)	<i>r</i> (in.)
36 × 16½	300	88.17	36.72	16.655	1.680	0.945	20290.2	1105.1	15.17	1225.2	147.1	3.73
	280	82.32	36.50	16.595	1.570	0.885	18819.3	1031.2	15.12	1127.5	135.9	3.70
	260	76.56	36.24	16.555	1.440	0.845	17233.8	951.1	15.00	1020.6	123.3	3.65
	245	72.03	36.06	16.512	1.350	0.802	16092.2	892.5	14.95	944.7	114.4	3.62
	230	67.73	35.88	16.475	1.260	0.765	14988.4	835.5	14.88	870.9	105.7	3.59
36 × 12	194	57.11	36.48	12.117	1.260	0.770	12103.4	663.6	14.56	355.4	58.7	2.49
	182	53.54	36.32	12.072	1.180	0.725	11281.5	621.2	14.52	327.7	54.3	2.47
	170	49.98	36.16	12.027	1.100	0.680	10470.0	579.1	14.47	300.6	50.0	2.45
	160	47.09	36.00	12.000	1.020	0.653	9738.8	541.0	14.38	275.4	45.9	2.42
	150	44.16	35.84	11.972	0.940	0.625	9012.1	502.9	14.29	250.4	41.8	2.38
33 × 15¾	240	70.52	33.50	15.865	1.400	0.830	13585.1	811.1	13.88	874.3	110.2	3.52
	220	64.73	33.25	15.810	1.275	0.775	12312.1	740.6	13.79	782.4	99.0	3.48
	200	58.79	33.00	15.750	1.150	0.715	11048.2	669.6	13.71	691.7	87.8	3.43
33 × 11½	152	44.71	33.50	11.565	1.055	0.635	8147.6	486.4	13.50	256.1	44.3	2.39
	141	41.51	33.31	11.535	0.960	0.605	7442.2	446.8	13.39	229.7	39.8	2.35
	130	38.26	33.10	11.510	0.855	0.580	6699.0	404.8	13.23	201.4	35.0	2.29
30 × 15	210	61.78	30.38	15.105	1.315	0.775	9872.4	649.9	12.64	707.9	93.7	3.38
	190	55.90	30.12	15.040	1.185	0.710	8825.9	586.1	12.57	624.6	83.1	3.34
	172	50.65	29.88	14.985	1.065	0.655	7891.5	528.2	12.48	550.1	73.4	3.30
30 × 10½	132	38.83	30.30	10.551	1.000	0.615	5753.1	379.7	12.17	185.0	35.1	2.18
	124	36.45	30.16	10.521	0.930	0.585	5347.1	354.6	12.11	169.7	32.3	2.16
	116	34.13	30.00	10.500	0.850	0.564	4919.1	327.9	12.00	153.2	29.2	2.12
	108	31.77	29.82	10.484	0.760	0.548	4461.0	299.2	11.85	135.1	25.8	2.06
27 × 14	177	52.10	27.31	14.090	1.190	0.725	6728.6	492.8	11.36	518.9	73.7	3.16
	160	47.04	27.08	14.023	1.075	0.658	6018.6	444.5	11.31	458.0	65.3	3.12
	145	42.68	26.88	13.965	0.975	0.600	5414.3	402.9	11.26	406.9	58.3	3.09
27 × 10	114	33.53	27.28	10.070	0.932	0.570	4080.5	299.2	11.03	149.6	29.7	2.11
	102	30.01	27.07	10.018	0.827	0.518	3604.1	266.3	10.96	129.5	25.9	2.08
	94	27.65	26.91	9.990	0.747	0.490	3266.7	242.8	10.87	115.1	23.0	2.04
24 × 14	160	47.04	24.72	14.091	1.135	0.656	5110.3	413.5	10.42	492.6	69.9	3.23
	145	42.62	24.49	14.043	1.020	0.608	4561.0	372.5	10.34	434.3	61.8	3.19
	130	38.21	24.25	14.000	0.900	0.565	4009.5	330.7	10.24	375.2	53.6	3.13
24 × 12	120	35.29	24.31	12.088	0.930	0.556	3635.3	299.1	10.15	254.0	42.0	2.68
	110	32.36	24.16	12.042	0.855	0.510	3315.0	274.4	10.12	229.1	38.0	2.66
	100	29.43	24.00	12.000	0.775	0.468	2987.3	248.9	10.08	203.5	33.9	2.63

TABLE 51.76 (Continued)

Nominal Size (in.)	Weight per Foot (lb)	Area (in. ²)	Depth (in.)	Flange		Web Thickness (in.)	Axis X-X			Axis Y-Y		
				Width (in.)	Thickness (in.)		<i>I</i> (in. ⁴)	<i>S</i> (in. ³)	<i>r</i> (in.)	<i>I</i> (in. ⁴)	<i>S</i> (in. ³)	<i>r</i> (in.)
24 × 9	94	27.63	24.29	9.061	0.872	0.516	2683.0	220.9	9.85	102.2	22.6	1.92
	84	24.71	24.09	9.015	0.772	0.470	2364.3	196.3	9.78	88.3	19.6	1.89
	76	22.37	23.91	8.985	0.682	0.440	2096.4	175.4	9.68	76.5	17.0	1.85
21 × 13	142	41.76	21.46	13.132	1.095	0.659	3403.1	317.2	9.03	385.9	58.8	3.04
	127	37.34	21.24	13.061	0.985	0.588	3017.2	284.1	8.99	338.6	51.8	3.01
	112	32.93	21.00	13.000	0.865	0.527	2620.6	249.6	8.92	289.7	44.6	2.96
21 × 9	96	28.21	21.14	9.038	0.935	0.575	2088.9	197.6	8.60	109.3	24.2	1.97
	82	24.10	20.86	8.962	0.795	0.499	1752.4	168.0	8.53	89.6	20.0	1.93
21 × 8¼	73	21.46	21.24	8.295	0.740	0.455	100.3	150.7	8.64	66.2	16.0	1.76
	68	20.02	21.13	8.270	0.685	0.430	1478.3	139.9	8.59	60.4	14.6	1.74
	62	18.23	20.99	8.240	0.615	0.400	1326.8	126.4	8.53	53.1	12.9	1.71
18 × 11¾	114	33.51	18.48	11.833	0.991	0.595	2033.8	220.1	7.79	255.6	43.2	2.76
	105	30.86	18.32	11.792	0.911	0.554	1852.5	202.2	7.75	231.0	39.2	2.73
	96	28.22	18.16	11.750	0.831	0.512	1674.7	184.4	7.70	206.8	35.2	2.71
18 × 8¾	85	24.97	18.32	8.838	0.911	0.526	1429.9	156.1	7.57	99.4	22.5	2.00
	77	22.63	18.16	8.787	0.831	0.475	1286.8	141.7	7.54	88.6	20.2	1.98
	70	20.56	18.00	8.750	0.751	0.438	1153.9	128.2	7.49	78.5	17.9	1.95
18 × 7½	64	18.80	17.87	8.715	0.686	0.403	1045.8	117.0	7.46	70.3	16.1	1.93
	60	17.64	18.25	7.558	0.695	0.416	984.0	107.8	7.47	47.1	12.5	1.63
	55	16.19	18.12	7.532	0.630	0.390	889.9	98.2	7.41	42.0	11.1	1.61
16 × 11½	50	14.71	18.00	7.500	0.570	0.358	800.6	89.0	7.38	37.2	9.9	1.59
	96	28.22	16.32	11.533	0.875	0.535	1355.1	166.1	6.93	207.2	35.9	2.71
	88	25.87	16.16	11.502	0.795	0.504	1222.6	151.3	6.87	185.2	32.2	2.67
16 × 8½	78	22.92	16.32	8.586	0.875	0.529	1042.6	127.8	6.74	87.5	20.4	1.95
	71	20.86	16.16	8.543	0.795	0.486	936.9	115.9	6.70	77.9	18.2	1.93
	64	18.80	16.00	8.500	0.715	0.443	833.8	104.2	6.66	68.4	16.1	1.91
16 × 7	58	17.04	15.86	8.464	0.645	0.407	746.4	94.1	6.62	60.5	14.3	1.88
	50	14.70	16.25	7.073	0.628	0.380	655.4	80.7	6.68	34.8	9.8	1.54
	45	13.24	16.12	7.039	0.563	0.346	583.3	72.4	6.64	30.5	8.7	1.52
14 × 16	40	11.77	16.00	7.000	0.503	0.307	515.5	64.4	6.62	26.5	7.6	1.50
	36	10.59	15.85	6.992	0.428	0.299	446.3	56.3	6.49	22.1	6.3	1.45
	426	125.25	18.69	16.695	3.033	1.875	6610.3	707.4	7.26	2359.5	282.7	4.34
14 × 16	398	116.98	18.31	16.590	2.843	1.770	6013.7	656.9	7.17	2169.7	261.6	4.31
	370	108.78	17.94	16.475	2.658	1.655	5454.2	608.1	7.08	1986.0	241.1	4.27
	342	100.59	17.56	16.365	2.468	1.545	4911.5	559.4	6.99	1806.9	220.8	4.24
	314	92.30	17.19	16.235	2.283	1.415	4399.4	511.9	6.90	1631.4	201.0	4.20
	287	84.37	16.81	16.130	2.093	1.310	3912.1	465.5	6.81	1466.5	181.8	4.17
	264	77.63	16.50	16.025	1.938	1.205	3526.0	427.4	6.74	1331.2	166.1	4.14
	246	72.33	16.25	15.945	1.813	1.125	3228.9	397.4	6.68	1226.6	153.9	4.12
	237	69.69	16.12	15.910	1.748	1.090	3080.9	382.2	6.65	1174.8	147.7	4.11
	228	67.06	16.00	15.865	1.688	1.045	2942.4	367.8	6.62	1124.8	141.8	4.10
	219	64.36	15.87	15.825	1.623	1.005	2798.2	352.6	6.59	1073.2	135.6	4.08
	211	62.07	15.75	15.800	1.563	0.980	2671.4	339.2	6.56	1028.6	130.2	4.07

(continued)

TABLE 51.76 (Continued)

Nominal Size (in.)	Weight per Foot (lb)	Area (in. ²)	Depth (in.)	Flange		Web Thickness (in.)	Axis X-X			Axis Y-Y		
				Width (in.)	Thickness (in.)		<i>I</i> (in. ⁴)	<i>S</i> (in. ³)	<i>r</i> (in.)	<i>I</i> (in. ⁴)	<i>S</i> (in. ³)	<i>r</i> (in.)
14 × 14½	202	59.39	15.63	15.750	1.503	0.930	2538.8	324.9	6.54	979.7	124.4	4.06
	193	56.73	15.50	15.710	1.438	0.890	2402.4	310.0	6.51	930.1	118.4	4.05
	184	54.07	15.38	15.660	1.378	0.840	2274.8	295.8	6.49	882.7	112.7	4.04
	176	51.73	15.25	15.640	1.313	0.820	2149.6	281.9	6.45	837.9	107.1	4.02
	167	49.09	15.12	15.600	1.248	0.780	2020.8	267.3	6.42	790.2	101.3	4.01
	158	46.47	15.00	15.550	1.188	0.730	1900.6	253.4	6.40	745.0	95.8	4.00
	150	44.08	14.88	15.515	1.128	0.695	1786.9	240.2	6.37	702.5	90.6	3.99
	142	41.85	14.75	15.500	1.063	0.680	1672.2	226.7	6.32	660.1	85.2	3.97
	320 ^a	94.12	16.81	16.710	2.093	1.890	4141.7	492.8	6.63	1635.1	195.7	4.17
	136	39.98	14.75	14.740	1.063	0.660	1593.0	216.0	6.31	567.7	77.0	3.77
	127	37.33	14.62	14.690	0.998	0.610	1476.7	202.0	6.29	527.6	71.8	3.76
	119	34.99	14.50	14.650	0.938	0.570	1373.1	189.4	6.26	491.8	67.1	3.75
	111	32.65	14.37	14.620	0.873	0.540	1266.5	176.3	6.23	454.9	62.2	3.73
	103	30.26	14.25	14.575	0.813	0.495	1165.8	163.6	6.21	419.7	57.6	3.72
	95	27.94	14.12	14.545	0.748	0.465	1063.5	150.6	6.17	383.7	52.8	3.71
14 × 12	87	25.56	14.00	14.500	0.688	0.420	966.9	138.1	6.15	349.7	48.2	3.70
	84	24.71	14.18	12.023	0.778	0.451	928.4	130.9	6.13	225.5	37.5	3.02
14 × 10	78	22.94	14.06	12.000	0.718	0.428	851.2	121.1	6.09	206.9	34.5	3.00
	74	21.76	14.19	10.072	0.783	0.450	796.8	112.3	6.05	133.5	26.5	2.48
14 × 8	68	20.00	14.06	10.040	0.718	0.418	724.1	103.0	6.02	121.2	24.1	2.46
	61	17.94	13.91	10.000	0.643	0.378	321.5	92.2	5.98	107.3	21.5	2.45
	53	15.59	13.94	8.062	0.658	0.370	542.1	77.8	5.90	57.5	14.3	1.92
14 × 6¾	48	14.11	13.81	8.031	0.593	0.339	484.9	70.2	5.86	51.3	12.8	1.91
	43	12.65	13.68	8.000	0.528	0.308	429.0	62.7	5.82	45.1	11.3	1.89
	38	11.17	14.12	6.776	0.513	0.313	385.3	54.6	5.87	24.6	7.3	1.49
	34	10.00	14.00	6.750	0.453	0.287	339.2	48.5	5.83	21.3	6.3	1.46
12 × 12	30	8.81	13.86	6.733	0.383	0.270	289.6	41.8	5.73	17.5	5.2	1.41
	190	55.86	14.38	12.670	1.736	1.060	1892.5	263.2	5.82	589.7	93.1	3.25
	161	47.38	13.88	12.515	1.486	0.905	1541.8	222.2	5.70	486.2	77.7	3.20
	133	39.11	13.38	12.365	1.236	0.755	1221.2	182.5	5.59	389.9	63.1	3.16
	120	35.31	13.12	12.320	1.106	0.710	1071.7	163.4	5.51	345.1	56.0	3.13
	106	31.19	12.88	12.230	0.986	0.620	930.7	144.5	5.46	300.9	49.2	3.11
	99	29.09	12.75	12.190	0.921	0.580	858.5	134.7	5.43	278.2	45.7	3.09
	92	27.06	12.62	12.155	0.856	0.545	788.9	125.0	5.40	256.4	42.2	3.08
	85	24.98	12.50	12.105	0.796	0.495	723.3	115.7	5.38	235.5	38.9	3.07
	79	23.22	12.38	12.080	0.736	0.470	663.0	107.1	5.34	216.4	35.8	3.05
12 × 10	72	21.16	12.25	12.040	0.671	0.430	597.4	97.5	5.31	195.3	32.4	3.04
	65	19.11	12.12	12.000	0.606	0.390	533.4	88.0	5.28	174.6	29.1	3.02
	58	17.06	12.19	10.014	0.641	0.359	476.1	78.1	5.28	107.4	21.4	2.51
	53	15.59	12.06	10.000	0.576	0.345	426.2	70.7	5.23	96.1	19.2	2.48
	50	14.71	12.19	8.077	0.641	0.371	394.5	64.7	5.18	56.4	14.0	1.96
12 × 8	45	13.24	12.06	8.042	0.576	0.336	350.8	58.2	5.15	50.0	12.4	1.94
	40	11.77	11.94	8.000	0.516	0.294	310.1	51.9	5.13	44.1	11.0	1.94

TABLE 51.76 (Continued)

Nominal Size (in.)	Weight per Foot (lb)	Area (in. ²)	Depth (in.)	Flange		Web Thickness (in.)	Axis X-X			Axis Y-Y		
				Width (in.)	Thickness (in.)		<i>I</i> (in. ⁴)	<i>S</i> (in. ³)	<i>r</i> (in.)	<i>I</i> (in. ⁴)	<i>S</i> (in. ³)	<i>r</i> (in.)
12 × 6½	36	10.59	12.24	6.565	0.540	0.305	280.8	45.9	5.15	23.7	7.2	1.50
	31	9.12	12.09	6.525	0.465	0.265	238.4	39.4	5.11	19.8	6.1	1.47
	27	7.97	11.95	6.500	0.400	0.240	204.1	34.1	5.06	16.6	5.1	1.44
10 × 10	112	32.92	11.38	10.415	1.248	0.755	718.7	126.3	4.67	235.4	45.2	2.67
	100	29.43	11.12	10.345	1.118	0.685	625.0	112.4	4.61	206.6	39.9	2.65
	89	26.19	10.88	10.275	0.998	0.615	542.4	99.7	4.55	180.6	35.2	2.63
	77	22.67	10.62	10.195	0.868	0.535	457.2	86.1	4.49	153.4	30.1	2.60
	72	21.18	10.50	10.170	0.808	0.510	420.7	80.1	4.46	141.8	27.9	2.59
	66	19.41	10.38	10.117	0.748	0.457	382.5	73.7	4.44	129.2	25.5	2.58
	60	17.66	10.25	10.075	0.683	0.415	343.7	67.1	4.41	116.5	23.1	2.57
	54	15.88	10.12	10.028	0.618	0.368	305.7	60.4	4.39	103.9	20.7	2.56
	49	14.40	10.00	10.000	0.558	0.340	272.9	54.6	4.35	93.0	18.6	2.54
	45	13.24	10.12	8.022	0.618	0.350	248.6	49.1	4.33	53.2	13.3	2.00
10 × 8	39	11.48	9.94	7.990	0.528	0.318	209.7	42.2	4.27	44.9	11.2	1.98
	33	9.71	9.75	7.964	0.433	0.292	170.9	35.0	4.20	36.5	9.2	1.94
	29	8.53	10.22	5.799	0.500	0.289	157.3	30.8	4.29	15.2	5.2	1.34
10 × 5¾	25	7.35	10.08	5.762	0.430	0.252	133.2	26.4	4.26	12.7	4.4	1.31
	21	6.19	9.90	5.750	0.340	0.240	106.3	21.5	4.14	9.7	3.4	1.25
	17	5.00	8.00	5.250	0.308	0.230	56.4	14.1	3.36	6.7	2.6	1.16
8 × 8	67	19.70	9.00	8.287	0.933	0.575	271.8	60.4	3.71	88.6	21.4	2.12
	58	17.06	8.75	8.222	0.808	0.510	227.3	52.0	3.65	74.9	18.2	2.10
	48	14.11	8.50	8.117	0.683	0.405	183.7	43.2	3.61	60.9	15.0	2.08
	40	11.76	8.25	8.077	0.558	0.365	146.3	35.5	3.53	49.0	12.1	2.04
	35	10.30	8.12	8.027	0.493	0.315	126.5	31.1	3.50	42.5	10.6	2.03
	31	9.12	8.00	8.000	0.433	0.288	109.7	27.4	3.47	37.0	9.2	2.01
8 × 6½	28	8.23	8.06	6.540	0.463	0.285	97.8	24.3	3.45	21.6	6.6	1.62
	24	7.06	7.93	6.500	0.398	0.245	82.5	20.8	3.42	18.2	5.6	1.61
8 × 5¼	20	5.88	8.14	5.268	0.378	0.248	69.2	17.0	3.43	8.5	3.2	1.20
	17	5.00	8.00	5.250	0.308	0.230	56.4	14.1	3.36	6.7	2.6	1.16

^aColumn core section.

Most of the sections can be supplied promptly steel mills. Owing to variations in the rolling practice of the different mills, their products are not identical, although their divergence from the values given in the tables is practically negligible. For standardization, only the lesser values are given, and therefore they are on the side of safety.

Further information on sections listed in the tables, together with information on other products and on the requirements for placing orders, may be gathered from mill catalogs.

TABLE 51.77 Properties of American Standard Beams



<div><div></div><div></div></div>												
Nominal Size (in.)	Weight per Foot (lb)	Area (in. ²)	Depth (in.)	Flange		Web Thickness (in.)	Axis X-X			Axis Y-Y		
				Width (in.)	Thickness (in.)		<i>I</i> (in. ⁴)	<i>S</i> (in. ³)	<i>r</i> (in.)	<i>I</i> (in. ⁴)	<i>S</i> (in. ³)	<i>r</i> (in.)
24 × 7 ⁷ / ₈	120.0	35.13	24.00	8.048	1.102	0.798	3010.8	250.9	9.26	84.9	21.1	1.56
	105.9	30.98	24.00	7.875	1.102	0.625	2811.5	234.3	9.53	78.9	20.0	1.60
24 × 7	100.0	29.25	24.00	7.247	0.871	0.747	2371.8	197.6	9.05	48.4	13.4	1.29
	90.0	26.30	24.00	7.124	0.871	0.624	2230.1	185.8	9.21	45.5	12.8	1.32
	79.9	23.33	24.00	7.000	0.871	0.500	2087.2	173.9	9.46	42.9	12.2	1.36
20 × 7	95.0	27.74	20.00	7.200	0.916	0.800	1599.7	160.0	7.59	50.5	14.0	1.35
	85.0	24.80	20.00	7.053	0.916	0.653	1501.7	150.2	7.78	47.0	13.3	1.38
20 × 6 ¹ / ₄	75.0	21.90	20.00	6.391	0.789	0.641	1263.5	126.3	7.60	30.1	9.4	1.17
	65.4	19.08	20.00	6.250	0.789	0.500	1169.5	116.9	7.83	27.9	8.9	1.21
18 × 6	70.0	20.46	18.00	6.251	0.691	0.711	917.5	101.9	6.70	24.5	7.8	1.09
	54.7	15.94	18.00	6.000	0.691	0.460	795.5	88.4	7.07	21.2	7.1	1.15
15 × 5 ¹ / ₂	50.0	14.59	15.00	5.640	0.622	0.550	481.1	64.2	5.74	16.0	5.7	1.05
	42.9	12.49	15.00	5.500	0.622	0.410	441.8	58.9	5.95	14.6	5.3	1.08
12 × 5 ¹ / ₄	50.0	14.57	12.00	5.477	0.659	0.687	301.6	50.3	4.55	16.0	5.8	1.05
	40.8	11.84	12.00	5.250	0.659	0.460	268.9	44.8	4.77	13.8	5.3	1.08
12 × 5	35.0	10.20	12.00	5.078	0.544	0.428	227.0	37.8	4.72	10.0	3.9	0.99
	31.8	9.26	12.00	5.000	0.544	0.350	215.8	36.0	4.83	9.5	3.8	1.01
10 × 4 ⁵ / ₈	35.0	10.22	10.0	4.944	0.491	0.594	145.8	29.2	3.78	8.5	3.4	0.91
	25.4	7.38	10.00	4.660	0.491	0.310	122.1	24.4	4.07	6.9	3.0	0.97
8 × 4	23.0	6.71	8.00	4.171	0.425	0.441	64.2	16.0	3.09	4.4	2.1	0.81
	18.4	5.34	8.00	4.000	0.425	0.270	56.9	14.2	3.26	3.8	1.9	0.84
7 × 3 ⁵ / ₈	20.0	5.83	7.00	3.860	0.392	0.450	41.9	12.0	2.68	3.1	1.6	0.74
	15.3	4.43	7.00	3.660	0.392	0.250	36.2	10.4	2.86	2.7	1.5	0.78
6 × 3 ³ / ₈	17.25	5.02	6.00	3.565	0.359	0.465	26.0	8.7	2.28	2.3	1.3	0.68
	12.5	3.61	6.00	3.330	0.359	0.230	21.8	7.3	2.46	1.8	1.1	0.72
5 × 3	14.75	4.29	5.00	3.284	0.326	0.494	15.0	6.0	1.87	1.7	1.0	0.63
	10.0	2.87	5.00	3.000	0.326	0.210	12.1	4.8	2.05	1.2	0.82	0.65
4 × 2 ⁵ / ₈	9.5	2.76	4.00	2.796	0.293	0.326	6.7	3.3	1.56	0.91	0.65	0.58
	7.7	2.21	4.00	2.660	0.293	0.190	6.0	3.0	1.64	0.77	0.58	0.59
3 × 2 ³ / ₈	7.5	2.17	3.00	2.509	0.260	0.349	2.9	1.9	1.15	0.59	0.47	0.52
	5.7	1.64	3.00	2.330	0.260	0.170	2.5	1.7	1.23	0.46	0.40	0.53

TABLE 51.78 Properties of American Standard Channels

<div><div></div><div></div><div></div></div>													
Nominal Size (in.)	Weight per Foot (lb)	Area (in. ²)	Depth (in.)	Flange		Web Thickness (in.)	Axis X-X			Axis Y-Y			
				Width (in.)	Thickness (in.)		<i>I</i> (in. ⁴)	<i>S</i> (in. ³)	<i>r</i> (in.)	<i>I</i> (in. ⁴)	<i>S</i> (in. ³)	<i>x</i> (in.)	<i>r</i> (in.)
18 × 4 ^a	58.0	16.98	18.00	4.200	0.625	0.700	670.7	74.5	6.29	18.5	5.6	1.04	0.88
	51.9	15.18	18.00	4.100	0.625	0.600	622.1	69.1	6.40	17.1	5.3	1.06	0.87
	45.8	13.38	18.00	4.000	0.625	0.500	573.5	63.7	6.55	15.8	5.1	1.09	0.89
	42.7	12.48	18.00	3.950	0.625	0.450	549.2	61.0	6.64	15.0	4.9	1.10	0.90
15 × 3 ³ / ₈	50.0	14.64	15.00	3.716	0.650	0.716	401.4	53.6	5.24	11.2	3.8	0.87	0.80
	40.0	11.70	15.00	3.520	0.650	0.520	346.3	46.2	5.44	9.3	3.4	0.89	0.78
	33.9	9.90	15.00	3.400	0.650	0.400	312.6	41.7	5.62	8.2	3.2	0.91	0.79
12 × 3	30.0	8.79	12.00	3.170	0.501	0.510	161.2	26.9	4.28	5.2	2.1	0.77	0.68
	25.0	7.32	12.00	3.047	0.501	0.387	143.5	23.9	4.43	4.5	1.9	0.79	0.68
	20.7	6.03	12.00	2.940	0.501	0.280	128.1	21.4	4.61	3.9	1.7	0.81	0.70
10 × 2 ⁵ / ₈	30.0	8.80	10.00	3.033	0.436	0.673	103.0	20.6	3.42	4.0	1.7	0.67	0.65
	25.0	7.33	10.00	2.886	0.436	0.526	90.7	18.1	3.52	3.4	1.5	0.68	0.62
	20.0	5.86	10.00	2.739	0.436	0.379	78.5	15.7	3.66	2.8	1.3	0.70	0.61
	15.3	4.47	10.00	2.600	0.436	0.240	66.9	13.4	3.87	2.3	1.2	0.72	0.64
9 × 2 ¹ / ₂	20.0	5.86	9.00	2.648	0.413	0.448	60.6	13.5	3.22	2.4	1.2	0.65	0.59
	15.0	4.39	9.00	2.485	0.413	0.285	50.7	11.3	3.40	1.9	1.0	0.67	0.59
	13.4	3.89	9.00	2.430	0.413	0.230	47.3	10.5	3.49	1.8	0.97	0.67	0.61
8 × 2 ¹ / ₄	18.75	5.49	8.00	2.527	0.390	0.487	43.7	10.9	2.82	2.0	1.0	0.60	0.57
	13.75	4.02	8.00	2.343	0.390	0.303	35.8	9.0	2.99	1.5	0.86	0.62	0.56
	11.5	3.36	8.00	2.260	0.390	0.220	32.3	8.1	3.10	1.3	0.79	0.63	0.58
7 × 2 ¹ / ₈	14.75	4.32	7.00	2.299	0.366	0.419	27.1	7.7	2.51	1.4	0.79	0.57	0.53
	12.25	3.58	7.00	2.194	0.366	0.314	24.1	6.9	2.59	1.2	0.71	0.58	0.53
	9.8	2.85	7.00	2.090	0.366	0.210	21.1	6.0	2.72	0.98	0.63	0.59	0.55
6 × 2	13.0	3.81	6.00	2.157	0.343	0.437	17.3	5.8	2.13	1.1	0.65	0.53	0.52
	10.5	3.07	6.00	2.034	0.343	0.314	15.1	5.0	2.22	0.87	0.57	0.53	0.50
	8.2	2.39	6.00	1.920	0.343	0.200	13.0	4.3	2.34	0.70	0.50	0.54	0.52
5 × 1 ³ / ₄	9.0	2.63	5.00	1.885	0.320	0.325	8.8	3.5	1.83	0.64	0.45	0.49	0.48
	6.7	1.95	5.00	1.750	0.320	0.190	7.4	3.0	1.95	0.48	0.38	0.50	0.49
4 × 1 ⁵ / ₈	7.25	2.12	4.00	1.720	0.296	0.320	4.5	2.3	1.47	0.44	0.35	0.46	0.46
	5.4	1.56	4.00	1.580	0.296	0.180	3.8	1.9	1.56	0.32	0.29	0.45	0.46
3 × 1 ¹ / ₂	6.0	1.75	3.00	1.596	0.273	0.356	2.1	1.4	1.08	0.31	0.27	0.42	0.46
	5.0	1.46	3.00	1.498	0.273	0.258	1.8	1.2	1.12	0.25	0.24	0.41	0.44
	4.1	1.19	3.00	1.410	0.273	0.170	1.6	1.1	1.17	0.20	0.21	0.41	0.44

^aCar and Shipbuilding Channel; not an American standard.

TABLE 51.79 Properties of Angles with Equal Legs


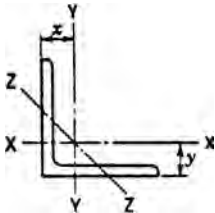
<div><div></div><div></div></div>								
Size (in.)	Thickness (in.)	Weight per Foot (lb)	Area (in. ²)	Axis X-X and Axis Y-Y				Axis Z-Z
				<i>I</i> (in. ⁴)	<i>S</i> (in. ³)	<i>r</i> (in.)	<i>x</i> or <i>y</i> (in.)	<i>r</i> (in.)
8 × 8	1 ¹ / ₈	56.9	16.73	98.0	17.5	2.42	2.41	1.56
	1	51.0	15.00	89.0	15.8	2.44	2.37	1.56
	⁷ / ₈	45.0	13.23	79.6	14.0	2.45	2.32	1.57
	³ / ₄	38.9	11.44	69.7	12.2	2.47	2.28	1.57
	⁵ / ₈	32.7	9.61	59.4	10.3	2.49	2.23	1.58
	⁹ / ₁₆	29.6	8.68	54.1	9.3	2.50	2.21	1.58
	¹ / ₂	26.4	7.75	48.6	8.4	2.50	2.19	1.59
6 × 6	1	37.4	11.00	35.5	8.6	1.80	1.86	1.17
	⁷ / ₈	33.1	9.73	31.9	7.6	1.81	1.82	1.17
	³ / ₄	28.7	8.44	28.2	6.7	1.83	1.78	1.17
	⁵ / ₈	24.2	7.11	24.2	5.7	1.84	1.73	1.18
	⁹ / ₁₆	21.9	6.43	22.1	5.1	1.85	1.71	1.18
	¹ / ₂	19.6	5.75	19.9	4.6	1.86	1.68	1.18
	⁷ / ₁₆	17.2	5.06	17.7	4.1	1.87	1.66	1.19
5 × 5	³ / ₈	14.9	4.36	15.4	3.5	1.88	1.64	1.19
	⁵ / ₁₆	12.5	3.66	13.0	3.0	1.89	1.61	1.19
	⁷ / ₈	27.2	7.98	17.8	5.2	1.49	1.57	0.97
	³ / ₄	23.6	6.94	15.7	4.5	1.51	1.52	0.97
	⁵ / ₈	20.0	5.86	13.6	3.9	1.52	1.48	0.98
	¹ / ₂	16.2	4.75	11.3	3.2	1.54	1.43	0.98
	⁷ / ₁₆	14.3	4.18	10.0	2.8	1.55	1.41	0.98
4 × 4	³ / ₈	12.3	3.61	8.7	2.4	1.56	1.39	0.99
	⁵ / ₁₆	10.3	3.03	7.4	2.0	1.57	1.37	0.99
	³ / ₄	18.5	5.44	7.7	2.8	1.19	1.27	0.78
	⁵ / ₈	15.7	4.61	6.7	2.4	1.20	1.23	0.78
	¹ / ₂	12.8	3.75	5.6	2.0	1.22	1.18	0.78
	⁷ / ₁₆	11.3	3.31	5.0	1.8	1.23	1.16	0.78
	³ / ₈	9.8	2.86	4.4	1.5	1.23	1.14	0.79

TABLE 51.79 (Continued)

Size (in.)	Thickness (in.)	Weight per Foot (lb)	Area (in. ²)	Axis X-X and Axis Y-Y				Axis Z-Z
				<i>I</i> (in. ⁴)	<i>S</i> (in. ³)	<i>r</i> (in.)	<i>x</i> or <i>y</i> (in.)	<i>r</i> (in.)
$3\frac{1}{2} \times 3\frac{1}{2}$	$\frac{5}{16}$	8.2	2.40	3.7	1.3	1.24	1.12	0.79
	$\frac{1}{4}$	6.6	1.94	3.0	1.1	1.25	1.09	0.80
	$\frac{1}{2}$	11.1	3.25	3.6	1.5	1.06	1.06	0.68
	$\frac{7}{16}$	9.8	2.87	3.3	1.3	1.07	1.04	0.68
	$\frac{3}{8}$	8.5	2.48	2.9	1.2	1.07	1.01	0.69
3×3	$\frac{5}{16}$	7.2	2.09	2.5	0.98	1.08	0.99	0.69
	$\frac{1}{4}$	5.8	1.69	2.0	0.79	1.09	0.97	0.69
	$\frac{1}{2}$	9.4	2.75	2.2	1.1	0.90	0.93	0.58
	$\frac{7}{16}$	8.3	2.43	2.0	0.95	0.91	0.91	0.58
	$\frac{3}{8}$	7.2	2.11	1.8	0.83	0.91	0.89	0.58
$2\frac{1}{2} \times 2\frac{1}{2}$	$\frac{5}{16}$	6.1	1.78	1.5	0.71	0.92	0.87	0.59
	$\frac{1}{4}$	4.9	1.44	1.2	0.58	0.93	0.84	0.59
	$\frac{3}{16}$	3.71	1.09	0.96	0.44	0.94	0.82	0.59
	$\frac{1}{2}$	7.7	2.25	1.2	0.72	0.74	0.81	0.49
	$\frac{3}{8}$	5.9	1.73	0.98	0.57	0.75	0.76	0.49
2×2	$\frac{5}{16}$	5.0	1.47	0.85	0.48	0.76	0.74	0.49
	$\frac{1}{4}$	4.1	1.19	0.70	0.39	0.77	0.72	0.49
	$\frac{3}{16}$	3.07	0.90	0.55	0.30	0.78	0.69	0.49
	$\frac{3}{8}$	4.7	1.36	0.48	0.35	0.59	0.64	0.39
	$\frac{5}{16}$	3.92	1.15	0.42	0.30	0.60	0.61	0.39
$1\frac{3}{4} \times 1\frac{3}{4}$	$\frac{1}{4}$	3.19	0.94	0.35	0.25	0.61	0.59	0.39
	$\frac{3}{16}$	2.44	0.71	0.27	0.19	0.62	0.57	0.39
	$\frac{1}{8}$	1.65	0.48	0.19	0.13	0.63	0.55	0.40
	$\frac{1}{4}$	2.77	0.81	0.23	0.19	0.53	0.53	0.34
	$\frac{3}{16}$	2.12	0.62	0.18	0.14	0.54	0.51	0.34
$1\frac{1}{2} \times 1\frac{1}{2}$	$\frac{1}{8}$	1.44	0.42	0.13	0.10	0.55	0.48	0.35
	$\frac{1}{4}$	2.34	0.69	0.14	0.13	0.45	0.47	0.29
	$\frac{3}{16}$	1.80	0.53	0.11	0.10	0.46	0.44	0.29
	$\frac{1}{8}$	1.23	0.36	0.08	0.07	0.47	0.42	0.30
	$\frac{1}{4}$	1.92	0.56	0.08	0.09	0.37	0.40	0.24
$1\frac{1}{4} \times 1\frac{1}{4}$	$\frac{3}{16}$	1.48	0.43	0.06	0.07	0.38	0.38	0.24
	$\frac{1}{8}$	1.01	0.30	0.04	0.05	0.38	0.36	0.25
	$\frac{1}{4}$	1.49	0.44	0.04	0.06	0.29	0.34	0.20
	$\frac{3}{16}$	1.16	0.34	0.03	0.04	0.30	0.32	0.19
	$\frac{1}{8}$	0.80	0.23	0.02	0.03	0.30	0.30	0.20

TABLE 51.80 Properties of Angles with Unequal Legs

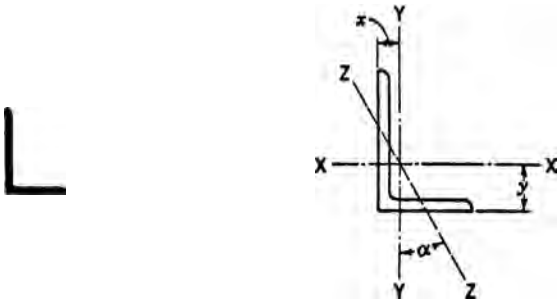
<div><div></div></div>													
Size (in.)	Thickness (in.)	Weight per Foot (lb)	Area (in. ²)	Axis X-X				Axis Y-Y				Axis Z-Z	
				<i>I</i> (in. ⁴)	<i>S</i> (in. ³)	<i>r</i> (in.)	<i>y</i> (in.)	<i>I</i> (in. ⁴)	<i>S</i> (in. ³)	<i>r</i> (in.)	<i>x</i> (in.)	<i>r</i> (in.)	tanα
9 × 4	1	40.8	12.00	97.0	17.6	2.84	3.50	12.0	4.0	1.00	1.00	0.83	0.203
	$\frac{7}{8}$	36.1	10.61	86.8	15.7	2.86	3.45	10.8	3.6	1.01	0.95	0.84	0.208
	$\frac{3}{4}$	31.3	9.19	76.1	13.6	2.88	3.41	9.6	3.1	1.02	0.91	0.84	0.212
	$\frac{5}{8}$	26.3	7.73	64.9	11.5	2.90	3.36	8.3	2.6	1.04	0.86	0.85	0.216
	$\frac{9}{16}$	23.8	7.00	59.1	10.4	2.91	3.33	7.6	2.4	1.04	0.83	0.85	0.218
	$\frac{1}{2}$	21.3	6.25	53.2	9.3	2.92	3.31	6.9	2.2	1.05	0.81	0.85	0.220
8 × 6	1	44.2	13.00	80.8	15.1	2.49	2.65	38.8	8.9	1.73	1.65	1.28	0.543
	$\frac{7}{8}$	39.1	11.48	72.3	13.4	2.51	2.61	34.9	7.9	1.74	1.61	1.28	0.547
	$\frac{3}{4}$	33.8	9.94	63.4	11.7	2.53	2.56	30.7	6.9	1.76	1.56	1.29	0.551
	$\frac{5}{8}$	28.5	8.36	54.1	9.9	2.54	2.52	26.3	5.9	1.77	1.52	1.29	0.554
	$\frac{9}{16}$	25.7	7.56	49.3	9.0	2.55	2.50	24.0	5.3	1.78	1.50	1.30	0.556
	$\frac{1}{2}$	23.0	6.75	44.3	8.0	2.56	2.47	21.7	4.8	1.79	1.47	1.30	0.558
8 × 4	$\frac{7}{16}$	20.2	5.93	39.2	7.1	2.57	2.45	19.3	4.2	1.80	1.45	1.31	0.560
	1	37.4	11.00	69.6	14.1	2.52	3.05	11.6	3.9	1.03	1.05	0.85	0.247
	$\frac{7}{8}$	33.1	9.73	62.5	12.5	2.53	3.00	10.5	3.5	1.04	1.00	0.85	0.253
	$\frac{3}{4}$	28.7	8.44	54.9	10.9	2.55	2.95	9.4	3.1	1.05	0.95	0.85	0.258
	$\frac{5}{8}$	24.2	7.11	46.9	9.2	2.57	2.91	8.1	2.6	1.07	0.91	0.86	0.262
	$\frac{9}{16}$	21.9	6.43	42.8	8.4	2.58	2.88	7.4	2.4	1.07	0.88	0.86	0.265
7 × 4	$\frac{1}{2}$	19.6	5.75	38.5	7.5	2.59	2.86	6.7	2.2	1.08	0.86	0.86	0.267
	$\frac{7}{16}$	17.2	5.06	34.1	6.6	2.60	2.83	6.0	1.9	1.09	0.83	0.87	0.269
	$\frac{7}{8}$	30.2	8.86	42.9	9.7	2.20	2.55	10.2	3.5	1.07	1.05	0.86	0.318
	$\frac{3}{4}$	26.2	7.69	37.8	8.4	2.22	2.51	9.1	3.0	1.09	1.01	0.86	0.324
	$\frac{5}{8}$	22.1	6.48	32.4	7.1	2.24	2.46	7.8	2.6	1.10	0.96	0.86	0.329
	$\frac{9}{16}$	20.0	5.87	29.6	6.5	2.24	2.44	7.2	2.4	1.11	0.94	0.87	0.332
6 × 4	$\frac{1}{2}$	17.9	5.25	26.7	5.8	2.25	2.42	6.5	2.1	1.11	0.92	0.87	0.335
	$\frac{7}{16}$	15.8	4.62	23.7	5.1	2.26	2.39	5.8	1.9	1.12	0.89	0.88	0.337
	$\frac{3}{8}$	13.6	3.98	20.6	4.4	2.27	2.37	5.1	1.6	1.13	0.87	0.88	0.339
	$\frac{7}{8}$	27.2	7.98	27.7	7.2	1.86	2.12	9.8	3.4	1.11	1.12	0.86	0.421
	$\frac{3}{4}$	23.6	6.94	24.5	6.3	1.88	2.08	8.7	3.0	1.12	1.08	0.86	0.428

TABLE 51.80 (Continued)

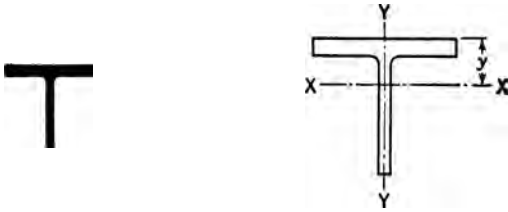
Size (in.)	Thickness (in.)	Weight per Foot (lb)	Area (in. ²)	Axis X-X				Axis Y-Y				Axis Z-Z	
				<i>I</i> (in. ⁴)	<i>S</i> (in. ³)	<i>r</i> (in.)	<i>y</i> (in.)	<i>I</i> (in. ⁴)	<i>S</i> (in. ³)	<i>r</i> (in.)	<i>x</i> (in.)	<i>r</i> (in.)	tanα
6 × 3½	$\frac{5}{8}$	20.0	5.86	21.1	5.3	1.90	2.03	7.5	2.5	1.13	1.03	0.86	0.435
	$\frac{9}{16}$	18.1	5.31	19.3	4.8	1.90	2.01	6.9	2.3	1.14	1.01	0.87	0.438
	$\frac{1}{2}$	16.2	4.75	17.4	4.3	1.91	1.99	6.3	2.1	1.15	0.99	0.87	0.440
	$\frac{7}{16}$	14.3	4.18	15.5	3.8	1.92	1.96	5.6	1.9	1.16	0.96	0.87	0.443
	$\frac{3}{8}$	12.3	3.61	13.5	3.3	1.93	1.94	4.9	1.6	1.17	0.94	0.88	0.446
	$\frac{5}{16}$	10.3	3.03	11.4	2.8	1.94	1.92	4.2	1.4	1.17	0.92	0.88	0.449
	$\frac{1}{2}$	15.3	4.50	16.6	4.2	1.92	2.08	4.3	1.6	0.97	0.83	0.76	0.344
	$\frac{3}{8}$	11.7	3.42	12.9	3.2	1.94	2.04	3.3	1.2	0.99	0.79	0.77	0.350
	$\frac{5}{16}$	9.8	2.87	10.9	2.7	1.95	2.01	2.9	1.0	1.00	0.76	0.77	0.352
	$\frac{1}{4}$	7.9	2.31	8.9	2.2	1.96	1.99	2.3	0.85	1.01	0.74	0.78	0.355
5 × 3½	$\frac{3}{4}$	19.8	5.81	13.9	4.3	1.55	1.75	5.6	2.2	0.98	1.00	0.75	0.464
	$\frac{5}{8}$	16.8	4.92	12.0	3.7	1.56	1.70	4.8	1.9	0.99	0.95	0.75	0.472
	$\frac{1}{2}$	13.6	4.00	10.0	3.0	1.58	1.66	4.1	1.6	1.01	0.91	0.75	0.479
	$\frac{7}{16}$	12.0	3.53	8.9	2.6	1.59	1.63	3.6	1.4	1.01	0.88	0.76	0.482
	$\frac{3}{8}$	10.4	3.05	7.8	2.3	1.60	1.61	3.2	1.2	1.02	0.86	0.76	0.486
	$\frac{5}{16}$	8.7	2.56	6.6	1.9	1.61	1.59	2.7	1.0	1.03	0.84	0.76	0.489
	$\frac{1}{4}$	7.0	2.06	5.4	1.6	1.61	1.56	2.2	0.83	1.04	0.81	0.76	0.492
	$\frac{1}{2}$	12.8	3.75	9.5	2.9	1.59	1.75	2.6	1.1	0.83	0.75	0.65	0.357
	$\frac{7}{16}$	11.3	3.31	8.4	2.6	1.60	1.73	2.3	1.0	0.84	0.73	0.65	0.361
	$\frac{3}{8}$	9.8	2.86	7.4	2.2	1.61	1.70	2.0	0.89	0.84	0.70	0.65	0.364
5 × 3	$\frac{5}{16}$	8.2	2.40	6.3	1.9	1.61	1.68	1.8	0.75	0.85	0.68	0.66	0.368
	$\frac{1}{4}$	6.6	1.94	5.1	1.5	1.62	1.66	1.4	0.61	0.86	0.66	0.66	0.371
	$\frac{5}{8}$	14.7	4.30	6.4	2.4	1.22	1.29	4.5	1.8	1.03	1.04	0.72	0.745
	$\frac{1}{2}$	11.9	3.50	5.3	1.9	1.23	1.25	3.8	1.5	1.04	1.00	0.72	0.750
	$\frac{7}{16}$	10.6	3.09	4.8	1.7	1.24	1.23	3.4	1.4	1.05	0.98	0.72	0.753
	$\frac{3}{8}$	9.1	2.67	4.2	1.5	1.25	1.21	3.0	1.2	1.06	0.96	0.73	0.755
	$\frac{5}{16}$	7.7	2.25	3.6	1.3	1.26	1.18	2.6	1.0	1.07	0.93	0.73	0.757
	$\frac{1}{4}$	6.2	1.81	2.9	1.0	1.27	1.16	2.1	0.81	1.07	0.91	0.73	0.759
	$\frac{5}{8}$	13.6	3.98	6.0	2.3	1.23	1.37	2.9	1.4	0.85	0.87	0.64	0.534
	$\frac{1}{2}$	11.1	3.25	5.1	1.9	1.25	1.33	2.4	1.1	0.86	0.83	0.64	0.543
4 × 3	$\frac{7}{16}$	9.8	2.87	4.5	1.7	1.25	1.30	2.2	1.0	0.87	0.80	0.64	0.547
	$\frac{3}{8}$	8.5	2.48	4.0	1.5	1.26	1.28	1.9	0.87	0.88	0.78	0.64	0.551
	$\frac{5}{16}$	7.2	2.09	3.4	1.2	1.27	1.26	1.7	0.73	0.89	0.76	0.65	0.554
	$\frac{1}{4}$	5.8	1.69	2.8	1.0	1.28	1.24	1.4	0.60	0.90	0.74	0.65	0.558
	$\frac{1}{2}$	10.2	3.00	3.5	1.5	1.07	1.13	2.3	1.1	0.88	0.88	0.62	0.714
	$\frac{7}{16}$	9.1	2.65	3.1	1.3	1.08	1.10	2.1	0.98	0.89	0.85	0.62	0.718
	$\frac{3}{8}$	7.9	2.30	2.7	1.1	1.09	1.08	1.9	0.85	0.90	0.83	0.62	0.721
	$\frac{5}{16}$	6.6	1.93	2.3	0.95	1.10	1.06	1.6	0.72	0.90	0.81	0.63	0.724
3½ × 3													

(continued)

TABLE 51.80 (Continued)

Size (in.)	Thickness (in.)	Weight per Foot (lb)	Area (in. ²)	Axis X-X				Axis Y-Y				Axis Z-Z	
				<i>I</i> (in. ⁴)	<i>S</i> (in. ³)	<i>r</i> (in.)	<i>y</i> (in.)	<i>I</i> (in. ⁴)	<i>S</i> (in. ³)	<i>r</i> (in.)	<i>x</i> (in.)	<i>r</i> (in.)	tanα
$3\frac{1}{2} \times 2\frac{1}{2}$	$\frac{1}{4}$	5.4	1.56	1.9	0.78	1.11	1.04	1.3	0.59	0.91	0.79	0.63	0.727
	$\frac{1}{2}$	9.4	2.75	3.2	1.4	1.09	1.20	1.4	0.76	0.70	0.70	0.53	0.486
	$\frac{7}{16}$	8.3	2.43	2.9	1.3	1.09	1.18	1.2	0.68	0.71	0.68	0.54	0.491
	$\frac{3}{8}$	7.2	2.11	2.6	1.1	1.10	1.16	1.1	0.59	0.72	0.66	0.54	0.496
	$\frac{5}{16}$	6.1	1.78	2.2	0.93	1.11	1.14	0.94	0.50	0.73	0.64	0.54	0.501
$3 \times 2\frac{1}{2}$	$\frac{1}{4}$	4.9	1.44	1.8	0.75	1.12	1.11	0.78	0.41	0.74	0.61	0.54	0.506
	$\frac{1}{2}$	8.5	2.50	2.1	1.0	0.91	1.00	1.3	0.74	0.72	0.75	0.52	0.667
	$\frac{7}{16}$	7.6	2.21	1.9	0.93	0.92	0.98	1.2	0.66	0.73	0.73	0.52	0.672
	$\frac{3}{8}$	6.6	1.92	1.7	0.81	0.93	0.96	1.0	0.58	0.74	0.71	0.52	0.676
	$\frac{5}{16}$	5.6	1.62	1.4	0.69	0.94	0.93	0.90	0.49	0.74	0.68	0.53	0.680
3×2	$\frac{1}{4}$	4.5	1.31	1.2	0.56	0.95	0.91	0.74	0.40	0.75	0.66	0.53	0.684
	$\frac{1}{2}$	7.7	2.25	1.9	1.0	0.92	1.08	0.67	0.47	0.55	0.58	0.43	0.414
	$\frac{7}{16}$	6.8	2.00	1.7	0.89	0.93	1.06	0.61	0.42	0.55	0.56	0.43	0.421
	$\frac{3}{8}$	5.9	1.73	1.5	0.78	0.94	1.04	0.54	0.37	0.56	0.54	0.43	0.428
	$\frac{5}{16}$	5.0	1.47	1.3	0.66	0.95	1.02	0.47	0.32	0.57	0.52	0.43	0.435
$2\frac{1}{2} \times 2$	$\frac{1}{4}$	4.1	1.19	1.1	0.54	0.95	0.99	0.39	0.26	0.57	0.49	0.43	0.440
	$\frac{3}{16}$	3.07	0.90	0.84	0.41	0.97	0.97	0.31	0.20	0.58	0.47	0.44	0.446
	$\frac{3}{8}$	5.3	1.55	0.91	0.55	0.77	0.83	0.51	0.36	0.58	0.58	0.42	0.614
	$\frac{5}{16}$	4.5	1.31	0.79	0.47	0.78	0.81	0.45	0.31	0.58	0.56	0.42	0.620
	$\frac{1}{4}$	3.62	1.06	0.65	0.38	0.78	0.79	0.37	0.25	0.59	0.54	0.42	0.626
$2\frac{1}{2} \times 1\frac{1}{2}$	$\frac{3}{16}$	2.75	0.81	0.51	0.29	0.79	0.76	0.29	0.20	0.60	0.51	0.43	0.631
	$\frac{3}{8}$	4.7	1.36	0.82	0.52	0.78	0.92	0.22	0.20	0.40	0.42	0.32	0.340
	$\frac{5}{16}$	3.92	1.15	0.71	0.44	0.79	0.90	0.19	0.17	0.41	0.40	0.32	0.349
	$\frac{1}{4}$	3.19	0.94	0.59	0.36	0.79	0.88	0.16	0.14	0.41	0.38	0.32	0.357
	$\frac{3}{16}$	2.44	0.72	0.46	0.28	0.80	0.85	0.13	0.11	0.42	0.35	0.33	0.364
$2 \times 1\frac{1}{2}$	$\frac{1}{4}$	2.77	0.81	0.32	0.24	0.62	0.66	0.15	0.14	0.43	0.41	0.32	0.543
	$\frac{3}{16}$	2.12	0.62	0.25	0.18	0.63	0.64	0.12	0.11	0.44	0.39	0.32	0.551
	$\frac{1}{8}$	1.44	0.42	0.17	0.13	0.64	0.62	0.09	0.08	0.45	0.37	0.33	0.558
$1\frac{3}{4} \times 1\frac{1}{4}$	$\frac{1}{4}$	2.34	0.69	0.20	0.18	0.54	0.60	0.09	0.10	0.35	0.35	0.27	0.486
	$\frac{3}{16}$	1.80	0.53	0.16	0.14	0.55	0.58	0.07	0.08	0.36	0.33	0.27	0.496
	$\frac{1}{8}$	1.23	0.36	0.11	0.09	0.56	0.56	0.05	0.05	0.37	0.31	0.27	0.506

TABLE 51.81 Properties and Dimensions of Tees



Section Number	Weight per Foot (lb)	Area (in. ²)	Depth of Tee (in.)	Flange		Stem Thickness (in.)	Axis X-X				Axis Y-Y		
				Width (in.)	Average Thickness (in.)		<i>I</i> (in. ⁴)	<i>S</i> (in. ³)	<i>r</i> (in.)	<i>y</i> (in.)	<i>I</i> (in. ⁴)	<i>S</i> (in. ³)	<i>r</i> (in.)
ST 18 WF ^a	150	44.09	18.36	16.655	1.680	0.945	1222.7	85.9	5.27	4.13	612.6	73.6	3.73
	140	41.16	18.25	16.595	1.570	0.885	1133.3	79.9	5.25	4.07	563.7	67.9	3.70
	130	38.28	18.12	16.555	1.440	0.845	1059.2	75.4	5.26	4.07	510.3	61.6	3.65
	122.5	36.01	18.03	16.512	1.350	0.802	994.3	71.1	5.25	4.04	472.3	57.2	3.62
	115	33.86	17.94	16.475	1.260	0.765	935.8	67.2	5.26	4.02	435.5	52.9	3.59
ST 18 WF	97	28.56	18.24	12.117	1.260	0.770	904.0	67.3	5.63	4.81	177.7	29.3	2.49
	91	26.77	18.16	12.072	1.180	0.725	844.0	63.0	5.61	4.77	163.9	27.1	2.47
	85	24.99	18.08	12.027	1.100	0.680	784.7	58.8	5.60	4.74	150.3	25.0	2.45
	80	23.54	18.00	12.000	1.020	0.653	741.0	56.0	5.61	4.76	137.7	22.9	2.42
	75	22.08	17.92	11.972	0.940	0.625	696.7	53.0	5.62	4.79	125.2	20.9	2.38
ST 16 WF	120	35.26	16.75	15.865	1.400	0.830	822.5	63.2	4.83	3.73	437.2	55.1	3.52
	110	32.36	16.63	15.810	1.275	0.775	754.1	58.4	4.83	3.71	391.2	49.5	3.48
	100	29.40	16.50	15.750	1.150	0.715	683.6	53.3	4.82	3.67	345.8	43.9	3.43
ST 16 WF	76	22.35	16.75	11.565	1.055	0.635	591.9	47.4	5.15	4.26	128.1	22.1	2.39
	70.5	20.76	16.66	11.535	0.960	0.603	551.8	44.7	5.16	4.30	114.9	19.9	2.35
	65	19.13	16.55	11.510	0.855	0.580	513.0	42.1	5.18	4.37	100.7	17.5	2.29
ST 15 WF	105	30.89	15.19	15.105	1.315	0.775	578.0	48.7	4.33	3.31	354.0	46.9	3.38
	95	27.95	15.06	15.040	1.185	0.710	520.4	44.1	4.31	3.26	312.3	41.5	3.34
	86	25.32	14.94	14.985	1.065	0.655	471.0	40.2	4.31	3.23	275.1	36.7	3.30
ST 15 WF	66	19.41	15.15	10.551	1.000	0.615	420.7	37.4	4.66	3.90	92.5	17.5	2.18
	62	18.22	15.08	10.521	0.930	0.585	394.8	35.3	4.65	3.90	84.8	16.1	2.16
	58.0	17.07	15.00	10.500	0.850	0.564	371.8	33.6	4.67	3.94	76.6	14.6	2.12
	54.0	15.88	14.91	10.484	0.760	0.548	349.5	32.1	4.69	4.03	67.6	12.9	2.06
ST 13 WF	88.5	26.05	13.66	14.090	1.190	0.725	391.8	36.7	3.88	2.97	259.4	36.8	3.16
	80	23.72	13.54	14.023	1.075	0.658	351.4	33.1	3.87	2.91	229.0	32.7	3.12
	72.5	21.34	13.44	13.965	0.975	0.600	316.3	29.9	3.85	2.85	203.5	29.1	3.09
ST 13 WF	57	16.77	13.64	10.070	0.932	0.570	288.9	28.3	4.15	3.42	74.8	14.9	2.11
	51	15.01	13.53	10.018	0.827	0.518	257.7	25.4	4.14	3.39	64.8	12.9	2.08
	47	13.83	13.45	9.990	0.747	0.490	238.5	23.7	4.15	3.41	57.5	11.5	2.04
ST 12 WF	80	23.54	12.36	14.091	1.135	0.656	271.6	27.6	3.40	2.51	246.3	35.0	3.23
	72.5	21.31	12.24	14.043	1.020	0.608	246.2	25.2	3.40	2.48	217.1	30.9	3.19
	65	19.11	12.13	14.000	0.900	0.565	222.6	23.1	3.41	2.47	187.6	26.8	3.13
ST 12 WF	60	17.64	12.16	12.088	0.930	0.556	213.6	22.4	3.48	2.62	127.0	21.0	2.68
	55	16.18	12.08	12.042	0.855	0.510	195.2	20.5	3.47	2.57	114.5	19.0	2.66
	50	14.71	12.00	12.000	0.775	0.468	176.7	18.7	3.46	2.54	101.8	17.0	2.63

(continued)

TABLE 51.81 (Continued)

Section Number	Weight per Foot (lb)	Area (in. ²)	Depth of Tee (in.)	Flange		Stem Thickness (in.)	Axis X-X				Axis Y-Y		
				Width (in.)	Average Thickness (in.)		<i>I</i> (in. ⁴)	<i>S</i> (in. ³)	<i>r</i> (in.)	<i>y</i> (in.)	<i>I</i> (in. ⁴)	<i>S</i> (in. ³)	<i>r</i> (in.)
ST 12 WF	47	13.81	12.15	9.061	0.872	0.516	185.9	20.3	3.67	2.99	51.1	11.3	1.92
	42	12.35	12.04	9.015	0.772	0.470	165.9	18.3	3.66	2.97	44.2	9.8	1.89
	38	11.18	11.95	8.985	0.682	0.440	151.1	16.9	3.68	3.00	38.3	8.5	1.85
ST 10 WF	71	20.88	10.73	13.132	1.095	0.659	177.3	20.8	2.91	2.18	193.0	29.4	3.04
	63.5	18.67	10.62	13.061	0.985	0.588	155.8	18.3	2.89	2.11	169.3	25.9	3.01
	56	16.47	10.50	13.000	0.865	0.527	136.4	16.2	2.88	2.06	144.8	22.3	2.96
ST 10 WF ^a	48	14.11	10.57	9.038	0.935	0.575	137.1	17.1	3.11	2.55	54.7	12.1	1.97
	41	12.05	10.43	8.962	0.795	0.499	115.4	14.5	3.09	2.48	44.8	10.0	1.93
ST 10 WF	36.5	10.73	10.62	8.295	0.740	0.455	110.2	13.7	3.21	2.60	33.1	7.98	1.76
	34	10.01	10.57	8.270	0.685	0.430	102.8	12.9	3.20	2.59	30.2	7.30	1.74
	31	9.12	10.49	8.240	0.615	0.400	93.7	11.9	3.21	2.59	26.6	6.45	1.71
ST 9 WF	57	16.77	9.24	11.833	0.991	0.595	102.6	13.9	2.47	1.85	127.8	21.6	2.76
	52.5	15.43	9.16	11.792	0.911	0.554	93.9	12.8	2.47	1.82	115.5	19.6	2.73
	48	14.11	9.08	11.750	0.831	0.512	85.3	11.7	2.46	1.78	103.4	17.6	2.71
ST 9 WF	42.5	12.49	9.16	8.838	0.911	0.526	84.4	11.9	2.60	2.05	49.7	11.3	2.00
	38.5	11.32	9.08	8.787	0.831	0.475	75.3	10.6	2.58	1.99	44.3	10.1	1.98
	35	10.28	9.00	8.750	0.751	0.438	68.1	9.67	2.57	1.96	39.2	8.97	1.95
	32	9.40	8.94	8.715	0.686	0.403	61.8	8.82	2.56	1.93	35.2	8.07	1.93
ST 9 WF	30	8.82	9.12	7.558	0.695	0.416	64.8	9.32	2.71	2.17	23.5	6.23	1.63
	27.5	8.09	9.06	7.532	0.630	0.390	59.6	8.63	2.71	2.16	21.0	5.57	1.61
	25	7.35	9.00	7.500	0.570	0.358	53.9	7.85	2.71	2.14	18.6	4.96	1.59
ST 8 WF	48	14.11	8.16	11.533	0.875	0.535	64.7	9.82	2.14	1.57	103.6	18.0	2.71
	44	12.94	8.08	11.502	0.795	0.504	59.5	9.11	2.14	1.55	92.6	16.1	2.67
ST 8 WF	39	11.46	8.16	8.586	0.875	0.529	60.0	9.45	2.28	1.81	43.8	10.2	1.95
	35.5	10.43	8.08	8.543	0.795	0.486	54.0	8.57	2.28	1.77	38.9	9.11	1.93
	32	9.40	8.00	8.500	0.715	0.443	48.3	7.71	2.27	1.73	34.2	8.05	1.91
	29	8.52	7.93	8.464	0.645	0.407	43.6	7.00	2.26	1.70	30.2	7.14	1.88
ST 8 WF	25	7.35	8.13	7.073	0.628	0.380	42.2	6.77	2.40	1.89	17.4	4.92	1.54
	22.5	6.62	8.06	7.039	0.563	0.346	37.8	6.10	2.39	1.87	15.2	4.33	1.52
	20	5.88	8.00	7.000	0.503	0.307	33.2	5.37	2.37	1.82	13.3	3.79	1.50
	18	5.30	7.93	6.992	0.428	0.299	30.7	5.10	2.41	1.90	11.1	3.17	1.45
ST 7 WF	105.5	31.04	7.88	15.800	1.563	0.980	102.2	16.2	1.81	1.57	514.3	65.1	4.07
	101	29.70	7.82	15.750	1.503	0.930	95.7	15.2	1.80	1.53	489.8	62.2	4.06
	96.5	28.36	7.75	15.710	1.438	0.890	90.1	14.4	1.78	1.49	465.1	59.2	4.05
	92	27.04	7.69	15.660	1.378	0.840	83.9	13.4	1.76	1.45	441.4	56.4	4.04
	88	25.87	7.63	15.640	1.313	0.820	80.2	12.9	1.76	1.42	418.9	53.6	4.02
	83.5	24.55	7.56	15.600	1.248	0.780	75.0	12.1	1.75	1.39	395.1	50.7	4.01
	79	23.24	7.50	15.550	1.188	0.730	69.3	11.3	1.73	1.34	372.5	47.9	4.00
	75	22.04	7.44	15.515	1.128	0.695	64.9	10.6	1.72	1.31	351.3	45.3	3.99
	71	20.92	7.38	15.500	1.063	0.680	62.1	10.2	1.72	1.29	330.1	42.6	3.97
	68	19.99	7.38	14.740	1.063	0.660	60.0	9.89	1.73	1.31	283.9	38.5	3.77
ST 7 WF	63.5	18.67	7.31	14.690	0.998	0.610	54.7	9.04	1.71	1.26	263.8	35.9	3.76
	59.5	17.49	7.25	14.650	0.938	0.570	50.4	8.36	1.70	1.22	245.9	33.6	3.75
	55.5	16.33	7.19	14.620	0.873	0.540	46.7	7.80	1.69	1.19	227.4	31.1	3.73
	51.5	15.13	7.13	14.575	0.813	0.495	42.4	7.10	1.67	1.15	209.9	28.8	3.72
	47.5	13.97	7.06	14.545	0.748	0.465	39.1	6.58	1.67	1.12	191.9	26.4	3.71
	43.5	12.78	7.00	14.5	0.688	0.420	34.9	5.88	1.65	1.08	174.8	24.1	3.70

TABLE 51.81 (Continued)

Section Number	Weight per Foot (lb)	Area (in. ²)	Depth of Tee (in.)	Flange		Stem Thickness (in.)	Axis X-X				Axis Y-Y		
				Width (in.)	Average Thickness (in.)		<i>I</i> (in. ⁴)	<i>S</i> (in. ³)	<i>r</i> (in.)	<i>y</i> (in.)	<i>I</i> (in. ⁴)	<i>S</i> (in. ³)	<i>r</i> (in.)
ST 7 WF	42	12.36	7.09	12.023	0.778	0.451	37.4	6.36	1.74	1.21	112.7	18.8	3.02
	39	11.47	7.03	12.000	0.718	0.428	34.8	5.96	1.74	1.19	103.5	17.2	3.00
ST 7 WF	37	10.88	7.10	10.072	0.783	0.450	36.1	6.26	1.82	1.32	66.7	13.3	2.48
	34	10.00	7.03	10.040	0.718	0.418	33.0	5.74	1.81	1.29	60.6	12.1	2.46
ST 7 WF	30.5	8.97	6.96	10.000	0.643	0.378	29.2	5.13	1.80	1.25	53.6	10.7	2.45
	26.5	7.79	6.97	8.062	0.658	0.370	27.7	4.95	1.88	1.38	28.8	7.14	1.92
	24	7.06	6.91	8.031	0.593	0.339	24.9	4.49	1.88	1.35	25.6	6.38	1.91
	21.5	6.32	6.84	8.000	0.528	0.308	22.2	4.02	1.87	1.33	22.6	5.64	1.89
ST 7 WF ^d	19	5.59	7.06	6.776	0.513	0.313	23.5	4.27	2.05	1.56	12.3	3.64	1.49
	17	5.00	7.00	6.750	0.453	0.287	21.1	3.86	2.05	1.55	10.6	3.15	1.46
	15	4.41	6.93	6.733	0.383	0.270	19.0	3.55	2.08	1.59	8.77	2.61	1.41
ST 6 WF	80.5	23.69	6.94	12.515	1.486	0.905	62.6	11.5	1.63	1.47	243.1	38.9	3.20
	66.5	19.56	6.69	12.365	1.236	0.755	48.4	9.03	1.57	1.33	195.0	31.5	3.16
	60	17.65	6.56	12.320	1.106	0.710	43.4	8.22	1.57	1.28	172.5	28.0	3.13
	53	15.59	6.44	12.230	0.986	0.620	36.7	7.01	1.53	1.20	150.4	24.6	3.11
	49.5	14.54	6.38	12.190	0.921	0.580	33.7	6.46	1.52	1.16	139.1	22.8	3.09
	46	13.53	6.31	12.155	0.856	0.545	31.0	5.98	1.51	1.13	128.2	21.1	3.08
	42.5	12.49	6.25	12.105	0.796	0.495	27.8	5.38	1.49	1.08	117.7	19.5	3.07
	39.5	11.61	6.19	12.080	0.736	0.470	25.8	5.02	1.48	1.06	108.2	17.9	3.05
	36	10.58	6.13	12.040	0.671	0.430	23.1	4.53	1.48	1.02	97.6	16.2	3.04
	32.5	9.55	6.06	12.000	0.606	0.390	20.6	4.06	1.47	0.98	87.3	14.6	3.02
ST 6 WF	29	8.53	6.10	10.014	0.641	0.359	19.0	3.75	1.49	1.03	53.7	10.7	2.51
	26.5	7.80	6.03	10.000	0.576	0.345	17.7	3.54	1.51	1.02	48.0	9.60	2.48
ST 6 WF	25	7.36	6.10	8.077	0.641	0.371	18.7	3.80	1.60	1.17	28.2	6.98	1.96
	22.5	6.62	6.03	8.042	0.576	0.336	16.6	3.40	1.59	1.13	25.0	6.20	1.94
	20	5.89	5.97	8.000	0.516	0.294	14.4	2.94	1.56	1.08	22.0	5.50	1.94
ST 6 WF	18	5.29	6.12	6.565	0.540	0.305	15.3	3.14	1.70	1.26	11.9	3.62	1.50
	15.5	4.56	6.04	6.525	0.465	0.265	13.0	2.69	1.69	1.22	9.9	3.04	1.47
	13.5	3.98	5.98	6.500	0.400	0.240	11.4	2.39	1.69	1.21	8.3	2.55	1.44
ST 6 WF	7	2.07	5.96	3.970	0.224	0.200	7.66	1.83	1.92	1.76	1.13	0.57	0.74
ST 6 I ^b	25	7.29	6.00	5.477	0.660	0.687	25.2	6.05	1.85	1.84	7.85	2.87	1.03
	20.4	5.92	6.00	5.250	0.660	0.460	18.8	4.26	1.77	1.57	6.77	2.58	1.06
ST 6 I	17.5	5.10	6.00	5.078	0.544	0.428	17.2	3.95	1.83	1.65	4.93	1.94	0.98
	15.9	4.63	6.00	5.000	0.544	0.350	14.9	3.31	1.78	1.51	4.68	1.87	1.00
ST 5 I	17.5	5.11	5.00	4.944	0.491	0.594	12.5	3.63	1.56	1.56	4.18	1.69	0.90
	12.7	3.69	5.00	4.660	0.491	0.310	7.81	2.05	1.45	1.20	3.39	1.46	0.95
ST 4 I	11.5	3.36	4.00	4.171	0.425	0.441	5.03	1.77	1.22	1.15	2.15	1.03	0.80
	9.2	2.67	4.00	4.000	0.425	0.270	3.50	1.14	1.14	0.94	1.86	0.93	0.83
ST 3.5 I	10	2.92	3.50	3.860	0.392	0.450	3.36	1.36	1.07	1.04	1.58	0.82	0.73
	7.65	2.22	3.50	3.660	0.392	0.250	2.18	0.81	0.99	0.81	1.32	0.72	0.77
ST 3 I	8.625	2.51	3.00	3.565	0.359	0.465	2.13	1.02	0.92	0.91	1.15	0.65	0.67
	6.25	1.81	3.00	3.330	0.359	0.230	1.27	0.55	0.83	0.69	0.93	0.56	0.71
ST 5 WF	56	16.46	5.69	10.415	1.248	0.755	28.8	6.42	1.32	1.21	117.7	22.6	2.67
	50	14.72	5.56	10.345	1.118	0.685	24.8	5.62	1.30	1.14	103.3	20.0	2.65
	44.5	13.09	5.44	10.275	0.998	0.615	21.3	4.88	1.28	1.07	90.3	17.6	2.63

(continued)

TABLE 51.81 (Continued)


Section Number	Weight per Foot (lb)	Area (in. ²)	Depth of Tee (in.)	Flange		Stem Thickness (in.)	Axis X-X				Axis Y-Y		
				Width (in.)	Average Thickness (in.)		<i>I</i> (in. ⁴)	<i>S</i> (in. ³)	<i>r</i> (in.)	<i>y</i> (in.)	<i>I</i> (in. ⁴)	<i>S</i> (in. ³)	<i>r</i> (in.)
ST 5 WF	38.5	11.33	5.31	10.195	0.868	0.535	17.7	4.10	1.25	1.00	76.7	15.1	2.60
	36	10.59	5.25	10.170	0.808	0.510	16.4	3.83	1.24	0.97	70.9	13.9	2.59
	33	9.70	5.19	10.117	0.748	0.457	14.5	3.39	1.22	0.92	64.6	12.8	2.58
	30	8.83	5.13	10.075	0.683	0.415	12.8	3.02	1.21	0.88	58.2	11.6	2.57
	27	7.94	5.06	10.028	0.618	0.368	11.2	2.64	1.18	0.84	51.95	10.4	2.56
	24.5	7.20	5.00	10.000	0.558	0.340	10.1	2.40	1.18	0.81	46.5	9.30	2.54
	22.5	6.62	5.06	8.022	0.618	0.350	10.3	2.48	1.25	0.91	26.6	6.63	2.00
	19.5	5.74	4.97	7.990	0.528	0.318	8.96	2.19	1.25	0.88	22.5	5.62	1.98
ST 5 WF ^a	16.5	4.85	4.88	7.964	0.433	0.292	7.80	1.95	1.27	0.88	18.2	4.58	1.94
	14.5	4.27	5.11	5.799	0.500	0.289	8.38	2.07	1.40	1.05	7.61	2.62	1.34
	12.5	3.67	5.04	5.762	0.430	0.252	7.12	1.77	1.39	1.02	6.34	2.20	1.31
ST 4 WF	10.5	3.10	4.95	5.750	0.340	0.240	6.31	1.62	1.43	1.06	4.87	1.69	1.25
	33.5	9.85	4.50	8.287	0.933	0.575	10.94	3.07	1.05	0.94	44.3	10.7	2.12
	29	8.53	4.38	8.222	0.808	0.510	9.11	2.60	1.03	0.87	37.5	9.10	2.10
	24	7.06	4.25	8.117	0.683	0.405	6.92	2.00	0.99	0.78	30.45	7.50	2.08
	20	5.88	4.13	8.077	0.558	0.365	5.80	1.71	0.99	0.74	24.5	6.05	2.04
ST 4 WF	17.5	5.15	4.06	8.027	0.493	0.315	4.88	1.45	0.97	0.69	21.25	5.30	2.03
	15.5	4.56	4.00	8.000	0.433	0.288	4.31	1.30	0.97	0.67	18.5	4.60	2.01
	14	4.11	4.03	6.540	0.463	0.285	4.22	1.28	1.01	0.73	10.8	3.30	1.62
	12	3.53	3.97	6.500	0.398	0.245	3.53	1.08	1.00	0.70	9.10	2.80	1.61
	10	2.94	4.07	5.268	0.378	0.248	3.66	1.13	1.12	0.83	4.25	1.61	1.20
ST 4 WF	8.5	2.50	4.00	5.250	0.308	0.230	3.21	1.01	1.13	0.84	3.36	1.28	1.16

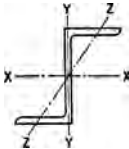
Nominal Size (in.)	Weight per Foot (lb)	Area (in. ²)	Depth (in.)	Dimensions			Axis X-X				Axis Y-Y		
				Width Flange (in.)	Minimum Flange (in.)	Thickness Stem (in.)	<i>I</i> (in. ⁴)	<i>S</i> (in. ³)	<i>r</i> (in.)	<i>y</i> (in.)	<i>I</i> (in. ⁴)	<i>S</i> (in. ³)	<i>r</i> (in.)
5 × 3 $\frac{1}{8}$	13.6	4.00	3 $\frac{1}{8}$	5	$\frac{1}{2}$	$\frac{13}{32}$	2.7	1.1	0.82	0.76	5.2	2.1	1.14
5 × 3	11.5	3.37	3	5	$\frac{3}{8}$	$\frac{13}{32}$	2.4	1.1	0.84	0.76	3.9	1.6	1.10
4 × 4 $\frac{1}{2}$	11.2	3.29	4 $\frac{1}{2}$	4	$\frac{3}{8}$	$\frac{3}{8}$	6.3	2.0	1.39	1.31	2.1	1.1	0.80
4 × 4	13.5	3.97	4	4	$\frac{1}{2}$	$\frac{1}{2}$	5.7	2.0	1.20	1.18	2.8	1.4	0.84
4 × 3	9.2	2.68	3	4	$\frac{3}{8}$	$\frac{3}{8}$	2.0	0.90	0.86	0.78	2.1	1.1	0.89
4 × 2 $\frac{1}{2}$	8.5	2.48	2 $\frac{1}{2}$	4	$\frac{3}{8}$	$\frac{3}{8}$	1.2	0.62	0.69	0.62	2.1	1.0	0.92
3 × 3	7.8	2.29	3	3	$\frac{3}{8}$	$\frac{3}{8}$	1.84	0.86	0.89	0.88	0.89	0.60	0.63
3 × 3	6.7	1.97	3	3	$\frac{5}{16}$	$\frac{5}{16}$	1.61	0.74	0.90	0.85	0.75	0.50	0.62
3 × 2 $\frac{1}{2}$	6.1	1.77	2 $\frac{1}{2}$	3	$\frac{5}{16}$	$\frac{5}{16}$	0.94	0.51	0.73	0.68	0.75	0.50	0.65
2 $\frac{1}{2}$ × 2 $\frac{1}{2}$	6.4	1.87	2 $\frac{1}{2}$	2 $\frac{1}{2}$	$\frac{3}{8}$	$\frac{3}{8}$	1.0	0.59	0.74	0.76	0.52	0.42	0.53
2 $\frac{1}{2}$ × 2 $\frac{1}{2}$	4.6	1.33	2 $\frac{1}{2}$	2 $\frac{1}{2}$	$\frac{1}{4}$	$\frac{1}{4}$	0.74	0.42	0.75	0.71	0.34	0.27	0.51
2 $\frac{1}{4}$ × 2 $\frac{1}{4}$	4.1	1.19	2 $\frac{1}{4}$	2 $\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}$	0.52	0.32	0.66	0.65	0.25	0.22	0.46
2 × 2	4.3	1.26	2	2	$\frac{5}{16}$	$\frac{5}{16}$	0.44	0.31	0.59	0.61	0.23	0.23	0.43
2 × 2	3.56	1.05	2	2	$\frac{1}{4}$	$\frac{1}{4}$	0.37	0.26	0.59	0.59	0.18	0.18	0.42

^aWF indicates structural tee cut from wide-flange section.

^bI indicates structural tee cut from standard beam section.

TABLE 51.82 Properties and Dimensions of Zees




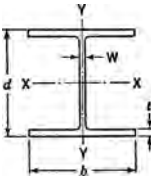


Zees are seldom used as structural framing members. When so used they are generally employed on short spans in flexure. This table lists a few selected sizes, the range of whose section moduli will cover all ordinary conditions. For sizes not listed, the catalogs of the respective rolling mills should be consulted.

Nominal Size (in.)	Weight per Foot (lb)	Area (in. ²)	Depth (in.)	Dimensions		Axis X-X			Axis Y-Y			Axis Z-Z
				Width of Flange (in.)	Thickness (in.)	<i>I</i> (in. ⁴)	<i>S</i> (in. ³)	<i>r</i> (in.)	<i>I</i> (in. ⁴)	<i>S</i> (in. ³)	<i>r</i> (in.)	<i>r</i> (in.)
6 × 3½	21.1	6.19	6⅛	3⅝	½	34.4	11.2	2.36	12.9	3.8	1.44	0.84
	15.7	4.59	6	3½	⅜	25.3	8.4	2.35	9.1	2.8	1.41	0.83
5 × 3¼	17.9	5.25	5	3¼	½	19.2	7.7	1.91	9.1	3.0	1.31	0.74
	16.4	4.81	5⅛	3⅜	⅞	19.1	7.4	1.99	9.2	2.9	1.38	0.77
	14.0	4.10	5⅒	3⅝	⅜	16.2	6.4	1.99	7.7	2.5	1.37	0.76
	11.6	3.40	5	3¼	⅞	13.4	5.3	1.98	6.2	2.0	1.35	0.75
4 × 3⅒	15.9	4.66	4⅒	3⅛	½	11.2	5.5	1.55	8.0	2.8	1.31	0.67
	12.5	3.66	4⅛	3⅒	⅜	9.6	4.7	1.62	6.8	2.3	1.36	0.69
	10.3	3.03	4⅒	3⅛	⅞	7.9	3.9	1.62	5.5	1.8	1.34	0.68
	8.2	2.41	4	3⅒	¼	6.3	3.1	1.62	4.2	1.4	1.33	0.67
3 × 2⅒	12.6	3.69	3	2⅒	½	4.6	3.1	1.12	4.9	2.0	1.15	0.53
	9.8	2.86	3	2⅒	⅜	3.9	2.6	1.16	3.9	1.6	1.17	0.54
	6.7	1.97	3	2⅒	¼	2.9	1.9	1.21	2.8	1.1	1.19	0.55

TABLE 51.83 Properties and Dimensions of H Bearing Piles





Section Number and Nominal Size	Weight per Foot (lb)	Area <i>A</i> (in. ²)	Depth <i>d</i> (in.)	Flange		Web Thickness <i>W</i> (in.)	Axis X-X			Axis Y-Y		
				Width <i>b</i> (in.)	Thickness <i>t</i> (in.)		<i>I</i> (in. ⁴)	<i>S</i> (in. ³)	<i>r</i> (in.)	<i>I</i> ' (in. ⁴)	<i>S</i> ' (in. ³)	<i>r</i> ' (in.)
BP 14,	117	34.44	14.234	14.885	0.805	0.805	1228.5	172.6	5.97	443.1	59.5	3.59
14 × 14½	102	30.01	14.032	14.784	0.704	0.704	1055.1	150.4	5.93	379.6	51.3	3.56
	89	26.19	13.856	14.696	0.616	0.616	909.1	131.2	5.89	326.2	44.4	3.53
	73	21.46	13.636	14.586	0.506	0.506	733.1	107.5	5.85	261.9	35.9	3.49
BP 12,	74	21.76	12.122	12.217	0.607	0.607	566.5	93.5	5.10	184.7	30.2	2.91
12 × 12	53	15.58	11.780	12.046	0.436	0.436	394.8	67.0	5.03	127.3	21.2	2.86
BP 10,	57	16.76	10.012	10.224	0.564	0.564	294.7	58.9	4.19	100.6	19.7	2.45
10 × 10	42	12.35	9.720	10.078	0.418	0.418	210.8	43.4	4.13	71.4	14.2	2.40
BP 8, 8 × 8	36	10.60	8.026	8.158	0.446	0.446	119.8	29.9	3.36	40.4	9.9	1.95

TABLE 51.84 Square and Round Bars^a

Square					Round				
Size (in.)	Weight/ft (lb)	Area (in. ²)	Weight/ft (lb)	Area (in. ²)	Size (in.)	Weight/ft (lb)	Area (in. ²)	Weight/ft (lb)	Area (in. ²)
0					$\frac{1}{4}$	17.213	5.0625	13.519	3.9761
$\frac{1}{16}$	0.013	0.0039	0.010	0.0031	$\frac{5}{16}$	18.182	5.3477	14.280	4.2000
$\frac{1}{8}$	0.053	0.0156	0.042	0.0123	$\frac{3}{8}$	19.178	5.6406	15.062	4.4301
$\frac{3}{16}$	0.120	0.0352	0.094	0.0276	$\frac{7}{16}$	20.201	5.9414	15.866	4.6664
$\frac{1}{4}$	0.213	0.0625	0.167	0.0491	$\frac{1}{2}$	21.250	6.2500	16.690	4.9087
$\frac{5}{16}$	0.332	0.0977	0.261	0.0767	$\frac{9}{16}$	22.326	6.5664	17.534	5.1572
$\frac{3}{8}$	0.478	0.1406	0.376	0.1105	$\frac{5}{8}$	23.428	6.8906	18.400	5.4119
$\frac{7}{16}$	0.651	0.1914	0.511	0.1503	$\frac{11}{16}$	24.557	7.2227	19.287	5.6727
$\frac{1}{2}$	0.850	0.2500	0.668	0.1963	$\frac{3}{4}$	25.713	7.5625	20.195	5.9396
$\frac{9}{16}$	1.076	0.3164	0.845	0.2485	$\frac{13}{16}$	26.895	7.9102	21.123	6.2126
$\frac{5}{8}$	1.328	0.3906	1.043	0.3068	$\frac{7}{8}$	28.103	8.2656	22.072	6.4918
$\frac{11}{16}$	1.607	0.4727	1.262	0.3712	$\frac{15}{16}$	29.338	8.6289	23.042	6.7771
$\frac{3}{4}$	1.913	0.5625	1.502	0.4418	3	30.60	9.000	24.03	7.069
$\frac{13}{16}$	2.245	0.6602	1.763	0.5185	$\frac{1}{16}$	31.89	9.379	25.05	7.366
$\frac{7}{8}$	2.603	0.7656	2.044	0.6013	$\frac{1}{8}$	33.20	9.766	26.08	7.670
$\frac{15}{16}$	2.988	0.8789	2.347	0.6903	$\frac{3}{16}$	34.54	10.160	27.13	7.980
1	3.400	1.0000	2.670	0.7854	$\frac{1}{4}$	35.91	10.563	28.21	8.296
$\frac{1}{16}$	3.838	1.1289	3.015	0.8866	$\frac{5}{16}$	37.31	10.973	29.30	8.618
$\frac{1}{8}$	4.303	1.2656	3.380	0.9940	$\frac{3}{8}$	38.73	11.391	30.42	8.946
$\frac{3}{16}$	4.795	1.4102	3.766	1.1075	$\frac{7}{16}$	40.18	11.816	31.55	9.281
$\frac{1}{4}$	5.313	1.5625	4.172	1.2272	$\frac{1}{2}$	41.65	12.250	32.71	9.621
$\frac{5}{16}$	5.857	1.7227	4.600	1.3530	$\frac{9}{16}$	43.15	12.691	33.89	9.968
$\frac{3}{8}$	6.428	1.8906	5.049	1.4849	$\frac{5}{8}$	44.68	13.141	35.09	10.321
$\frac{7}{16}$	7.026	2.0664	5.518	1.6230	$\frac{11}{16}$	46.23	13.598	36.31	10.680
$\frac{1}{2}$	7.650	2.2500	6.008	1.7671	$\frac{3}{4}$	47.81	14.063	37.55	11.045
$\frac{9}{16}$	8.301	2.4414	6.519	1.9175	$\frac{13}{16}$	49.42	14.535	38.81	11.416
$\frac{5}{8}$	8.978	2.6406	7.051	2.0739	$\frac{7}{8}$	51.05	15.016	40.10	11.793
$\frac{11}{16}$	9.682	2.8477	7.604	2.2365	$\frac{15}{16}$	52.71	15.504	41.40	12.177
$\frac{3}{4}$	10.413	3.0625	8.178	2.4053	4	54.40	16.000	42.73	12.566
$\frac{13}{16}$	11.170	3.2852	8.773	2.5802	$\frac{1}{16}$	56.11	16.504	44.07	12.962
$\frac{7}{8}$	11.953	3.5156	9.388	2.7612	$\frac{1}{8}$	57.85	17.016	45.44	13.364
$\frac{15}{16}$	12.763	3.7539	10.024	2.9483	$\frac{3}{16}$	59.62	17.535	46.83	13.772
2	13.600	4.0000	10.681	3.1416	$\frac{1}{4}$	61.41	18.063	48.23	14.186
$\frac{1}{16}$	14.463	4.2539	11.359	3.3410	$\frac{5}{16}$	63.23	18.598	49.66	14.607
$\frac{1}{8}$	15.353	4.5156	12.058	3.5466	$\frac{3}{8}$	65.08	19.141	51.11	15.033
$\frac{3}{16}$	16.270	4.7852	12.778	3.7583					

TABLE 51.84 (Continued)

Size (in.)	Square		Round		Size (in.)	Square		Round	
	Weight/ft (lb)	Area (in. ²)	Weight/ft (lb)	Area (in. ²)		Weight/ft (lb)	Area (in. ²)	Weight/ft (lb)	Area (in. ²)
$\frac{7}{16}$	66.95	19.691	52.58	15.466	$\frac{1}{4}$	132.81	39.063	104.31	30.680
$\frac{1}{2}$	68.85	20.250	54.07	15.904	$\frac{5}{16}$	135.48	39.848	106.41	31.296
$\frac{9}{16}$	70.78	20.816	55.59	16.349	$\frac{3}{8}$	138.18	40.641	108.53	31.919
$\frac{5}{8}$	72.73	21.391	57.12	16.800	$\frac{7}{16}$	140.90	41.441	110.66	32.548
$\frac{11}{16}$	74.71	21.973	58.67	17.257	$\frac{1}{2}$	143.65	42.250	112.82	33.183
$\frac{3}{4}$	76.71	22.563	60.25	17.721	$\frac{9}{16}$	146.43	43.066	115.00	33.824
$\frac{13}{16}$	78.74	23.160	61.85	18.190	$\frac{5}{8}$	149.23	43.891	117.20	34.472
$\frac{7}{8}$	80.80	23.766	63.46	18.665	$\frac{11}{16}$	152.06	44.723	119.43	35.125
$\frac{15}{16}$	82.89	24.379	65.10	19.147	$\frac{3}{4}$	154.91	45.563	121.67	35.785
5	85.00	25.000	66.76	19.635	$\frac{13}{16}$	157.79	46.410	123.93	36.450
$\frac{1}{16}$	87.14	25.629	68.44	20.129	$\frac{7}{8}$	160.70	47.266	126.22	37.122
$\frac{1}{8}$	89.30	26.266	70.14	20.629	$\frac{15}{16}$	163.64	48.129	128.52	37.800
$\frac{3}{16}$	91.49	26.910	71.86	21.135	7	166.60	49.000	130.85	38.485
$\frac{1}{4}$	93.71	27.563	73.60	21.648	$\frac{1}{16}$	169.59	49.879	133.19	39.175
$\frac{5}{16}$	95.96	28.223	75.36	22.166	$\frac{1}{8}$	172.60	50.766	135.56	39.871
$\frac{3}{8}$	98.23	28.891	77.15	22.691	$\frac{3}{16}$	175.64	51.660	137.95	40.574
$\frac{7}{16}$	100.53	29.566	78.95	23.221	$\frac{1}{4}$	178.71	52.563	140.36	41.282
$\frac{1}{2}$	102.85	30.250	80.78	23.758	$\frac{5}{16}$	181.81	53.473	142.79	41.997
$\frac{9}{16}$	105.20	30.941	82.62	24.301	$\frac{3}{8}$	184.93	54.391	145.24	42.718
$\frac{5}{8}$	107.58	31.641	84.49	24.850	$\frac{7}{16}$	188.07	55.316	147.71	43.445
$\frac{11}{16}$	109.98	32.348	86.38	25.406	$\frac{1}{2}$	191.25	56.250	150.21	44.179
$\frac{3}{4}$	112.41	33.063	88.29	25.967	$\frac{9}{16}$	194.45	57.191	152.72	44.918
$\frac{13}{16}$	114.87	33.785	90.22	26.535	$\frac{5}{8}$	197.68	58.141	155.26	45.664
$\frac{7}{8}$	117.35	34.516	92.17	27.109	$\frac{11}{16}$	200.93	59.098	157.81	46.415
$\frac{15}{16}$	119.86	35.254	94.14	27.688	$\frac{3}{4}$	204.21	60.063	160.39	47.173
6	122.40	36.000	96.13	28.274	$\frac{13}{16}$	207.52	61.035	162.99	47.937
$\frac{1}{16}$	124.96	36.754	98.15	28.866	$\frac{7}{8}$	210.85	62.016	165.60	48.707
$\frac{1}{8}$	127.55	37.516	100.18	29.465	$\frac{15}{16}$	214.21	63.004	168.24	49.483
$\frac{3}{16}$	130.17	38.285	102.23	30.069	8	217.60	64.000	170.90	50.265

^aOne cubic inch of rolled steel is assumed to weigh 0.2833 lb.

TABLE 51.85 Dimensions of Ferrous Pipe

Nominal Pipe Size (in.)	Outside Diameter (in.)	Schedule No.	Wall Thickness (in.)	Inside Diameter (in.)	Cross-Sectional Area		Circumference, ft, or surface, ft ² /ft of Length		Capacity at 1 ft/sec Velocity		Weight of Plain-End Pipe (lb/ft)
					Metal (in. ²)	Flow (ft ²)	Outside	Inside	U.S. gal/min	lb/hr water	
1/2	0.405	10S	0.049	0.307	0.055	0.00051	0.106	0.0804	0.231	115.5	0.19
		40ST, 40S	0.068	0.269	0.072	0.00040	0.106	0.0705	0.179	89.5	0.24
		80XS, 80S	0.095	0.215	0.093	0.00025	0.106	0.0563	0.113	56.5	0.31
1	0.540	10S	0.065	0.410	0.097	0.00092	0.141	0.107	0.412	206.5	0.33
		40ST, 40S	0.088	0.364	0.125	0.00072	0.141	0.095	0.323	161.5	0.42
		80XS, 80S	0.119	0.302	0.157	0.00050	0.141	0.079	0.224	112.0	0.54
3/4	0.675	10S	0.065	0.545	0.125	0.00162	0.177	0.143	0.727	363.5	0.42
		40ST, 40S	0.091	0.493	0.167	0.00133	0.177	0.129	0.596	298.0	0.57
		80XS, 80S	0.126	0.423	0.217	0.00098	0.177	0.111	0.440	220.0	0.74
1	0.840	5S	0.065	0.710	0.158	0.00275	0.220	0.186	1.234	617.0	0.54
		10S	0.083	0.674	0.197	0.00248	0.220	0.176	1.112	556.0	0.67
		40ST, 40S	0.109	0.622	0.250	0.00211	0.220	0.163	0.945	472.0	0.85
		80XS, 80S	0.147	0.546	0.320	0.00163	0.220	0.143	0.730	365.0	1.09
		160	0.188	0.464	0.385	0.00117	0.220	0.122	0.527	263.5	1.31
		XX	0.294	0.252	0.504	0.00035	0.220	0.066	0.155	77.5	1.71
1 1/2	1.050	5S	0.065	0.920	0.201	0.00461	0.275	0.241	2.072	1036.0	0.69
		10S	0.083	0.884	0.252	0.00426	0.275	0.231	1.903	951.5	0.86
		40ST, 40S	0.113	0.824	0.333	0.00371	0.275	0.216	1.665	832.5	1.13
		80XS, 80S	0.154	0.742	0.433	0.00300	0.275	0.194	1.345	672.5	1.47
		160	0.219	0.612	0.572	0.00204	0.275	0.160	0.917	458.5	1.94
		XX	0.308	0.434	0.718	0.00103	0.275	0.114	0.461	230.5	2.44
1	1.315	5S	0.065	1.185	0.255	0.00768	0.344	0.310	3.449	1725	0.87
		10S	0.109	1.097	0.413	0.00656	0.344	0.287	2.946	1473	1.40
		40ST, 40S	0.133	1.049	0.494	0.00600	0.344	0.275	2.690	1345	1.68
		80XS, 80S	0.179	0.957	0.639	0.00499	0.344	0.250	2.240	1120	2.17

1½	1.660	160	0.250	0.815	0.836	0.00362	0.344	0.213	1.625	812.5	2.84
		XX	0.358	0.599	1.076	0.00196	0.344	0.157	0.878	439.0	3.66
		5S	0.065	1.530	0.326	0.01277	0.435	0.401	5.73	2865	1.11
		10S	0.109	1.442	0.531	0.01134	0.435	0.378	5.09	2545	1.81
		40ST, 40S	0.140	1.380	0.668	0.01040	0.435	0.361	4.57	2285	2.27
		80XS, 80S	0.191	1.278	0.881	0.00891	0.435	0.335	3.99	1995	3.00
		160	0.250	1.160	1.107	0.00734	0.435	0.304	3.29	1645	3.76
		XX	0.382	0.896	1.534	0.00438	0.435	0.235	1.97	985	5.21
		5S	0.065	1.770	0.375	0.01709	0.497	0.463	7.67	3835	1.28
		10S	0.109	1.682	0.614	0.01543	0.497	0.440	6.94	3465	2.09
2	2.375	40ST, 40S	0.145	1.610	0.800	0.01414	0.497	0.421	6.34	3170	2.72
		80SX, 80S	0.200	1.500	1.069	0.01225	0.497	0.393	5.49	2745	3.63
		160	0.281	1.338	1.429	0.00976	0.497	0.350	4.38	2190	4.86
		XX	0.400	1.100	1.885	0.00660	0.497	0.288	2.96	1480	6.41
		5S	0.065	2.245	0.472	0.02749	0.622	0.588	12.34	6170	1.61
		10S	0.109	2.157	0.776	0.02538	0.622	0.565	11.39	5695	2.64
		40ST, 40S	0.154	2.067	1.075	0.02330	0.622	0.541	10.45	5225	3.65
		80ST, 80S	0.218	1.939	1.477	0.02050	0.622	0.508	9.20	4600	5.02
		160	0.344	1.687	2.195	0.01552	0.622	0.436	6.97	3485	7.46
		XX	0.436	1.503	2.656	0.01232	0.622	0.393	5.53	2765	9.03
2½	2.875	5S	0.083	2.709	0.728	0.04003	0.753	0.709	17.97	8985	2.48
		10S	0.120	2.635	1.039	0.03787	0.753	0.690	17.00	8500	3.53
		40ST, 40S	0.203	2.469	1.704	0.03322	0.753	0.647	14.92	7460	5.79
		80XS, 80S	0.276	2.323	2.254	0.02942	0.753	0.608	13.20	6600	7.66
		160	0.375	2.125	2.945	0.02463	0.753	0.556	11.07	5535	10.01
		XX	0.552	1.771	4.028	0.01711	0.753	0.464	7.68	3840	13.70
		5S	0.083	3.334	0.891	0.06063	0.916	0.873	27.21	13,605	3.03
		10S	0.120	3.260	1.274	0.05796	0.916	0.853	26.02	13,010	4.33
		40ST, 40S	0.216	3.068	2.228	0.05130	0.916	0.803	23.00	11,500	7.58
		80XS, 80S	0.300	2.900	3.016	0.04587	0.916	0.759	20.55	10,275	10.25
3	3.500	160	0.438	2.624	4.213	0.03755	0.916	0.687	16.86	8430	14.31
		XX	0.600	2.300	5.466	0.02885	0.916	0.602	12.95	6475	18.58

(continued)

TABLE 51.85 (Continued)

Nominal Pipe Size (in.)	Outside Diameter (in.)	Schedule No.	Wall Thickness (in.)	Inside Diameter (in.)	Cross-Sectional Area		Circumference, ft, or surface, ft ² /ft of Length		Capacity at 1 ft/sec Velocity		Weight of Plain-End Pipe (lb/ft)
					Metal (in. ²)	Flow (ft ²)	Outside	Inside	U.S. gal/min	lb/hr water	
3½	4.0	5S	0.083	3.834	1.021	0.08017	1.047	1.004	35.98	17,990	3.48
		10S	0.120	3.760	1.463	0.07711	1.047	0.984	34.61	17,305	4.97
		40ST, 40S	0.226	3.548	2.680	0.06870	1.047	0.929	30.80	15,400	9.11
		80XS, 80S	0.318	3.364	3.678	0.06170	1.047	0.881	27.70	13,850	12.51
4	4.5	5S	0.083	4.334	1.152	0.10245	1.178	1.135	46.0	23,000	3.92
		10S	0.120	4.260	1.651	0.09898	1.178	1.115	44.4	22,200	5.61
		40ST, 40S	0.237	4.026	3.17	0.08840	1.178	1.054	39.6	19,800	10.79
		80XS, 80S	0.337	3.826	4.41	0.07986	1.178	1.002	35.8	17,900	14.98
		120	0.438	3.624	5.58	0.07170	1.178	0.949	32.2	16,100	19.01
		160	0.531	3.438	6.62	0.06647	1.178	0.900	28.9	14,450	22.52
5	5.563	XX	0.674	3.152	8.10	0.05419	1.178	0.825	24.3	12,150	27.54
		5S	0.109	5.345	1.87	0.1558	1.456	1.399	69.9	34,950	6.36
		10S	0.134	5.295	2.29	0.1529	1.456	1.386	68.6	34,300	7.77
		40ST, 40S	0.258	5.047	4.30	0.1390	1.456	1.321	62.3	31,150	14.62
		80XS, 80S	0.375	4.813	6.11	0.1263	1.456	1.260	57.7	28,850	20.78
		120	0.500	4.563	7.95	0.1136	1.456	1.195	51.0	25,500	27.04
6	6.625	160	0.625	4.313	9.70	0.1015	1.456	1.129	45.5	22,750	32.96
		XX	0.750	4.063	11.34	0.0900	1.456	1.064	40.4	20,200	38.55
		5S	0.109	6.407	2.23	0.2239	1.734	1.677	100.5	50,250	7.60
		10S	0.134	6.357	2.73	0.2204	1.734	1.664	98.9	49,450	9.29
		40ST, 40S	0.280	6.065	5.58	0.2006	1.734	1.588	90.0	45,000	18.97
		80XS, 80S	0.432	5.761	8.40	0.1810	1.734	1.508	81.1	40,550	28.57
		120	0.562	5.501	10.70	0.1650	1.734	1.440	73.9	36,950	36.42
		160	0.719	5.187	13.34	0.1467	1.734	1.358	65.9	32,950	45.34
		XX	0.864	4.897	15.64	0.1308	1.734	1.282	58.7	29,350	53.16

8	8.625	5S	0.109	8.407	2.915	0.3855	2.258	2.201	173.0	86,500	9.93
		10S	0.148	8.329	3.941	0.3784	2.258	2.180	169.8	84,900	13.40
		20	0.250	8.125	6.578	0.3601	2.258	2.127	161.5	80,750	22.36
		30	0.277	8.071	7.260	0.3553	2.258	2.113	159.4	79,700	24.70
		40ST, 40S	0.322	7.981	8.396	0.3474	2.258	2.089	155.7	77,850	28.55
		60	0.406	7.813	10.48	0.3329	2.258	2.045	149.4	74,700	35.66
		80XS, 80S	0.500	7.625	12.76	0.3171	2.258	1.996	142.3	71,150	43.39
		100	0.594	7.437	14.99	0.3017	2.258	1.947	135.4	67,700	50.93
		120	0.719	7.187	17.86	0.2817	2.258	1.882	126.4	63,200	60.69
		140	0.812	7.001	19.93	0.2673	2.258	1.833	120.0	60,000	67.79
10	10.75	XX	0.875	6.875	21.30	0.2578	2.258	1.800	115.7	57,850	72.42
		160	0.906	6.813	21.97	0.2532	2.258	1.784	113.5	56,750	74.71
		5S	0.134	10.842	4.47	0.5993	2.814	2.744	269.0	134,500	15.23
		10S	0.165	10.420	5.49	0.5922	2.814	2.728	265.8	132,900	18.70
		20	0.250	10.250	8.25	0.5731	2.814	2.685	257.0	128,500	28.04
		30	0.307	10.136	10.07	0.5603	2.814	2.655	252.0	126,000	34.24
		40ST, 40S	0.365	10.020	11.91	0.5475	2.814	2.620	246.0	123,000	40.48
		80S, 60XS	0.500	9.750	16.10	0.5185	2.814	2.550	233.0	116,500	54.74
		80	0.594	9.562	18.95	0.4987	2.814	2.503	223.4	111,700	64.40
		100	0.719	9.312	22.66	0.4729	2.814	2.438	212.3	106,150	77.00
12	12.75	120	0.844	9.062	26.27	0.4479	2.814	2.372	201.0	100,500	89.27
		140, XX	1.000	8.750	30.63	0.4176	2.814	2.291	188.0	94,000	104.13
		160	1.125	8.500	34.02	0.3941	2.814	2.225	177.0	88,500	115.65
		5S	0.156	12.438	6.17	0.8438	3.338	3.26	378.7	189,350	22.22
		10S	0.180	12.390	7.11	0.8373	3.338	3.24	375.8	187,900	24.20
		20	0.250	12.250	9.82	0.8185	3.338	3.21	367.0	183,500	33.38
		30	0.330	12.090	12.88	0.7972	3.338	3.17	358.0	179,000	43.77
		ST, 40S	0.375	12.000	14.58	0.7854	3.338	3.14	352.5	176,250	49.56
		40	0.406	11.938	15.74	0.7773	3.338	3.13	349.0	174,500	53.56
		XS, 80S	0.500	11.750	19.24	0.7530	3.338	3.08	338.0	169,000	65.42
1861		60	0.562	11.626	21.52	0.7372	3.338	3.04	331.0	165,500	73.22
		80	0.688	11.374	26.07	0.7056	3.338	2.98	316.7	158,350	88.57

(continued)

TABLE 51.85 (Continued)

Nominal Pipe Size (in.)	Outside Diameter (in.)	Schedule No.	Wall Thickness (in.)	Inside Diameter (in.)	Cross-Sectional Area		Circumference, ft, or surface, ft ² /ft of Length		Capacity at 1 ft/sec Velocity		Weight of Plain-End Pipe (lb/ft)
					Metal (in. ²)	Flow (ft ²)	Outside	Inside	U.S. gal/min	lb/hr water	
14	14	100	0.844	11.062	31.57	0.6674	3.338	2.90	299.6	149,800	107.29
		120, XX	1.000	10.750	36.91	0.6303	3.338	2.81	283.0	141,500	125.49
		140	1.125	10.500	41.09	0.6013	3.338	2.75	270.0	135,000	139.68
		160	1.312	10.126	47.14	0.5592	3.338	2.65	251.0	125,500	160.33
		5S	0.156	13.688	6.78	1.0219	3.665	3.58	459	229,500	22.76
		10S	0.188	13.624	8.16	1.0125	3.665	3.57	454	227,000	27.70
		10	0.250	13.500	10.80	0.9940	3.665	3.53	446	223,000	36.71
		20	0.312	13.376	13.42	0.9750	3.665	3.50	438	219,000	45.68
		30, ST	0.375	13.250	16.05	0.9575	3.665	3.47	430	215,000	54.57
		40	0.438	13.124	18.66	0.9397	3.665	3.44	422	211,000	63.37
16	16	XS	0.500	13.000	21.21	0.9218	3.665	3.40	414	207,000	72.09
		60	0.594	12.812	25.02	0.8957	3.665	3.35	402	201,000	85.01
		80	0.750	12.500	31.22	0.8522	3.665	3.27	382	191,000	106.13
		100	0.938	12.124	38.49	0.8017	3.665	3.17	360	180,000	130.79
		120	1.094	11.812	44.36	0.7610	3.665	3.09	342	171,000	150.76
		140	1.250	11.500	50.07	0.7213	3.665	3.01	324	162,000	170.22
		160	1.406	11.188	55.63	0.6827	3.665	2.93	306	153,000	189.12
		5S	0.165	15.670	8.18	1.3393	4.189	4.10	601	300,500	27.87
		10S	0.188	15.624	9.34	1.3314	4.189	4.09	598	299,000	31.62
		10	0.250	15.500	12.37	1.3104	4.189	4.06	587	293,500	42.05
16	16	20	0.312	15.376	15.38	1.2985	4.189	4.03	578	289,000	52.36
		30, ST	0.375	15.250	18.41	1.2680	4.189	3.99	568	284,000	62.58
		40, XS	0.500	15.000	24.35	1.2272	4.189	3.93	550	275,000	82.77
		60	0.656	14.688	31.62	1.1766	4.189	3.85	528	264,000	107.54
		80	0.844	14.312	40.19	1.1171	4.189	3.75	501	250,500	136.58

100	1.031	13.938	48.48	1.0596	4.189	3.65	474	237,000	164.86
120	1.219	13.562	56.61	1.0032	4.189	3.55	450	225,000	192.40
140	1.438	13.124	65.79	0.9394	4.189	3.44	422	211,000	223.57
160	1.594	12.812	72.14	0.8953	4.189	3.35	402	201,000	245.22
5S	0.165	17.670	9.25	1.7029	4.712	4.63	74	382,000	31.32
10S	0.188	17.624	10.52	1.6941	4.712	4.61	760	379,400	35.48
10	0.250	17.500	13.94	1.6703	4.712	4.58	750	375,000	47.39
20	0.312	17.376	17.34	1.6468	4.712	4.55	739	369,500	59.03
ST	0.375	17.250	20.76	1.6230	4.712	4.52	728	364,000	70.59
30	0.438	17.124	24.16	1.5993	4.712	4.48	718	359,000	82.06
XS	0.500	17.000	27.49	1.5763	4.712	4.45	707	353,500	93.45
40	0.562	16.876	30.79	1.5533	4.712	4.42	697	348,500	104.76
60	0.750	16.500	40.64	1.4849	4.712	4.32	666	333,000	138.17
80	0.938	16.124	50.28	1.4180	4.712	4.22	636	318,000	170.75
100	1.156	15.688	61.17	1.3423	4.712	4.11	602	301,000	208.00
120	1.375	15.250	71.82	1.2684	4.712	3.99	569	284,500	244.14
140	1.562	14.876	80.66	1.2070	4.712	3.89	540	270,000	274.30
160	1.781	14.438	90.75	1.1370	4.712	3.78	510	255,000	308.55
5S	0.188	19.624	11.70	2.1004	5.236	5.14	943	471,500	39.76
10S	0.218	19.564	13.55	2.0878	5.236	5.12	937	467,500	45.98
10	0.250	19.500	15.51	2.0740	5.236	5.11	930	465,000	52.73
20, ST	0.375	19.250	23.12	2.0211	5.236	5.04	902	451,000	78.60
30, XS	0.500	19.000	30.63	1.9689	5.236	4.97	883	441,500	104.13
40	0.594	18.812	36.21	1.9302	5.236	4.92	866	433,000	123.06
60	0.812	18.376	48.95	1.8417	5.236	4.81	826	413,000	166.50
80	1.031	17.938	61.44	1.7550	5.236	4.70	787	393,500	208.92
100	1.281	17.438	75.33	1.6585	5.236	4.57	744	372,000	256.15
120	1.500	17.000	87.18	1.5763	5.236	4.45	707	353,500	296.37
140	1.750	16.500	100.3	1.4849	5.236	4.32	665	332,500	341.10
160	1.969	16.062	111.5	1.4071	5.236	4.21	632	316,000	379.14

(continued)

TABLE 51.85 (Continued)

Nominal Pipe Size (in.)	Outside Diameter (in.)	Schedule No.	Wall Thickness (in.)	Inside Diameter (in.)	Cross-Sectional Area		Circumference, ft, or surface, ft ² /ft of Length		Capacity at 1 ft/sec Velocity		Weight of Plain-End Pipe (lb/ft)
					Metal (in. ²)	Flow (ft ²)	Outside	Inside	U.S. gal/min	lb/hr water	
24	24	5S	0.218	23.564	16.29	3.0285	6.283	6.17	1359	679,500	55.08
		10, 10S	0.250	23.500	18.65	3.012	6.283	6.15	1350	675,000	63.41
		20, ST	0.375	23.250	27.83	2.948	6.283	6.09	1325	662,500	94.62
		XS	0.500	23.000	36.90	2.885	6.283	6.02	1295	642,500	125.49
		30	0.562	22.876	41.39	2.854	6.283	5.99	1281	640,500	140.80
		40	0.688	22.624	50.39	2.792	6.283	5.92	1253	626,500	171.17
		60	0.969	22.062	70.11	2.655	6.283	5.78	1192	596,000	238.29
		80	1.219	21.562	87.24	2.536	6.283	5.64	1138	569,000	296.53
		100	1.531	20.938	108.1	2.391	6.283	5.48	1073	536,500	367.45
		120	1.812	20.376	126.3	2.264	6.283	5.33	1016	508,000	429.50
		140	2.062	19.876	142.1	2.155	6.283	5.20	965	482,500	483.24
		160	2.344	19.312	159.5	2.034	6.283	5.06	913	456,500	542.09
30	30	5S	0.250	29.500	23.37	4.746	7.854	7.72	2130	1,065,000	79.43
		10, 10S	0.312	29.376	29.10	4.707	7.854	7.69	2110	1,055,000	99.08
		ST	0.375	29.250	34.90	4.666	7.854	7.66	2094	1,048,000	118.65
		20, XS	0.500	29.000	46.34	4.587	7.854	7.59	2055	1,027,500	157.53
		30	0.625	28.750	57.68	4.508	7.854	7.53	2020	1,010,000	196.08

Schedule Nos. 5S, 10S, and 40S American National Standards Institute (ANSI)/American Society of Mechanical Engineers (ASME) B.36.19-1985, "Stainless Steel Pipe." ST = standard wall, XS = extra strong wall, XX = double extra strong wall are all taken from ANSI/ASME, B.36.10M-1985, "Welded and Seamless Wrought-steel Pipe." Wrought-iron pipe has slightly thicker walls, approximately 3%, but the same weight per foot, because of lower density. Decimal thicknesses for respective pipe sizes represent their nominal or average wall dimensions. Mill tolerances as high as 12½% are permitted.

Plain-end pipe is produced by a square cut. Pipe is also shipped from the mills threaded, with a threaded coupling on one end, or with the ends beveled for welding, or grooved or sized for patented couplings. Weights per foot for threaded and coupled pipe are slightly greater because of the weight of the coupling, but it is not available larger than 12 in., or lighter than Schedule 30 sizes 8 through 12 in., or Schedule 40 6 in. and smaller.

Source: From *Chemical Engineer's Handbook*, 4th ed., New York, McGraw-Hill, 1963. Used by permission.

TABLE 51.86 Properties and Dimensions of Steel Pipe^a

Nominal Diameter (in.)	Dimensions				Weight per Foot (lb)			Couplings			Properties		
	Outside Diameter (in.)	Inside Diameter (in.)	Thickness (in.)	Thread and Coupling		Threads per Inch	Outside Diameter (in.)	Length (in.)	Weight (lb)	I (in. ⁴)	A (in. ²)	k (in.)	
				Plain Ends	Thread and Coupling								
Schedule 40ST													
$\frac{1}{8}$	0.405	0.269	0.068	0.24	0.25	27	0.562	$\frac{7}{8}$	0.03	0.001	0.072	0.12	
$\frac{1}{4}$	0.540	0.364	0.088	0.42	0.43	18	0.685	1	0.04	0.003	0.125	0.16	
$\frac{3}{8}$	0.675	0.493	0.091	0.57	0.57	18	0.848	$1\frac{1}{8}$	0.07	0.007	0.167	0.21	
$\frac{1}{2}$	0.840	0.622	0.109	0.85	0.85	14	1.024	$1\frac{3}{8}$	0.12	0.017	0.250	0.26	
$\frac{3}{4}$	1.050	0.824	0.113	1.13	1.13	14	1.281	$1\frac{5}{8}$	0.21	0.037	0.333	0.33	
1	1.315	1.049	0.133	1.68	1.68	$11\frac{1}{2}$	1.576	$1\frac{7}{8}$	0.35	0.087	0.494	0.42	
$1\frac{1}{4}$	1.660	1.380	0.140	2.27	2.28	$11\frac{1}{2}$	1.950	$2\frac{1}{8}$	0.55	0.195	0.669	0.54	
$1\frac{1}{2}$	1.900	1.610	0.145	2.72	2.73	$11\frac{1}{2}$	2.218	$2\frac{3}{8}$	0.76	0.310	0.799	0.62	
2	2.375	2.067	0.154	3.65	3.68	$11\frac{1}{2}$	2.760	$2\frac{5}{8}$	1.23	0.666	1.075	0.79	
$2\frac{1}{2}$	2.875	2.469	0.203	5.79	5.82	8	3.276	$2\frac{7}{8}$	1.76	1.530	1.704	0.95	
3	3.500	3.068	0.216	7.58	7.62	8	3.948	$3\frac{1}{8}$	2.55	3.017	2.228	1.16	
$3\frac{1}{2}$	4.000	3.548	0.226	9.11	9.20	8	4.591	$3\frac{5}{8}$	4.33	4.788	2.680	1.34	
4	4.500	4.026	0.237	10.79	10.89	8	5.091	$3\frac{7}{8}$	5.41	7.233	3.174	1.51	
5	5.563	5.047	0.258	14.62	14.81	8	6.296	$4\frac{1}{8}$	9.16	15.16	4.300	1.88	
6	6.625	6.065	0.280	18.97	19.19	8	7.358	$4\frac{3}{8}$	10.82	28.14	5.581	2.25	
8	8.625	8.071	0.277	24.70	25.00	8	9.420	$4\frac{5}{8}$	15.84	63.35	7.265	2.95	
8	8.625	7.981	0.322	28.55	28.81	8	9.420	$4\frac{5}{8}$	15.84	72.49	8.399	2.94	
10	10.750	10.192	0.279	31.20	32.00	8	11.721	$6\frac{1}{8}$	33.92	125.4	9.178	3.70	
10	10.750	10.136	0.307	34.24	35.00	8	11.721	$6\frac{1}{8}$	33.92	137.4	10.07	3.69	

(continued)

TABLE 51.86 (Continued)

Dimensions					Weight per Foot (lb)		Couplings			Properties			
Nominal Diameter (in.)	Outside Diameter (in.)	Inside Diameter (in.)	Thickness (in.)	Plain Ends	Thread and Coupling		Threads per Inch	Outside Diameter (in.)	Length (in.)	Weight (lb)	I (in. ⁴)	A (in. ²)	k (in.)
10	10.750	10.020	0.365	40.48	41.13		8	11.721	6 $\frac{1}{8}$	33.92	160.7	11.91	3.67
12	12.750	12.090	0.330	43.77	45.00		8	13.958	6 $\frac{1}{8}$	48.27	248.5	12.88	4.39
12	12.750	12.000	0.375	49.56	50.71		8	13.958	6 $\frac{1}{8}$	48.27	279.3	14.38	4.38
Schedule 80XS													
$\frac{1}{8}$	0.405	0.215	0.095	0.31	0.32		27	0.582	1 $\frac{1}{8}$	0.05	0.001	0.093	0.12
$\frac{1}{4}$	0.540	0.302	0.119	0.54	0.54		18	0.724	1 $\frac{3}{8}$	0.07	0.004	0.157	0.16
$\frac{3}{8}$	0.675	0.423	0.126	0.74	0.75		18	0.898	1 $\frac{5}{8}$	0.13	0.009	0.217	0.20
$\frac{1}{2}$	0.840	0.546	0.147	1.09	1.10		14	1.085	1 $\frac{7}{8}$	0.22	0.020	0.320	0.25
$\frac{3}{4}$	1.050	0.742	0.154	1.47	1.49		14	1.316	2 $\frac{1}{8}$	0.33	0.045	0.433	0.32
1	1.315	0.957	0.179	2.17	2.20		11 $\frac{1}{2}$	1.575	2 $\frac{3}{8}$	0.47	0.106	0.639	0.41
1 $\frac{1}{4}$	1.660	1.278	0.191	3.00	3.05		11 $\frac{1}{2}$	2.054	2 $\frac{7}{8}$	1.04	0.242	0.881	0.52
1 $\frac{1}{2}$	1.900	1.500	0.200	3.63	3.69		11 $\frac{1}{2}$	2.294	2 $\frac{7}{8}$	1.17	0.391	1.068	0.61
2	2.375	1.939	0.218	5.02	5.13		11 $\frac{1}{2}$	2.870	3 $\frac{5}{8}$	2.17	0.868	1.477	0.77
2 $\frac{1}{2}$	2.875	2.323	0.276	7.66	7.83		8	3.389	4 $\frac{1}{8}$	3.43	1.924	2.254	0.92
3	3.500	2.900	0.300	10.25	10.46		8	4.014	4 $\frac{1}{8}$	4.13	3.894	3.016	1.14
3 $\frac{1}{2}$	4.000	3.364	0.318	12.51	12.82		8	4.628	4 $\frac{5}{8}$	6.29	6.280	3.678	1.31
4	4.500	3.826	0.337	14.98	15.39		8	5.233	4 $\frac{5}{8}$	8.16	9.610	4.407	1.48
5	5.563	4.813	0.375	20.78	21.42		8	6.420	5 $\frac{1}{8}$	12.87	20.67	6.112	1.84
6	6.625	5.761	0.432	28.57	29.33		8	7.482	5 $\frac{1}{8}$	15.18	40.49	8.405	2.20
8	8.625	7.625	0.500	43.39	44.72		8	9.596	6 $\frac{1}{8}$	26.63	105.7	12.76	2.88
10	10.750	9.750	0.500	54.74	56.94		8	11.958	6 $\frac{5}{8}$	44.16	211.9	16.10	3.63
12	12.750	11.750	0.500	65.42	68.02		8	13.958	6 $\frac{5}{8}$	51.99	361.5	19.24	4.34

Schedule XX												
$\frac{1}{2}$	0.840	0.252	0.294	1.71	1.73	14	1.085	$1\frac{7}{8}$	0.22	0.024	0.504	0.22
$\frac{3}{4}$	1.050	0.434	0.308	2.44	2.46	14	1.316	$2\frac{1}{8}$	0.33	0.058	0.718	0.28
1	1.315	0.599	0.358	3.66	3.68	$11\frac{1}{2}$	1.575	$2\frac{3}{8}$	0.47	0.140	1.076	0.36
$1\frac{1}{4}$	1.660	0.896	0.382	5.21	5.27	$11\frac{1}{2}$	2.054	$2\frac{7}{8}$	1.04	0.341	1.534	0.47
$1\frac{1}{2}$	1.900	1.100	0.400	6.41	6.47	$11\frac{1}{2}$	2.294	$2\frac{7}{8}$	1.17	0.568	1.885	0.55
2	2.375	1.503	0.436	9.03	9.14	$11\frac{1}{2}$	2.870	$3\frac{5}{8}$	2.17	1.311	2.656	0.70
$2\frac{1}{2}$	2.875	1.771	0.552	13.70	13.87	8	3.389	$4\frac{1}{8}$	3.43	2.871	4.028	0.84
3	3.500	2.300	0.600	18.58	18.79	8	4.014	$4\frac{1}{8}$	4.13	5.992	5.466	1.05
$3\frac{1}{2}$	4.000	2.728	0.636	22.85	23.16	8	4.628	$4\frac{5}{8}$	6.29	9.848	6.721	1.21
4	4.500	3.152	0.674	27.54	27.95	8	5.233	$4\frac{5}{8}$	8.16	15.28	8.101	1.37
5	5.563	4.063	0.750	38.55	39.20	8	6.420	$5\frac{1}{8}$	12.87	33.64	11.34	1.72
6	6.625	4.897	0.864	53.16	53.92	8	7.482	$5\frac{1}{8}$	15.18	66.33	15.64	2.06
8	8.625	6.875	0.875	72.42	73.76	8	9.596	$6\frac{1}{8}$	26.63	162.0	21.30	2.76
Large Outside Diameter Pipe												

Pipe 14 in. and larger is sold by actual outside step diameter and thickness.

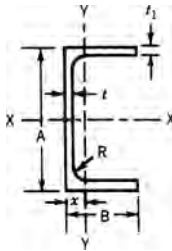
Sizes 14, 15, and 16 in. are available regularly in thicknesses varying by $\frac{1}{16}$ in. from $\frac{1}{4}$ to 1 in., inclusive.

All pipe is furnished random length unless otherwise ordered, viz: 12-22 ft with privilege of furnishing 5 % in 6-12-ft lengths. Pipe railing is most economically detailed with slip joints and random lengths between couplings.

^aSteel Construction, 1980, A.I.S.C.

51.6.6 Standard Structural Shapes—Aluminum¹²

TABLE 51.87 Aluminum Association Standard Channels—Dimensions, Areas, Weights, and Section Properties^a



Size		Section Properties ^d												
Depth A (in.)	Width B (in.)	Area ^b (in. ²)	Weight ^c (lb/ft)	Fange Thickness t ₁ (in.)	Web Thickness t (in.)	Fillet Radius R (in.)	Axis X-X				Axis Y-Y			
							I (in. ⁴)	S (in. ³)	r (in.)	I (in. ⁴)	S (in. ³)	r (in.)	x (in.)	
2.00	1.00	0.491	0.557	0.13	0.13	0.10	0.288	0.288	0.766	0.045	0.064	0.303	0.298	
2.00	1.25	0.911	1.071	0.26	0.17	0.15	0.546	0.546	0.774	0.139	0.178	0.391	0.471	
3.00	1.50	0.965	1.135	0.20	0.13	0.25	1.41	0.94	1.21	0.22	0.22	0.47	0.49	
3.00	1.75	1.358	1.597	0.26	0.17	0.25	1.97	1.31	1.20	0.42	0.37	0.55	0.62	
4.00	2.00	1.478	1.738	0.23	0.15	0.25	3.91	1.95	1.63	0.60	0.45	0.64	0.65	
4.00	2.25	1.982	2.331	0.29	0.19	0.25	5.21	2.60	1.62	1.02	0.69	0.72	0.78	
5.00	2.25	1.881	2.212	0.26	0.15	0.30	7.88	3.15	2.05	0.98	0.64	0.72	0.73	
5.00	2.75	2.627	3.089	0.32	0.19	0.30	11.14	4.45	2.06	2.05	1.14	0.88	0.95	
6.00	2.50	2.410	2.834	0.29	0.17	0.30	14.35	4.78	2.44	1.53	0.90	0.80	0.79	
6.00	3.25	3.427	4.030	0.35	0.21	0.30	21.04	7.01	2.48	3.76	1.76	1.05	1.12	
7.00	2.75	2.725	3.205	0.29	0.17	0.30	22.09	6.31	2.85	2.10	1.10	0.88	0.84	
7.00	3.50	4.009	4.715	0.38	0.21	0.30	33.79	9.65	2.90	5.13	2.23	1.13	1.20	
8.00	3.00	3.526	4.147	0.35	0.19	0.30	37.40	9.35	3.26	3.25	1.57	0.96	0.93	
8.00	3.75	4.923	5.789	0.41	0.25	0.35	52.69	13.17	3.27	7.13	2.82	1.20	1.22	
9.00	3.25	4.237	4.983	0.35	0.23	0.35	54.41	12.09	3.58	4.40	1.89	1.02	0.93	
9.00	4.00	5.927	6.970	0.44	0.29	0.35	78.31	17.40	3.63	9.61	3.49	1.27	1.25	
10.00	3.50	5.218	6.136	0.41	0.25	0.35	83.22	16.64	3.99	6.33	2.56	1.10	1.02	
10.00	4.25	7.109	8.360	0.50	0.31	0.40	116.15	23.23	4.04	13.02	4.47	1.35	1.34	
12.00	4.00	7.036	8.274	0.47	0.29	0.40	159.76	26.63	4.77	11.03	3.86	1.25	1.14	
12.00	5.00	10.053	11.822	0.62	0.35	0.45	239.69	39.95	4.88	25.74	7.60	1.60	1.61	

^aUsers are encouraged to ascertain current availability of particular structural shapes through inquiries to their suppliers.

^bAreas listed are based on nominal dimensions.

^cWeights per foot are based on nominal dimensions and a density of 0.098 lb/in.³, which is the density of alloy 6061.

^dI = moment of inertia; S = section modulus; r = radius of gyration.

¹²Tables 51.87–51.101 are from *Aluminum Standards and Data*. Copyright © 1984 The Aluminum Association.

TABLE 51.88 Aluminum Association Standard I Beams—Dimensions, Areas, Weights, and Section Properties^a

Size		Section Properties ^d									
Depth	Width	Flange		Web		Fillet		Axis X-X		Axis Y-Y	
A (in.)	B (in.)	Area ^b (in. ²)	Weight ^c (lb/ft)	Thickness t ₁ (in.)	Thickness t (in.)	Radius R (in.)		I (in. ⁴)	S (in. ³)	I (in. ⁴)	S (in. ³)
3.00	2.50	1.392	1.637	0.20	0.13	0.25		2.24	1.49	0.52	0.42
3.00	2.50	1.726	2.030	0.26	0.15	0.25		2.71	1.81	0.68	0.54
4.00	3.00	1.965	2.311	0.23	0.15	0.25		5.62	2.81	1.04	0.69
4.00	3.00	2.375	2.793	0.29	0.17	0.25		6.71	3.36	1.31	0.87
5.00	3.50	3.146	3.700	0.32	0.19	0.30		13.94	5.58	2.11	1.31
6.00	4.00	3.427	4.030	0.29	0.19	0.30		21.99	7.33	2.53	1.55
6.00	4.00	3.990	4.692	0.35	0.21	0.30		25.50	8.50	2.53	1.87
7.00	4.50	4.932	5.800	0.38	0.23	0.30		42.89	12.25	2.95	2.57
8.00	5.00	5.256	6.181	0.35	0.23	0.30		59.69	14.92	3.37	2.92
8.00	5.00	5.972	7.023	0.41	0.25	0.30		67.78	16.94	3.37	3.42
9.00	5.50	7.110	8.361	0.44	0.27	0.30		102.02	22.67	3.79	4.44
10.00	6.00	7.352	8.646	0.41	0.25	0.40		132.09	26.42	4.24	4.93
10.00	6.00	8.747	10.286	0.50	0.29	0.40		155.79	31.16	4.22	6.01
12.00	7.00	9.925	11.672	0.47	0.29	0.40		255.57	42.60	5.07	7.69
12.00	7.00	12.153	14.292	0.62	0.31	0.40		317.33	52.89	5.11	10.14

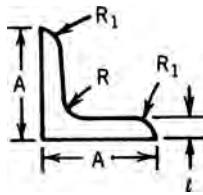
^aUsers are encouraged to ascertain current availability of particular structural shapes through inquiries to their suppliers.

^bAreas listed are based on nominal dimensions.

^cWeights per foot are based on nominal dimensions and a density of 0.098 lb/in.³, which is the density of alloy 6061.

^dI = moment of inertia; S = section modulus; r = radius of gyration.

TABLE 51.89 Standard Structural Shapes—Equal Angles^a



A	t	R	R ₁	Area ^b (in. ²)	Weight per Foot ^c (lb)
$\frac{3}{4}$	$\frac{1}{8}$	$\frac{1}{8}$	$\frac{3}{32}$	0.171	0.201
$\frac{3}{4}$	$\frac{3}{16}$	$\frac{1}{8}$	$\frac{3}{32}$	0.246	0.289
1	$\frac{3}{32}$	$\frac{1}{8}$	$\frac{3}{32}$	0.179	0.211
1	$\frac{1}{8}$	$\frac{1}{8}$	$\frac{3}{32}$	0.234	0.275
1	$\frac{3}{16}$	$\frac{1}{8}$	$\frac{3}{32}$	0.340	0.400
1	$\frac{1}{4}$	$\frac{1}{8}$	$\frac{3}{32}$	0.437	0.514
$1\frac{1}{4}$	$\frac{1}{8}$	$\frac{3}{16}$	$\frac{1}{8}$	0.292	0.343
$1\frac{1}{4}$	$\frac{3}{16}$	$\frac{3}{16}$	$\frac{1}{8}$	0.434	0.510
$1\frac{1}{4}$	$\frac{1}{4}$	$\frac{3}{16}$	$\frac{1}{8}$	0.558	0.656
$1\frac{1}{2}$	$\frac{1}{8}$	$\frac{3}{16}$	$\frac{1}{8}$	0.360	0.423
$1\frac{1}{2}$	$\frac{3}{16}$	$\frac{3}{16}$	$\frac{1}{8}$	0.529	0.619
$1\frac{1}{2}$	$\frac{1}{4}$	$\frac{3}{16}$	$\frac{1}{8}$	0.688	0.809
$1\frac{3}{4}$	$\frac{1}{8}$	$\frac{3}{16}$	$\frac{1}{8}$	0.423	0.497
$1\frac{3}{4}$	$\frac{3}{16}$	$\frac{3}{16}$	$\frac{1}{8}$	0.622	0.731
$1\frac{3}{4}$	$\frac{1}{4}$	$\frac{3}{16}$	$\frac{1}{8}$	0.813	0.956
$1\frac{3}{4}$	$\frac{5}{16}$	$\frac{3}{16}$	$\frac{1}{8}$	0.996	1.171
2	$\frac{1}{8}$	$\frac{1}{4}$	$\frac{1}{8}$	0.491	0.577
2	$\frac{3}{16}$	$\frac{1}{4}$	$\frac{1}{8}$	0.723	0.850
2	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{8}$	0.944	1.110
2	$\frac{5}{16}$	$\frac{1}{4}$	$\frac{1}{8}$	1.160	1.364
2	$\frac{3}{8}$	$\frac{1}{4}$	$\frac{1}{8}$	1.366	1.606
$2\frac{1}{2}$	$\frac{1}{8}$	$\frac{1}{4}$	$\frac{1}{8}$	0.616	0.724
$2\frac{1}{2}$	$\frac{3}{16}$	$\frac{1}{4}$	$\frac{1}{8}$	0.910	1.070
$2\frac{1}{2}$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{8}$	1.194	1.404
$2\frac{1}{2}$	$\frac{5}{16}$	$\frac{1}{4}$	$\frac{1}{8}$	1.470	1.729
$2\frac{1}{2}$	$\frac{3}{8}$	$\frac{1}{4}$	$\frac{1}{8}$	1.714	2.047
3	$\frac{3}{16}$	$\frac{5}{16}$	$\frac{1}{4}$	1.084	1.275
3	$\frac{1}{4}$	$\frac{5}{16}$	$\frac{1}{4}$	1.432	1.684
3	$\frac{5}{16}$	$\frac{5}{16}$	$\frac{1}{4}$	1.770	2.082
3	$\frac{3}{8}$	$\frac{5}{16}$	$\frac{1}{4}$	2.104	2.474
3	$\frac{7}{16}$	$\frac{5}{16}$	$\frac{1}{4}$	2.428	2.855
3	$\frac{1}{2}$	$\frac{5}{16}$	$\frac{1}{4}$	2.744	3.227
$3\frac{1}{2}$	$\frac{1}{4}$	$\frac{3}{8}$	$\frac{1}{4}$	1.691	1.989
$3\frac{1}{2}$	$\frac{5}{16}$	$\frac{3}{8}$	$\frac{1}{4}$	2.093	2.461
$3\frac{1}{2}$	$\frac{3}{8}$	$\frac{3}{8}$	$\frac{1}{4}$	2.488	2.926
$3\frac{1}{2}$	$\frac{1}{2}$	$\frac{3}{8}$	$\frac{1}{4}$	3.253	3.826

TABLE 51.89 *(Continued)*

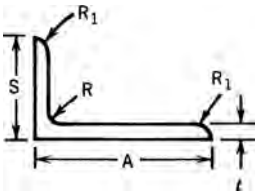
<i>A</i>	<i>t</i>	<i>R</i>	<i>R</i> ₁	Area ^{<i>b</i>} (in. ²)	Weight per Foot ^{<i>c</i>} (lb)
4	$\frac{1}{4}$	$\frac{3}{8}$	$\frac{1}{4}$	1.941	2.283
4	$\frac{5}{16}$	$\frac{3}{8}$	$\frac{1}{4}$	2.406	2.829
4	$\frac{3}{8}$	$\frac{3}{8}$	$\frac{1}{4}$	2.862	3.366
4	$\frac{7}{16}$	$\frac{3}{8}$	$\frac{1}{4}$	3.310	3.893
4	$\frac{1}{2}$	$\frac{3}{8}$	$\frac{1}{4}$	3.753	4.414
4	$\frac{9}{16}$	$\frac{3}{8}$	$\frac{1}{4}$	4.187	4.924
4	$\frac{5}{8}$	$\frac{3}{8}$	$\frac{1}{4}$	4.613	5.425
4	$\frac{11}{16}$	$\frac{3}{8}$	$\frac{1}{4}$	5.032	5.918
4	$\frac{3}{4}$	$\frac{3}{8}$	$\frac{1}{4}$	5.441	6.399
5	$\frac{3}{8}$	$\frac{1}{2}$	$\frac{3}{8}$	3.603	4.237
5	$\frac{7}{16}$	$\frac{1}{2}$	$\frac{3}{8}$	4.177	4.912
5	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{3}{8}$	4.743	5.578
5	$\frac{5}{8}$	$\frac{1}{2}$	$\frac{3}{8}$	5.853	6.883
6	$\frac{3}{8}$	$\frac{1}{2}$	$\frac{3}{8}$	4.353	5.119
6	$\frac{7}{16}$	$\frac{1}{2}$	$\frac{3}{8}$	5.052	5.941
6	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{3}{8}$	5.743	6.754
6	$\frac{5}{8}$	$\frac{1}{2}$	$\frac{3}{8}$	7.102	8.352
8	$\frac{1}{2}$	$\frac{5}{8}$	$\frac{3}{8}$	7.773	9.141
8	$\frac{3}{4}$	$\frac{5}{8}$	$\frac{3}{8}$	11.461	13.478
8	1	$\frac{5}{8}$	$\frac{3}{8}$	15.023	17.667

^{*a*}Users are encouraged to ascertain current availability of particular structural shapes through inquiries to their suppliers.

^{*b*}Areas listed are based on nominal dimensions.

^{*c*}Weights per foot are based on nominal dimensions and a density of 0.098 lb/in.³, which is the density of alloy 6061.

TABLE 51.90 **Standard Structural Shapes—Unequal Angles^{*a*}**



<i>A</i>	<i>B</i>	<i>t</i>	<i>R</i>	<i>R</i> ₁	Area ^{<i>b</i>} (in. ²)	Weight per Foot ^{<i>c</i>} (lb)
$1\frac{1}{4}$	$\frac{3}{4}$	$\frac{3}{32}$	$\frac{3}{32}$	$\frac{3}{64}$	0.180	0.212
$1\frac{1}{4}$	1	$\frac{1}{8}$	$\frac{1}{8}$	$\frac{1}{16}$	0.267	0.314
$1\frac{1}{2}$	$\frac{3}{4}$	$\frac{1}{8}$	$\frac{1}{8}$	$\frac{1}{16}$	0.267	0.314
$1\frac{1}{2}$	$\frac{5}{4}$	$\frac{3}{16}$	$\frac{1}{8}$	$\frac{3}{32}$	0.386	0.454
$1\frac{1}{2}$	1	$\frac{5}{32}$	$\frac{5}{32}$	$\frac{5}{64}$	0.368	0.433
$1\frac{1}{2}$	1	$\frac{1}{4}$	$\frac{3}{16}$	$\frac{1}{8}$	0.563	0.662
$1\frac{1}{2}$	$1\frac{1}{4}$	$\frac{1}{8}$	$\frac{3}{16}$	$\frac{1}{8}$	0.329	0.387

(continued)

TABLE 51.90 (Continued)

A	B	t	R	R_1	Area ^b (in. ²)	Weight per Foot ^c (lb)
$1\frac{1}{2}$	$1\frac{1}{4}$	$\frac{3}{16}$	$\frac{3}{16}$	$\frac{1}{8}$	0.481	0.566
$1\frac{1}{2}$	$1\frac{1}{4}$	$\frac{1}{4}$	$\frac{3}{16}$	$\frac{1}{8}$	0.624	0.734
$1\frac{3}{4}$	$1\frac{1}{4}$	$\frac{1}{8}$	$\frac{3}{16}$	$\frac{1}{8}$	0.358	0.421
$1\frac{3}{4}$	$1\frac{1}{4}$	$\frac{3}{16}$	$\frac{3}{16}$	$\frac{1}{8}$	0.528	0.621
$1\frac{3}{4}$	$1\frac{1}{4}$	$\frac{1}{4}$	$\frac{3}{16}$	$\frac{1}{8}$	0.688	0.809
2	$1\frac{1}{2}$	$\frac{1}{8}$	$\frac{3}{16}$	$\frac{1}{8}$	0.422	0.496
2	$1\frac{1}{2}$	$\frac{3}{16}$	$\frac{3}{16}$	$\frac{1}{8}$	0.622	0.731
2	$1\frac{1}{2}$	$\frac{1}{4}$	$\frac{3}{16}$	$\frac{1}{8}$	0.813	0.956
2	$1\frac{1}{2}$	$\frac{3}{8}$	$\frac{3}{16}$	$\frac{1}{8}$	1.172	1.378
$2\frac{1}{2}$	$1\frac{1}{2}$	$\frac{3}{16}$	$\frac{1}{4}$	$\frac{1}{8}$	0.723	0.850
$2\frac{1}{2}$	$1\frac{1}{2}$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{8}$	0.944	1.110
$2\frac{1}{2}$	$1\frac{1}{2}$	$\frac{5}{16}$	$\frac{3}{16}$	$\frac{1}{8}$	1.152	1.355
$2\frac{1}{2}$	2	$\frac{1}{8}$	$\frac{1}{4}$	$\frac{1}{8}$	0.554	0.652
$2\frac{1}{2}$	2	$\frac{3}{16}$	$\frac{1}{4}$	$\frac{1}{8}$	0.817	0.961
$2\frac{1}{2}$	2	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{8}$	1.069	1.257
$2\frac{1}{2}$	2	$\frac{5}{16}$	$\frac{1}{4}$	$\frac{1}{8}$	1.314	1.545
$2\frac{1}{2}$	2	$\frac{3}{8}$	$\frac{1}{4}$	$\frac{1}{8}$	1.554	1.828
3	2	$\frac{3}{16}$	$\frac{5}{16}$	$\frac{3}{16}$	0.911	1.071
3	2	$\frac{1}{4}$	$\frac{5}{16}$	$\frac{3}{16}$	1.193	1.403
3	2	$\frac{5}{16}$	$\frac{5}{16}$	$\frac{3}{16}$	1.471	1.730
3	2	$\frac{3}{8}$	$\frac{5}{16}$	$\frac{3}{16}$	1.740	2.046
3	2	$\frac{7}{16}$	$\frac{5}{16}$	$\frac{3}{16}$	2.001	2.353
3	$2\frac{1}{2}$	$\frac{1}{4}$	$\frac{5}{16}$	$\frac{1}{4}$	1.307	1.537
3	$2\frac{1}{2}$	$\frac{5}{16}$	$\frac{5}{16}$	$\frac{1}{4}$	1.614	1.898
3	$2\frac{1}{2}$	$\frac{3}{8}$	$\frac{5}{16}$	$\frac{1}{4}$	1.916	2.253
$3\frac{1}{2}$	$2\frac{1}{2}$	$\frac{1}{4}$	$\frac{5}{16}$	$\frac{1}{4}$	1.432	1.684
$3\frac{1}{2}$	$2\frac{1}{2}$	$\frac{5}{16}$	$\frac{5}{16}$	$\frac{1}{4}$	1.770	2.082
$3\frac{1}{2}$	$2\frac{1}{2}$	$\frac{3}{8}$	$\frac{5}{16}$	$\frac{1}{4}$	2.104	2.474
$3\frac{1}{2}$	$2\frac{1}{2}$	$\frac{1}{2}$	$\frac{5}{16}$	$\frac{1}{4}$	2.744	3.227
$3\frac{1}{2}$	3	$\frac{1}{4}$	$\frac{3}{8}$	$\frac{1}{4}$	1.566	1.842
$3\frac{1}{2}$	3	$\frac{5}{16}$	$\frac{3}{8}$	$\frac{1}{4}$	1.937	2.278
$3\frac{1}{2}$	3	$\frac{3}{8}$	$\frac{3}{8}$	$\frac{1}{4}$	2.300	2.705
$3\frac{1}{2}$	3	$\frac{1}{2}$	$\frac{3}{8}$	$\frac{1}{4}$	3.003	3.532
4	3	$\frac{1}{4}$	$\frac{3}{8}$	$\frac{1}{4}$	1.691	1.988
4	3	$\frac{5}{16}$	$\frac{3}{8}$	$\frac{1}{4}$	2.091	2.459
4	3	$\frac{3}{8}$	$\frac{3}{8}$	$\frac{1}{4}$	2.488	2.926
4	3	$\frac{7}{16}$	$\frac{3}{8}$	$\frac{1}{4}$	2.874	3.380
4	3	$\frac{1}{2}$	$\frac{3}{8}$	$\frac{1}{4}$	3.253	3.826
4	3	$\frac{5}{8}$	$\frac{3}{8}$	$\frac{1}{4}$	3.988	4.690
4	$3\frac{1}{2}$	$\frac{3}{8}$	$\frac{3}{8}$	$\frac{5}{16}$	2.660	3.128
4	$3\frac{1}{2}$	$\frac{1}{2}$	$\frac{3}{8}$	$\frac{5}{16}$	3.488	4.102
5	3	$\frac{3}{8}$	$\frac{3}{8}$	$\frac{5}{16}$	2.848	3.349

TABLE 51.90 (Continued)

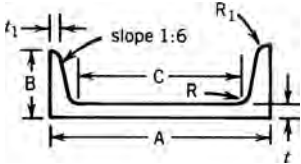
A	B	t	R	R ₁	Area ^b (in. ²)	Weight per Foot ^c (lb)
5	3	$\frac{1}{2}$	$\frac{3}{8}$	$\frac{5}{16}$	3.738	4.396
5	$3\frac{1}{2}$	$\frac{5}{16}$	$\frac{7}{16}$	$\frac{5}{16}$	2.558	3.008
5	$3\frac{1}{2}$	$\frac{3}{8}$	$\frac{7}{16}$	$\frac{5}{16}$	3.046	3.582
5	$3\frac{1}{2}$	$\frac{7}{16}$	$\frac{7}{16}$	$\frac{5}{16}$	3.527	4.148
5	$3\frac{1}{2}$	$\frac{1}{2}$	$\frac{7}{16}$	$\frac{5}{16}$	4.000	4.704
5	$3\frac{1}{2}$	$\frac{5}{8}$	$\frac{7}{16}$	$\frac{5}{16}$	4.921	5.787
6	$3\frac{1}{2}$	$\frac{5}{16}$	$\frac{1}{2}$	$\frac{5}{16}$	2.878	3.385
6	$3\frac{1}{2}$	$\frac{3}{8}$	$\frac{1}{2}$	$\frac{5}{16}$	3.433	4.037
6	$3\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{5}{16}$	4.512	5.306
6	4	$\frac{3}{8}$	$\frac{1}{2}$	$\frac{3}{8}$	3.603	4.237
6	4	$\frac{7}{16}$	$\frac{1}{2}$	$\frac{3}{8}$	4.179	4.915
6	4	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{3}{8}$	4.743	5.578
6	4	$\frac{9}{16}$	$\frac{1}{2}$	$\frac{3}{8}$	5.298	6.230
6	4	$\frac{5}{8}$	$\frac{1}{2}$	$\frac{3}{8}$	5.853	6.883
6	4	$\frac{3}{4}$	$\frac{1}{2}$	$\frac{3}{8}$	6.931	8.151
8	6	$\frac{5}{8}$	$\frac{1}{2}$	$\frac{5}{16}$	8.371	9.844
8	6	$\frac{11}{16}$	$\frac{1}{2}$	$\frac{3}{8}$	9.152	10.763
8	6	$\frac{3}{4}$	$\frac{1}{2}$	$\frac{3}{8}$	9.931	11.679

^aUsers are encouraged to ascertain current availability of particular structural shapes through inquiries to their suppliers.

^bAreas listed are based on nominal dimensions.

^cWeights per foot are based on nominal dimensions and a density of 0.098lb/in.³, which is the density of alloy 6061.

TABLE 51.91 Channels, American Standard^a

								
A	B	C	t	t ₁	R	R ₁	Area ^b (in. ²)	Weight per Foot ^c (lb)
3	1.410	$1\frac{3}{4}$	0.170	0.170	0.270	0.100	1.205	1.417
3	1.498	$1\frac{3}{4}$	0.258	0.170	0.270	0.100	1.470	1.729
3	1.596	$1\frac{3}{4}$	0.356	0.170	0.270	0.100	1.764	2.074
4	1.580	$2\frac{3}{4}$	0.180	0.180	0.280	0.110	1.570	1.846
4	1.647	$2\frac{3}{4}$	0.247	0.180	0.280	0.110	1.838	2.161
4	1.720	$2\frac{3}{4}$	0.320	0.180	0.280	0.110	2.129	2.504
5	1.750	$3\frac{3}{4}$	0.190	0.190	0.290	0.110	1.969	2.316
5	1.885	$3\frac{3}{4}$	0.325	0.190	0.290	0.110	2.643	3.108
5	2.032	$3\frac{3}{4}$	0.472	0.190	0.290	0.110	3.380	3.975
6	1.920	$4\frac{1}{2}$	0.200	0.200	0.300	0.120	2.403	2.826
6	1.945	$4\frac{1}{2}$	0.225	0.200	0.300	0.120	2.553	3.002

(continued)

TABLE 51.91 (Continued)

A	B	C	t	t ₁	R	R ₁	Area ^b (in. ²)	Weight per Foot ^c (lb)
6	2.034	4½	0.314	0.200	0.300	0.120	3.088	3.631
6	2.157	4½	0.437	0.200	0.300	0.120	3.825	4.498
7	2.110	5½	0.230	0.210	0.310	0.130	3.011	3.541
7	2.194	5½	0.314	0.210	0.310	0.130	3.599	4.232
7	2.299	5½	0.419	0.210	0.310	0.130	4.334	5.097
8	2.290	6¼	0.250	0.220	0.320	0.130	3.616	4.252
8	2.343	6¼	0.303	0.220	0.320	0.130	4.040	4.751
8	2.435	6¼	0.395	0.220	0.320	0.130	4.776	5.617
8	2.527	6¼	0.487	0.220	0.320	0.130	5.514	6.484
9	2.430	7¼	0.230	0.230	0.330	0.140	3.915	4.604
9	2.648	7¼	0.448	0.230	0.330	0.140	5.877	6.911
10	2.600	8¼	0.240	0.240	0.340	0.140	4.488	5.278
10	2.886	8¼	0.526	0.240	0.340	0.140	7.348	8.641
12	2.960	10	0.300	0.280	0.380	0.170	6.302	7.411
12	3.047	10	0.387	0.280	0.380	0.170	7.346	8.639
12	3.170	10	0.510	0.280	0.380	0.170	8.822	10.374
15	3.400	12⅜	0.400	0.400	0.500	0.240	9.956	11.708
15	3.716	12⅜	0.716	0.400	0.500	0.240	14.696	17.282

^aUsers are encouraged to ascertain current availability of particular structural shapes through inquiries to their suppliers.

^bAreas listed are based on nominal dimensions.

^cWeights per foot are based on nominal dimensions and a density of 0.098 lb/in.³, which is the density of alloy 6061.

TABLE 51.92 Channels, Shipbuilding, and Carbuilding^a

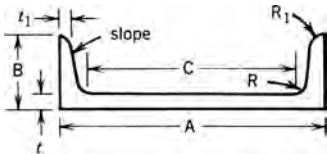
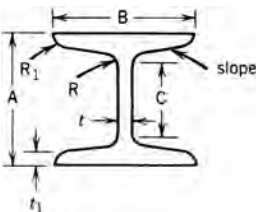
									
A	B	C	t	t ₁	R	R ₁	Slope	Area ^b (in. ²)	Weight per Foot ^c (lb)
3	2	1¾	0.250	0.250	0.250	0	12:12.1	1.900	2.234
3	2	1⅞	0.375	0.375	0.188	0.375	0	2.298	2.702
4	2½	2⅜	0.318	0.313	0.375	0.125	1:34.9	2.825	3.322
5	2⅞	3	0.438	0.438	0.250	0.094	1:9.8	4.950	5.821
6	3	4½	0.500	0.375	0.375	0.250	0	4.909	5.773
6	3½	4	0.375	0.412	0.480	0.420	1:49.6	5.044	5.932
8	3	5¾	0.380	0.380	0.550	0.220	1:14.43	5.600	6.586
8	3½	5¾	0.425	0.471	0.525	0.375	1:28.5	6.682	7.858
10	3½	7½	0.375	0.375	0.625	0.188	1:9	7.298	8.581
10	3⅞	7½	0.438	0.375	0.625	0.188	1:9	7.928	9.323
10	3⅝	7½	0.500	0.375	0.625	0.188	1:9	8.548	10.052

TABLE 51.93 H Beams^a



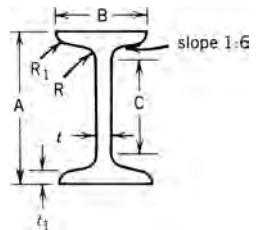
A	B	C	t	t ₁	R	R ₁	Slope	Area ^b (in. ²)	Weight per Foot ^c (lb)
4	4	2 $\frac{3}{8}$	0.313	0.290	0.313	0.145	1:11.3	4.046	4.758
5	5	3 $\frac{3}{8}$	0.313	0.330	0.313	0.165	1:13.6	5.522	6.494
6	5.938	4 $\frac{3}{8}$	0.250	0.360	0.313	0.180	1:15.6	6.678	7.853
8	7.938	6 $\frac{1}{4}$	0.313	0.358	0.313	0.179	1:18.9	9.554	11.263
8	8.125	6 $\frac{1}{4}$	0.500	0.358	0.313	0.179	1:18.9	11.050	12.995

^aUsers are encouraged to ascertain current availability of particular structural shapes through inquiries to their suppliers.

^bAreas listed are based on nominal dimensions.

^cWeights per foot are based on nominal dimensions and a density of 0.098 lb/in.³, which is the density of alloy 6061.

TABLE 51.94 I Beams^a



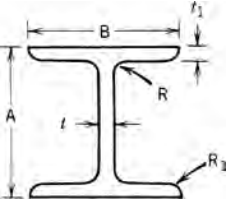
A	B	C	t	t ₁	R	R ₁	Area ^b (in. ²)	Weight per Foot ^c (lb)
3	2.330	1 $\frac{3}{4}$	0.170	0.170	0.270	0.100	1.669	1.963
3	2.509	1 $\frac{3}{4}$	0.349	0.170	0.270	0.100	2.203	2.591
4	2.660	2 $\frac{3}{4}$	0.190	0.190	0.290	0.110	2.249	2.644
4	2.796	2 $\frac{3}{4}$	0.326	0.190	0.290	0.110	2.792	3.283
5	3	3 $\frac{1}{2}$	0.210	0.210	0.310	0.130	2.917	3.430
5	3.284	3 $\frac{1}{2}$	0.494	0.210	0.310	0.130	4.337	5.100
6	3.330	4 $\frac{1}{2}$	0.230	0.230	0.330	0.140	3.658	4.302
6	3.443	4 $\frac{1}{2}$	0.343	0.230	0.330	0.140	4.336	5.099
7	3.755	5 $\frac{1}{4}$	0.345	0.250	0.350	0.150	5.147	6.053
8	4	6 $\frac{1}{4}$	0.270	0.270	0.370	0.160	5.398	6.348
8	4.262	6 $\frac{1}{4}$	0.532	0.270	0.370	0.160	7.494	8.813
10	4.660	8	0.310	0.310	0.410	0.190	7.452	8.764
12	5	9 $\frac{3}{4}$	0.350	0.350	0.450	0.210	9.349	10.994

^aUsers are encouraged to ascertain current availability of particular structural shapes through inquiries to their suppliers.

^bAreas listed are based on nominal dimensions.

^cWeights per foot are based on nominal dimensions and a density of 0.098 lb/in.³, which is the density of alloy 6061.

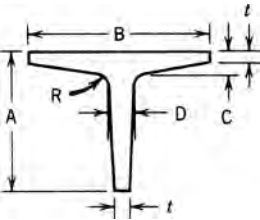
TABLE 51.95 Wide-Flange Beams^a



A	B	t	t ₁	R	R ₁	Area ^b (in. ²)	Weight per Foot ^c (lb)
6.000	4.000	0.230	0.279	0.250	—	3.538	4.161
6.000	6.000	0.240	0.269	0.250	—	4.593	5.401
8.000	5.250	0.230	0.308	0.320	—	5.020	5.904
8.000	6.500	0.245	0.398	0.400	—	7.076	8.321
8.000	8.000	0.288	0.433	0.400	—	9.120	10.725
9.750	7.964	0.292	0.433	0.500	—	9.706	11.414
9.900	5.750	0.240	0.340	0.312	0.031	6.205	7.297
11.940	8.000	0.294	0.516	0.600	—	11.772	13.844
12.060	10.000	0.345	0.576	0.600	—	15.593	18.337

^aUsers are encouraged to ascertain current availability of particular structural shapes through inquiries to their suppliers.
^bAreas listed are based on nominal dimensions.
^cWeights per foot are based on nominal dimensions and a density of 0.098 lb/in.³, which is the density of alloy 6061.

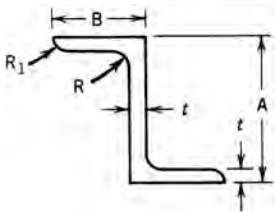
TABLE 51.96 Tees^a



A	B	C	D	t	R	Area ^b (in. ²)	Weight per Foot ^c (lb)
2	2	0.312	0.312	0.250	0.250	1.071	1.259
2 1/4	2 1/4	0.312	0.312	0.250	0.250	1.208	1.421
2 1/2	2 1/2	0.375	0.375	0.312	0.250	1.626	1.912
3	3	0.438	0.438	0.375	0.312	2.310	2.717
4	4	0.438	0.438	0.375	0.500	3.183	3.743

^aUsers are encouraged to ascertain current availability of particular structural shapes through inquiries to their suppliers.
^bAreas listed are based on nominal dimensions.
^cWeights per foot are based on nominal dimensions and a density of 0.098 lb/in.³, which is the density of alloy 6061.

TABLE 51.97 Zees^a



A	B	t	R	R ₁	Area ^b (in. ²)	Weight per Foot ^c (lb)
3	2 ¹¹ / ₁₆	0.250	0.312	0.250	1.984	2.333
3	2 ¹¹ / ₁₆	0.375	0.312	0.250	2.875	3.381
4	3 ¹ / ₁₆	0.250	0.312	0.250	2.422	2.848
4 ¹ / ₁₆	3 ¹ / ₈	0.312	0.312	0.250	3.040	3.575
4 ¹ / ₄	3 ³ / ₁₆	0.375	0.312	0.250	3.672	4.318
5	3 ¹ / ₄	0.500	0.312	0.250	5.265	6.192
5 ¹ / ₁₆	3 ⁵ / ₁₆	0.375	0.312	0.250	4.093	4.813

^aUsers are encouraged to ascertain current availability of particular structural shapes through inquiries to their suppliers.

^bAreas listed are based on nominal dimensions.

^cWeights per foot are based on nominal dimensions and a density of 0.098 lb/in.³, which is the density of alloy 6061.

TABLE 51.98 Aluminum Pipe—Diameters, Wall Thicknesses, and Weights

Nominal Pipe Size ^a (in.)	Schedule Number ^a	Outside Diameter (in.)			Inside Diameter (in.)		Wall Thickness (in.)			Weight per Foot (lb)	
		Nom ^a	Min ^b	Max ^b	Nom		Nom ^a	Min ^b	Max ^b	Nom ^d	Min ^b
¹ / ₈	40	0.405	0.374	0.420	0.269		0.068	0.060	—	0.085	0.091
	80	0.405	0.374	0.420	0.215		0.095	0.083	—	0.109	0.118
¹ / ₄	40	0.540	0.509	0.555	0.364		0.088	0.077	—	0.147	0.159
	80	0.540	0.509	0.555	0.302		0.119	0.104	—	0.185	0.200
³ / ₈	40	0.675	0.644	0.690	0.493		0.091	0.080	—	0.196	0.212
	80	0.675	0.644	0.690	0.493		0.091	0.080	—	0.196	0.212
	5	0.840	0.809	0.855	0.710		0.065	0.053	0.077	0.186	—
	10	0.840	0.809	0.855	0.674		0.083	0.071	0.095	0.232	—
¹ / ₂	40	0.840	0.809	0.855	0.622		0.109	0.095	—	0.294	0.318
	80	0.840	0.809	0.855	0.546		0.147	0.129	—	0.376	0.406
	160	0.840	0.809	0.855	0.464		0.188	0.164	—	0.453	0.489
	5	1.050	1.019	1.065	0.920		0.065	0.053	0.077	0.237	—
³ / ₄	10	1.050	1.019	1.065	0.884		0.083	0.071	0.095	0.297	—
	40	1.050	1.019	1.065	0.824		0.113	0.099	—	0.391	0.422
	80	1.050	1.019	1.065	0.742		0.154	0.135	—	0.510	0.551
	160	1.050	1.019	1.065	0.612		0.219	0.192	—	0.672	0.726
	5	1.315	1.284	1.330	1.185		0.065	0.053	0.077	0.300	—
	10	1.315	1.284	1.330	1.097		0.109	0.095	0.123	0.486	—

(continued)

TABLE 51.98 (Continued)

Nominal Pipe Size ^a (in.)	Schedule Number ^a	Outside			Inside				Weight per	
		Diameter (in.)			Diameter (in.)	Wall Thickness (in.)			Foot (lb)	
		Nom ^a	Min ^b	Max ^b		Nom	Nom ^a	Min ^b	Max ^b	Nom ^d
1	40	1.315	1.284	1.330	1.049	0.133	0.116	—	0.581	0.627
	80	1.315	1.284	1.330	0.957	0.179	0.157	—	0.751	0.811
	160	1.315	1.284	1.330	0.815	0.250	0.219	—	0.984	1.062
	5	1.660	1.629	1.675	1.530	0.065	0.053	0.077	0.383	—
	10	1.660	1.629	1.675	1.442	0.109	0.095	0.123	0.625	—
1 ¹ / ₄	40	1.660	1.629	1.675	1.380	0.140	0.122	—	0.786	0.849
	80	1.660	1.629	1.675	1.278	0.191	0.167	—	1.037	1.120
	160	1.660	1.629	1.675	1.160	0.250	0.219	—	1.302	1.407
	5	1.900	1.869	1.915	1.770	0.065	0.053	0.077	0.441	—
	10	1.900	1.869	1.915	1.682	0.109	0.095	0.123	0.721	—
1 ¹ / ₂	40	1.900	1.869	1.915	1.610	0.145	0.127	—	0.940	1.015
	80	1.900	1.869	1.915	1.500	0.200	0.175	—	1.256	1.357
	160	1.900	1.869	1.915	1.338	0.281	0.246	—	1.681	1.815
	5	2.375	2.344	2.406	2.245	0.065	0.053	0.077	0.555	—
	10	2.375	2.344	2.406	2.157	0.109	0.095	0.123	0.913	—
2	40	2.375	2.351	2.399	2.067	0.154	0.135	—	1.264	1.365
	80	2.375	2.351	2.399	1.939	0.218	0.191	—	1.737	1.876
	160	2.375	2.351	2.399	1.687	0.344	0.301	—	2.581	2.788
	5	2.875	2.844	2.906	2.709	0.083	0.071	0.095	0.856	—
	10	2.875	2.844	2.906	2.635	0.120	0.105	0.135	1.221	—
2 ¹ / ₂	40	2.875	2.846	2.904	2.469	0.203	0.178	—	2.004	2.164
	80	2.875	2.846	2.904	2.323	0.276	0.242	—	2.650	2.862
	160	2.875	2.846	2.904	2.125	0.375	0.328	—	3.464	3.741
	5	3.500	3.469	3.531	3.334	0.083	0.071	0.095	1.048	—
	10	3.500	3.469	3.531	3.260	0.120	0.105	0.135	1.498	—
3	40	3.500	3.465	3.535	3.068	0.216	0.189	—	2.621	2.830
	80	3.500	3.465	3.535	2.900	0.300	0.262	—	3.547	3.830
	160	3.500	3.465	3.535	2.624	0.438	0.383	—	4.955	5.351
3 ¹ / ₂	5	4.000	3.969	4.031	3.834	0.083	0.071	0.095	1.201	—
	10	4.000	3.969	4.031	3.760	0.120	0.105	0.135	1.720	—
	40	4.000	3.960	4.040	3.548	0.226	0.198	—	3.151	3.403
	80	4.000	3.960	4.040	3.364	0.318	0.278	—	4.326	4.672
4	5	4.500	4.469	4.531	4.334	0.083	0.071	0.095	1.354	—
	10	4.500	4.469	4.531	4.160	0.120	0.105	0.135	1.942	—
	40	4.500	4.455	4.545	4.026	0.237	0.207	—	3.733	4.031
	80	4.500	4.455	4.545	3.826	0.337	0.295	—	5.183	5.598
	120	4.500	4.455	4.545	3.624	0.438	0.383	—	6.573	7.099
5	160	4.500	4.455	4.545	3.438	0.531	0.465	—	7.786	8.409
	5.563	5.532	5.625	5.345	0.109	0.095	0.123	2.196	—	—
	10	5.563	5.532	5.625	5.295	0.134	0.117	0.151	2.688	—
	40	5.563	5.507	5.619	5.047	0.258	0.226	—	7.057	5.461
	80	5.563	5.507	5.619	4.813	0.375	0.328	—	7.188	7.763
	120	5.563	5.507	5.619	4.563	0.500	0.438	—	9.353	10.10
	160	5.563	5.507	5.619	4.313	0.625	0.547	—	11.40	12.31

TABLE 51.98 *(Continued)*

Nominal Pipe Size ^a (in.)	Schedule Number ^a	Outside Diameter (in.)			Inside Diameter (in.)	Wall Thickness (in.)			Weight per Foot (lb)	
		Nom ^a	Min ^b	Max ^b	Nom	Nom ^a	Min ^b	Max ^b	Nom ^d	Min ^b
6	5	6.625	6.594	6.687	6.407	0.109	0.095	0.123	2.624	—
	10	6.625	6.594	6.687	6.357	0.134	0.117	0.151	3.213	—
	40	6.625	6.559	6.691	6.065	0.280	0.245	—	6.564	7.089
	80	6.625	6.559	6.691	5.761	0.432	0.378	—	9.884	10.67
	120	6.625	6.559	6.691	5.501	0.562	0.492	—	12.59	13.60
	160	6.625	6.559	6.691	5.187	0.719	0.629	—	15.69	16.94
	5	8.625	8.594	8.718	8.407	0.109	0.095	0.123	3.429	—
	10	8.625	8.594	8.718	8.329	0.148	0.130	0.166	4.635	—
	20	8.625	8.539	8.711	8.125	0.250	0.219	—	7.735	8.354
	30	8.625	8.539	8.711	8.071	0.277	0.242	—	8.543	9.227
8	40	8.625	8.539	8.711	7.981	0.322	0.282	—	9.878	10.67
	60	8.625	8.539	8.711	7.813	0.406	0.355	—	12.33	13.31
	80	8.625	8.539	8.711	7.625	0.500	0.438	—	15.01	16.21
	100	8.625	8.539	8.711	7.437	0.594	0.520	—	17.62	19.03
	120	8.625	8.539	8.711	7.187	0.719	0.629	—	21.00	22.68
	140	8.625	8.539	8.711	7.001	0.812	0.710	—	23.44	25.31
	160	8.625	8.539	8.711	6.813	0.906	0.793	—	25.84	27.90
10	5	10.750	10.719	10.843	10.482	0.134	0.117	0.151	5.256	—
	10	10.750	10.719	10.843	10.420	0.165	0.144	0.186	6.453	—
	20	10.750	10.642	10.858	10.250	0.250	0.219	—	9.698	10.47
	30	10.750	10.642	10.858	10.136	0.307	0.269	—	11.84	12.69
	40	10.750	10.642	10.858	10.020	0.365	0.319	—	14.00	15.12
	60	10.750	10.642	10.858	9.750	0.500	0.438	—	18.93	24.07
	80	10.750	10.642	10.858	9.562	0.594	0.520	—	22.29	28.78
	100	10.750	10.642	10.858	9.312	0.719	0.629	—	26.65	28.78
	5	12.750	12.719	12.843	12.438	0.156	0.136	0.176	7.258	—
	10	12.750	12.719	12.843	12.390	0.1580	0.158	0.202	8.359	—
12	20	12.750	12.622	12.878	12.250	0.250	0.219	—	11.55	12.47
	30	12.750	12.622	12.878	12.090	0.330	0.289	—	15.14	16.35
	40	12.750	12.622	12.878	11.938	0.406	0.355	—	18.52	20.00
	60	12.750	12.622	12.878	11.626	0.562	0.492	—	25.31	27.33
	80	12.750	12.622	12.878	11.374	0.688	0.602	—	30.66	33.11

^aIn accordance with ANSI Standards B36.10 and B36.19.

^bBased on standard tolerances for pipe.

^cFor schedules 5 and 10 these values apply to mean outside diameters.

^dBased on nominal dimensions, plain ends, and a density of 0.098 lb/in.³, the density of 6061 alloy. For alloy 6063 multiply by 0.99, and for alloy 3003 multiply by 1.01.

TABLE 51.99 Aluminum Electrical Conduit—Designed Dimensions and Weights

Nominal or Trade Size of Conduit (in.)	Nominal Inside Diameter (in.)	Outside Diameter (in.)	Nominal Wall Thickness (in.)	Length without Coupling (ft and in.)	Minimum Weight of 10 Unit Lengths with Couplings Attached (lb)
$\frac{1}{4}$	0.364	0.540	0.088	9 – 11 $\frac{1}{2}$	13.3
$\frac{3}{8}$	0.493	0.675	0.091	9 – 11 $\frac{1}{2}$	17.8
$\frac{1}{2}$	0.622	0.840	0.109	9 – 11 $\frac{1}{4}$	27.4
$\frac{3}{4}$	0.824	1.050	0.113	9 – 11 $\frac{1}{4}$	36.4
1	1.049	1.315	0.133	9–11	53.0
1 $\frac{1}{4}$	1.380	1.660	0.140	9–11	69.6
1 $\frac{1}{2}$	1.610	1.900	0.145	9–11	86.2
2	2.067	2.375	0.154	9–11	115.7
2 $\frac{1}{2}$	2.469	2.875	0.203	9 – 10 $\frac{1}{2}$	182.5
3	3.068	3.500	0.216	9 – 10 $\frac{1}{2}$	238.9
3 $\frac{1}{2}$	3.548	4.000	0.226	9 – 10 $\frac{1}{4}$	287.7
4	4.026	4.500	0.237	9 – 10 $\frac{1}{4}$	340.0
5	5.047	5.563	0.258	9–10	465.4
6	6.065	6.625	0.280	9–10	612.5

TABLE 51.100 Equivalent Resistivity Values

Volume Conductivity, Percent International Amended Copper Standard at 68°F	Equivalent Resistivity at 68°F	
	Volume	
	Ohm – Circular Mil/ft	Microhm – in.
52.5	19.754	1.2929
53.5	19.385	1.2687
53.8	19.277	1.2617
53.9	19.241	1.2593
54.0	19.206	1.2570
54.3	19.099	1.2501
55.0	18.856	1.2341
56.0	18.520	1.2121
56.5	18.356	1.2014
57.0	18.195	1.1908
59.0	17.578	1.1505
59.5	17.430	1.1408
61.0	17.002	1.1128
61.2	16.946	1.1091
61.3	16.918	1.1073
61.4	16.891	1.1055
61.5	16.863	1.1037
61.8	16.782	1.0983
62.0	16.727	1.0948
62.1	16.700	1.0931
62.2	16.674	1.0913
62.3	16.647	1.0896
62.4	16.620	1.0878

TABLE 51.101 Property Limits—Wire (Up to 0.374 in. Diameter)

Alloy and Temper	Ultimate Strength (ksi)		Electrical Conductivity ^a percent IACS at 68°F min
	Min	Max	
	1350		
1350-O	8.5	14.0	61.8
1350-H12 and H22	12.0	17.0	61.0
1350-H14 and H24	15.0	20.0	61.0
1350-H16 and H26	17.0	22.0	61.0
	8017		
8017-H212 ^b	15.0	21.0	61.0
	8030		
8030-H221	15.0	22.0	61.0
	8176		
8176-H24	15.0	20.0	61.0
	8177		
8177-H221	15.0	22.0	61.0

Alloy and Temper	Specified Diameter (in.)	Ultimate Strength (ksi min)		Elongation Percent min in 10 in.		Electrical Conductivity ^a min percent IACS at 68°F
		Individual ^a	Average ^c	Individual ^a	Average ^c	
<i>1350</i>						
1350-H19	0.0105–0.0500	23.0	25.0	—	—	61.0
	0.0501–0.0600	27.0	29.0	1.2	1.4	
	0.0601–0.0700	27.0	28.5	1.3	1.5	
	0.0701–0.0800	26.5	28.0	1.4	1.6	
	0.0801–0.0900	26.0	27.5	1.5	1.6	
	0.0901–0.1000	25.5	27.0	1.5	1.6	
	0.1001–0.1100	24.5	26.0	1.5	1.6	
	0.1101–0.1200	24.0	25.5	1.6	1.7	
	0.1201–0.1400	23.5	25.0	1.7	1.8	
	0.1401–0.1500	23.5	24.5	1.8	1.9	
	0.1501–0.1800	23.0	24.0	1.9	2.0	
	0.1801–0.2100	23.0	24.0	2.0	2.1	
0.2101–0.2600	22.5	23.5	2.2	2.3		
<i>5005</i>						
5005-H19	0.0601–0.0700	38.0	40.0	1.3	—	53.5
	0.0701–0.0800	37.5	39.5	1.4	—	
	0.0801–0.0900	37.0	39.0	1.5	—	
	0.0901–0.1000	36.5	38.5	1.5	—	
5005-H19	0.1001–0.1100	36.0	38.0	1.5	—	
	0.1101–0.1200	35.5	37.5	1.6	—	
	0.1201–0.1400	35.0	37.0	1.7	—	
	0.1401–0.1500	35.0	36.5	1.8	—	
	0.1501–0.1600	34.5	36.0	1.9	—	
	0.1601–0.2100	32.5	34.0	2.0	—	
	0.2101–0.2600	31.5	33.0	2.2	—	

TABLE 51.101 (Continued)

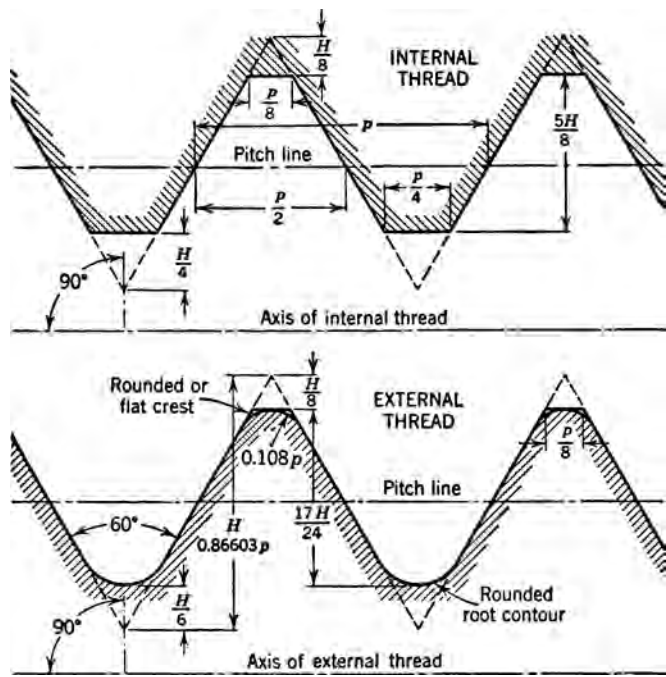
Alloy and Temper	Specified Diameter (in.)	Ultimate Strength (ksi min)		Elongation Percent min in 10 in.		Electrical Conductivity ^a min percent IACS at 68°F
		Individual ^a	Average ^c	Individual ^a	Average ^c	
6201-T81	0.0612–0.1327	46.0	48.0	6201		52.5
				3.0	—	
	0.1328–0.1878	44.0	46.0	3.0	—	
8176-H24	0.0500–0.2040	15.0	17.0	8176		61.0
				10.0	—	

^aTo convert conductivity to maximum resistivity use Table 51.100.

^bApplicable up to 0.250 in.

^cAverage of all tests in a lot.

51.7 STANDARD SCREWS¹³



Standard Screw Threads

The Unified and American Screw Threads included in Table 51.102 are taken from the publication of the American Standards Association, ASA B 1.1—1949. *The coarse-thread series* is the former United States Standard Series. It is recommended for general use in engineering work where conditions do not require the use of a fine thread.

¹³This section is extracted, with permission, from EMPIS Materials Selector. Copyright © 1982 General Electric Co.

TABLE 51.102 Standard Screw Threads

Sizes	Basic Major Diameter <i>D</i> (in.)	Threads per Inch <i>n</i>	Basic Pitch Diameter ^{<i>a</i>} <i>E</i> (in.)	Minor Diameter External Threads <i>K_s</i> (in.)	Minor Diameter Internal Threads <i>K_n</i> (in.)	Section at Minor Diameter at <i>D</i> – 2 <i>h_b</i>) (in. ²)	Stress Area ^{<i>b</i>} (in. ²)
<i>Coarse-thread Series – UNC and NC (Basic Dimensions)</i>							
1 (0.073)	0.0730	64	0.0629	0.0538	0.0561	0.0022	0.0026
2 (0.086)	0.0860	56	0.0744	0.0641	0.0667	0.0031	0.0036
3 (0.099)	0.0990	48	0.0855	0.0734	0.0764	0.0041	0.0048
4 (0.112)	0.1120	40	0.0958	0.0813	0.0849	0.0050	0.0060
5 (0.125)	0.1250	40	0.1088	0.0943	0.0979	0.0067	0.0079
6 (0.138)	0.1380	32	0.1177	0.0997	0.1042	0.0075	0.0090
8 (0.164)	0.1640	32	0.1437	0.1257	0.1302	0.0120	0.0139
10 (0.190)	0.1900	24	0.1629	0.1389	0.1449	0.0145	0.0174
12 (0.216)	0.2160	24	0.1889	0.1649	0.1709	0.0206	0.0240
$\frac{1}{4}$	0.2500	20	0.2175	0.1887	0.1959	0.0269	0.0317
$\frac{5}{16}$	0.3125	18	0.2764	0.2443	0.2524	0.0454	0.0522
$\frac{3}{8}$	0.3750	16	0.3344	0.2983	0.3073	0.0678	0.0773
$\frac{7}{16}$	0.4375	14	0.3911	0.3499	0.3602	0.0933	0.1060
$\frac{1}{2}$	0.5000	13	0.4500	0.4056	0.4167	0.1257	0.1416
$\frac{1}{2}$	0.5000	12	0.4459	0.3978	0.4098	0.1205	0.1374
$\frac{9}{16}$	0.5625	12	0.5084	0.4603	0.4723	0.1620	0.1816
$\frac{5}{8}$	0.6250	11	0.5660	0.5135	0.5266	0.2018	0.2256
$\frac{3}{4}$	0.7500	10	0.6850	0.6273	0.6417	0.3020	0.3340
$\frac{7}{8}$	0.8750	9	0.8028	0.7387	0.7547	0.4193	0.4612
1	1.0000	8	0.9188	0.8466	0.8647	0.5510	0.6051
$1\frac{1}{8}$	1.1250	7	1.0322	0.9497	0.9704	0.6931	0.7627
$1\frac{1}{4}$	1.2500	7	1.1572	1.0747	1.0954	0.8898	0.9684
$1\frac{3}{8}$	1.3750	6	1.2667	1.1705	1.1946	1.0541	1.1538
$1\frac{1}{2}$	1.5000	6	1.3917	1.2955	1.3196	1.2938	1.4041
$1\frac{3}{4}$	1.7500	5	1.6201	1.5046	1.5335	1.7441	1.8983
2	2.0000	$4\frac{1}{2}$	1.8557	1.7274	1.7594	2.3001	2.4971
$2\frac{1}{4}$	2.2500	$4\frac{1}{2}$	2.1057	1.9774	2.0094	3.0212	3.2464
$2\frac{1}{2}$	2.5000	4	2.3376	2.1933	2.2294	3.7161	3.9976
$2\frac{3}{4}$	2.7500	4	2.5876	2.4433	2.4794	4.6194	4.9326
3	3.0000	4	2.8376	2.6933	2.7294	5.6209	5.9659
$3\frac{1}{4}$	3.2500	4	3.0876	2.9433	2.9794	6.7205	7.0992
$3\frac{1}{2}$	3.5000	4	3.3376	3.1933	3.2294	7.9183	8.3268
$3\frac{3}{4}$	3.7500	4	3.5876	3.4433	3.4794	9.2143	9.6546
4	4.0000	4	3.8376	3.6933	3.7294	10.6084	11.0805

(continued)

TABLE 51.102 (Continued)

Sizes	Basic Major Diameter <i>D</i> (in.)	Threads per Inch <i>n</i>	Basic Pitch Diameter ^{<i>a</i>} <i>E</i> (in.)	Minor Diameter External Threads <i>K_s</i> (in.)	Minor Diameter Internal Threads <i>K_n</i> (in.)	Section at Minor Diameter at <i>D</i> − 2 <i>h_b</i>) (in. ²)	Stress Area ^{<i>b</i>} (in. ²)
<i>Fine-Thread Series – UNF and NF (Basic Dimensions)</i>							
0 (0.060)	0.0600	80	0.0519	0.0447	0.0465	0.0015	0.0018
1 (0.073)	0.0730	72	0.0640	0.0560	0.0580	0.0024	0.0027
2 (0.086)	0.0860	64	0.0759	0.0668	0.0691	0.0034	0.0039
3 (0.099)	0.0990	56	0.0874	0.0771	0.0797	0.0045	0.0052
4 (0.112)	0.1120	48	0.0985	0.0864	0.0894	0.0057	0.0065
5 (0.125)	0.1250	44	0.1102	0.0971	0.1004	0.0072	0.0082
6 (0.138)	0.1380	40	0.1218	0.1073	0.1109	0.0087	0.0101
8 (0.164)	0.1640	36	0.1460	0.1299	0.1339	0.0128	0.0146
10 (0.190)	0.1900	32	0.1697	0.1517	0.1562	0.0175	0.0199
12 (0.216)	0.2160	28	0.1928	0.1722	0.1773	0.0226	0.0257
$\frac{1}{4}$	0.2500	28	0.2268	0.2062	0.2113	0.0326	0.0362
$\frac{5}{16}$	0.3125	24	0.2854	0.2614	0.2674	0.0524	0.0579
$\frac{3}{8}$	0.3750	24	0.3479	0.3239	0.3299	0.0809	0.0876
$\frac{7}{16}$	0.4375	20	0.4050	0.3762	0.3834	0.1090	0.1185
$\frac{1}{2}$	0.5000	20	0.4675	0.4387	0.4459	0.1486	0.1597
$\frac{9}{16}$	0.5625	18	0.5264	0.4943	0.5024	0.1888	0.2026
$\frac{5}{8}$	0.6250	18	0.5889	0.5568	0.5649	0.2400	0.2555
$\frac{3}{4}$	0.7500	16	0.7094	0.6733	0.6823	0.3513	0.3724
$\frac{7}{8}$	0.8750	14	0.8286	0.7874	0.7977	0.4805	0.5088
1	1.0000	12	0.9459	0.8978	0.9098	0.6245	0.6624
$1\frac{1}{8}$	1.1250	12	1.0709	1.0228	1.0348	0.8118	0.8549
$1\frac{1}{4}$	1.2500	12	1.1959	1.1478	1.1598	1.0237	1.0721
$1\frac{3}{8}$	1.3750	12	1.3209	1.2728	1.2848	1.2602	1.3137
$1\frac{1}{2}$	1.5000	12	1.4459	1.3978	1.4098	1.5212	1.5799
<i>Extra-Fine-Thread Series – NEF (Basic Dimensions)</i>							
12 (0.216)	0.2160	32	0.1957	0.1777	0.1822	0.0242	0.0269
$\frac{1}{4}$	0.2500	32	0.2297	0.2117	0.2162	0.0344	0.0377
$\frac{5}{16}$	0.3125	32	0.2922	0.2742	0.2787	0.0581	0.0622
$\frac{3}{8}$	0.3750	32	0.3547	0.3367	0.3412	0.0878	0.0929
$\frac{7}{16}$	0.4375	28	0.4143	0.3937	0.3988	0.1201	0.1270
$\frac{1}{2}$	0.5000	28	0.4768	0.4562	0.4613	0.1616	0.1695
$\frac{9}{16}$	0.5625	24	0.5354	0.5114	0.5174	0.2030	0.2134
$\frac{5}{8}$	0.6250	24	0.5979	0.5739	0.5799	0.2560	0.2676
$\frac{11}{16}$	0.6875	24	0.6604	0.6364	0.6424	0.3151	0.3280
$\frac{3}{4}$	0.7500	20	0.7175	0.6887	0.6959	0.3685	0.3855
$\frac{13}{16}$	0.8125	20	0.7800	0.7512	0.7584	0.4388	0.4573

TABLE 51.102 (Continued)

Sizes	Basic Major Diameter <i>D</i> (in.)	Threads per Inch <i>n</i>	Basic Pitch Diameter ^a <i>E</i> (in.)	Minor Diameter External Threads <i>K_s</i> (in.)	Minor Diameter Internal Threads <i>K_n</i> (in.)	Section at Minor Diameter at <i>D</i> – 2 <i>h_b</i>) (in. ²)	Stress Area ^b (in. ²)
$\frac{7}{8}$	0.8750	20	0.8425	0.8137	0.8209	0.5153	0.5352
$\frac{15}{16}$	0.9375	20	0.9050	0.8762	0.8834	0.5979	0.6194
<i>Fine-Thread Series – UNF and NF (Basic Dimensions)</i>							
1	1.0000	20	0.9675	0.9387	0.9459	0.6866	0.7095
$1\frac{1}{16}$	1.0625	18	1.0264	0.9943	1.0024	0.7702	0.7973
$1\frac{1}{8}$	1.1250	18	1.0889	1.0568	1.0649	0.8705	0.8993
$1\frac{3}{16}$	1.1875	18	1.1514	1.1193	1.1274	0.9770	1.0074
$1\frac{1}{4}$	1.2500	18	1.2139	1.1818	1.1899	1.0895	1.1216
$1\frac{5}{16}$	1.3125	18	1.2764	1.2443	1.2524	1.2082	1.2420
$1\frac{3}{8}$	1.3750	18	1.3389	1.3068	1.3149	1.3330	1.3684
$1\frac{7}{16}$	1.4375	18	1.4014	1.3693	1.3774	1.4640	1.5010
$1\frac{1}{2}$	1.5000	18	1.4639	1.4318	1.4399	1.6011	1.6397
$1\frac{9}{16}$	1.5625	18	1.5264	1.4943	1.5024	1.7444	1.7846
$1\frac{5}{8}$	1.6250	18	1.5889	1.5568	1.5649	1.8937	1.9357
$1\frac{11}{16}$	1.6875	18	1.6514	1.6193	1.6274	2.0493	2.0929
$1\frac{3}{4}$	1.7500	16	1.7094	1.6733	1.6823	2.1873	2.2382
2	2.0000	16	1.9594	1.9233	1.9323	2.8917	2.9501

Note: Bold type indicates unified threads – UNC and UNF.
^aBritish: effective diameter.
^bThe stress area is the assumed area of an externally threaded part which is used for the purpose of computing the tensile strength.

TABLE 51.103 ASA^a Standard Bolts and Nuts

Nominal Size	Across Flats (in.)	Across Square Corners (in.)	Across Hex Corners (in.)	Thickness	
				Unfinished (in.)	Semifinished (in.)
Regular Bolt Heads					
$\frac{1}{4}$	$\frac{3}{8}$	0.498	0.413	$\frac{11}{64}$	$\frac{5}{32}$
$\frac{5}{16}$	$\frac{1}{2}$	0.665	0.552	$\frac{13}{64}$	$\frac{3}{16}$
$\frac{3}{8}$	$\frac{9}{16}$	0.747	0.620	$\frac{1}{4}$	$\frac{15}{64}$
$\frac{7}{16}$	$\frac{5}{8}$	0.828	0.687	$\frac{19}{64}$	$\frac{9}{32}$
$\frac{1}{2}$	$\frac{3}{4}$	0.995	0.826	$\frac{21}{64}$	$\frac{19}{64}$
$\frac{9}{16}$	$\frac{7}{8}$	1.163	0.966	$\frac{3}{8}$	$\frac{11}{32}$
$\frac{5}{8}$	$\frac{15}{16}$	1.244	1.033	$\frac{27}{64}$	$\frac{25}{64}$

(continued)

TABLE 51.103 (Continued)

Nominal Size	Across Flats (in.)	Across Square Corners (in.)	Across Hex Corners (in.)	Thickness	
				Unfinished (in.)	Semifinished (in.)
$\frac{3}{4}$	$1\frac{1}{8}$	1.494	1.240	$\frac{1}{2}$	$\frac{15}{32}$
$\frac{7}{8}$	$1\frac{5}{16}$	1.742	1.447	$\frac{19}{32}$	$\frac{9}{16}$
1	$1\frac{1}{2}$	1.991	1.653	$\frac{21}{32}$	$\frac{19}{32}$
$1\frac{1}{8}$	$1\frac{11}{16}$	2.239	1.859	$\frac{3}{4}$	$\frac{11}{16}$
$1\frac{1}{4}$	$1\frac{7}{8}$	2.489	2.066	$\frac{27}{32}$	$\frac{25}{32}$
$1\frac{3}{8}$	$2\frac{1}{16}$	2.738	2.273	$\frac{29}{32}$	$\frac{27}{32}$
$1\frac{1}{2}$	$2\frac{1}{4}$	2.986	2.480	1	$\frac{15}{16}$
$1\frac{5}{8}$	$2\frac{7}{16}$	3.235	2.686	$1\frac{3}{32}$	$1\frac{1}{32}$
$1\frac{3}{4}$	$2\frac{5}{8}$	3.485	2.893	$1\frac{5}{32}$	$1\frac{3}{32}$
$1\frac{7}{8}$	$2\frac{13}{16}$	3.733	3.100	$1\frac{1}{4}$	$1\frac{3}{16}$
2	3	3.982	3.306	$1\frac{11}{32}$	$1\frac{7}{32}$
$2\frac{1}{4}$	$3\frac{3}{8}$	4.479	3.719	$1\frac{1}{2}$	$1\frac{3}{8}$
$2\frac{1}{2}$	$3\frac{3}{4}$	4.977	4.133	$1\frac{21}{32}$	$1\frac{17}{32}$
$2\frac{3}{4}$	$4\frac{1}{8}$	5.476	4.546	$1\frac{53}{64}$	$1\frac{11}{16}$
3	$4\frac{1}{2}$	5.973	4.959	2	$1\frac{7}{8}$
Heavy Bolt Heads					
$\frac{1}{2}$	$\frac{7}{8}$	1.167	0.969	$\frac{7}{16}$	$\frac{13}{32}$
$\frac{9}{16}$	$\frac{15}{16}$	1.249	1.037	$\frac{15}{32}$	$\frac{7}{16}$
$\frac{5}{8}$	$1\frac{1}{16}$	1.416	1.175	$\frac{17}{32}$	$\frac{1}{2}$
$\frac{3}{4}$	$1\frac{1}{4}$	1.665	1.383	$\frac{5}{8}$	$\frac{19}{32}$
$\frac{7}{8}$	$1\frac{7}{16}$	1.914	1.589	$\frac{23}{32}$	$\frac{11}{16}$
1	$1\frac{5}{8}$	2.162	1.796	$\frac{13}{16}$	$\frac{3}{4}$
$1\frac{1}{8}$	$1\frac{13}{16}$	2.411	2.002	$\frac{29}{32}$	$\frac{27}{32}$
$1\frac{1}{4}$	2	2.661	2.209	1	$\frac{15}{16}$
$1\frac{3}{8}$	$2\frac{3}{16}$	2.909	2.416	$1\frac{3}{32}$	$1\frac{1}{32}$
$1\frac{1}{2}$	$2\frac{3}{8}$	3.158	2.622	$1\frac{3}{16}$	$1\frac{1}{8}$
$1\frac{5}{8}$	$2\frac{9}{16}$	3.406	2.828	$1\frac{9}{32}$	$1\frac{7}{32}$
$1\frac{3}{4}$	$2\frac{3}{4}$	3.655	3.036	$1\frac{3}{8}$	$1\frac{5}{16}$
$1\frac{7}{8}$	$2\frac{15}{16}$	3.905	3.242	$1\frac{15}{32}$	$1\frac{13}{32}$
2	$3\frac{1}{8}$	4.153	3.449	$1\frac{9}{16}$	$1\frac{7}{16}$
$2\frac{1}{4}$	$3\frac{1}{2}$	4.652	3.862	$1\frac{3}{4}$	$1\frac{5}{8}$
$2\frac{1}{2}$	$3\frac{7}{8}$	5.149	4.275	$1\frac{15}{16}$	$1\frac{13}{16}$
$2\frac{3}{4}$	$4\frac{1}{4}$	5.646	4.688	$2\frac{1}{8}$	2
3	$4\frac{5}{8}$	6.144	5.102	$2\frac{5}{16}$	$2\frac{3}{16}$

TABLE 51.103 (Continued)

Nominal Size	Width Across Flats (in.)	Width Across Corners		Thickness Unfinished, Regular		Thickness Semifinished, Regular	
		Square (in.)	Hex (in.)	Nuts (in.)	Jam Nuts (in.)	Nuts (in.)	Jam Nuts (in.)
Regular Nuts and Regular Jam Nuts							
$\frac{1}{4}$	$\frac{7}{16}$	0.584	0.484	$\frac{7}{32}$	$\frac{5}{32}$	$\frac{13}{64}$	$\frac{9}{64}$
$\frac{5}{16}$	$\frac{9}{16}$	0.751	0.624	$\frac{17}{64}$	$\frac{3}{16}$	$\frac{1}{4}$	$\frac{11}{64}$
$\frac{3}{8}$	$\frac{5}{8}$	0.832	0.691	$\frac{21}{64}$	$\frac{7}{32}$	$\frac{5}{16}$	$\frac{13}{64}$
$\frac{7}{16}$	$\frac{3}{4}$	1.000	0.830	$\frac{3}{8}$	$\frac{1}{4}$	$\frac{23}{64}$	$\frac{15}{64}$
$\frac{1}{2}$	$\frac{13}{16}$	1.082	0.898	$\frac{7}{16}$	$\frac{5}{16}$	$\frac{27}{64}$	$\frac{19}{64}$
$\frac{9}{16}$	$\frac{7}{8}$	1.163	0.966	$\frac{1}{2}$	$\frac{11}{32}$	$\frac{31}{64}$	$\frac{21}{64}$
$\frac{5}{8}$	1	1.330	1.104	$\frac{35}{64}$	$\frac{3}{8}$	$\frac{17}{32}$	$\frac{23}{64}$
$\frac{3}{4}$	$1\frac{1}{8}$	1.494	1.240	$\frac{21}{32}$	$\frac{7}{16}$	$\frac{41}{64}$	$\frac{27}{64}$
$\frac{7}{8}$	$1\frac{5}{16}$	1.742	1.447	$\frac{49}{64}$	$\frac{1}{2}$	$\frac{3}{4}$	$\frac{31}{64}$
1	$1\frac{1}{2}$	1.991	1.653	$\frac{7}{8}$	$\frac{9}{16}$	$\frac{55}{64}$	$\frac{35}{64}$
$1\frac{1}{8}$	$1\frac{11}{16}$	2.239	1.859	1	$\frac{5}{8}$	$\frac{31}{32}$	$\frac{39}{64}$
$1\frac{1}{4}$	$1\frac{7}{8}$	2.489	2.066	$1\frac{3}{32}$	$\frac{3}{4}$	$1\frac{1}{16}$	$\frac{23}{32}$
$1\frac{3}{8}$	$2\frac{1}{16}$	2.738	2.273	$1\frac{13}{64}$	$\frac{13}{16}$	$1\frac{11}{64}$	$\frac{25}{32}$
$1\frac{1}{2}$	$2\frac{1}{4}$	2.986	2.480	$1\frac{5}{16}$	$\frac{7}{8}$	$1\frac{9}{32}$	$\frac{27}{32}$
$1\frac{5}{8}$	$2\frac{7}{16}$	3.235	2.686	$1\frac{27}{64}$	$\frac{15}{16}$	$1\frac{25}{64}$	$\frac{29}{32}$
$1\frac{3}{4}$	$2\frac{5}{8}$	3.485	2.893	$1\frac{17}{32}$	1	$1\frac{1}{2}$	$\frac{31}{32}$
$1\frac{7}{8}$	$2\frac{13}{16}$	3.733	3.100	$1\frac{41}{64}$	$1\frac{1}{16}$	$1\frac{39}{64}$	$1\frac{1}{32}$
2	3	3.982	3.306	$1\frac{3}{4}$	$1\frac{1}{8}$	$1\frac{23}{32}$	$1\frac{3}{32}$
$2\frac{1}{4}$	$3\frac{3}{8}$	4.479	3.719	$1\frac{31}{32}$	$1\frac{1}{4}$	$1\frac{59}{64}$	$1\frac{13}{64}$
$2\frac{1}{2}$	$3\frac{3}{4}$	4.977	4.133	$2\frac{3}{16}$	$1\frac{1}{2}$	$2\frac{9}{64}$	$1\frac{29}{64}$
$2\frac{3}{4}$	$4\frac{1}{8}$	5.476	4.546	$2\frac{13}{32}$	$1\frac{5}{8}$	$2\frac{23}{64}$	$1\frac{37}{64}$
3	$4\frac{1}{2}$	5.973	4.959	$2\frac{5}{8}$	$1\frac{3}{4}$	$2\frac{37}{64}$	$1\frac{45}{64}$
Heavy Nuts and Heavy Jam Nuts							
$\frac{1}{4}$	$\frac{1}{2}$	0.670	0.556	$\frac{1}{4}$	$\frac{3}{16}$	$\frac{15}{64}$	$\frac{11}{64}$
$\frac{5}{16}$	$\frac{19}{32}$	0.794	0.659	$\frac{5}{16}$	$\frac{7}{32}$	$\frac{19}{64}$	$\frac{13}{64}$
$\frac{3}{8}$	$\frac{11}{16}$	0.919	0.763	$\frac{3}{8}$	$\frac{1}{4}$	$\frac{23}{64}$	$\frac{15}{64}$
$\frac{7}{16}$	$\frac{25}{32}$	1.042	0.865	$\frac{7}{16}$	$\frac{9}{32}$	$\frac{27}{64}$	$\frac{17}{64}$
$\frac{1}{2}$	$\frac{7}{8}$	1.167	0.969	$\frac{1}{2}$	$\frac{5}{16}$	$\frac{31}{64}$	$\frac{19}{64}$
$\frac{9}{16}$	$\frac{15}{16}$	1.249	1.037	$\frac{9}{16}$	$\frac{11}{32}$	$\frac{35}{64}$	$\frac{21}{64}$
$\frac{5}{8}$	$1\frac{1}{16}$	1.416	1.175	$\frac{5}{8}$	$\frac{3}{8}$	$\frac{39}{64}$	$\frac{23}{64}$
$\frac{3}{4}$	$1\frac{1}{4}$	1.665	1.382	$\frac{3}{4}$	$\frac{7}{16}$	$\frac{47}{64}$	$\frac{27}{64}$
$\frac{7}{8}$	$1\frac{7}{16}$	1.914	1.589	$\frac{7}{8}$	$\frac{1}{2}$	$\frac{55}{64}$	$\frac{31}{64}$
1	$1\frac{5}{8}$	2.162	1.796	1	$\frac{9}{16}$	$\frac{63}{64}$	$\frac{35}{64}$

(continued)

TABLE 51.103 (Continued)

Nominal Size	Width Across Flats (in.)	Width Across Corners		Thickness Unfinished, Regular		Thickness Semifinished, Regular	
		Square (in.)	Hex (in.)	Nuts (in.)	Jam Nuts (in.)	Nuts (in.)	Jam Nuts (in.)
$1\frac{1}{8}$	$1\frac{13}{16}$	2.411	2.002	$1\frac{1}{8}$	$\frac{5}{8}$	$1\frac{7}{64}$	$\frac{39}{64}$
$1\frac{1}{4}$	2	2.661	2.209	$1\frac{1}{4}$	$\frac{3}{4}$	$1\frac{7}{32}$	$\frac{23}{32}$
$1\frac{3}{8}$	$2\frac{3}{16}$	2.909	2.416	$1\frac{3}{8}$	$\frac{13}{16}$	$1\frac{11}{32}$	$\frac{25}{32}$
$1\frac{1}{2}$	$2\frac{3}{8}$	3.158	2.622	$1\frac{1}{2}$	$\frac{7}{8}$	$1\frac{15}{32}$	$\frac{27}{32}$
$1\frac{5}{8}$	$2\frac{9}{16}$	3.406	2.828	$1\frac{5}{8}$	$\frac{15}{16}$	$1\frac{19}{32}$	$\frac{29}{32}$
$1\frac{3}{4}$	$2\frac{3}{4}$	3.656	3.035	$1\frac{3}{4}$	1	$1\frac{23}{32}$	$\frac{31}{32}$
$1\frac{7}{8}$	$2\frac{15}{16}$	3.905	3.242	$1\frac{7}{8}$	$1\frac{1}{16}$	$1\frac{27}{32}$	$1\frac{1}{32}$
2	$3\frac{1}{8}$	4.153	3.449	2	$1\frac{1}{8}$	$1\frac{31}{32}$	$1\frac{3}{32}$
$2\frac{1}{4}$	$3\frac{1}{2}$	4.652	3.862	$2\frac{1}{4}$	$1\frac{1}{4}$	$2\frac{13}{64}$	$1\frac{13}{64}$
$2\frac{1}{2}$	$3\frac{7}{8}$	5.149	4.275	$2\frac{1}{2}$	$1\frac{1}{2}$	$2\frac{29}{64}$	$1\frac{29}{64}$
$2\frac{3}{4}$	$4\frac{1}{4}$	5.646	4.688	$2\frac{3}{4}$	$1\frac{5}{8}$	$2\frac{45}{64}$	$1\frac{37}{64}$
3	$4\frac{5}{8}$	6.144	5.102	3	$1\frac{3}{4}$	$2\frac{61}{64}$	$1\frac{45}{64}$
$3\frac{1}{4}$	5	6.643	5.515	$3\frac{1}{4}$	$1\frac{7}{8}$	$3\frac{3}{16}$	$1\frac{13}{16}$
$3\frac{1}{2}$	$5\frac{3}{8}$	7.140	5.928	$3\frac{1}{2}$	2	$3\frac{7}{16}$	$1\frac{15}{16}$
$3\frac{3}{4}$	$5\frac{3}{4}$	7.637	6.341	$3\frac{3}{4}$	$2\frac{1}{8}$	$3\frac{11}{16}$	$2\frac{1}{16}$
4	$6\frac{1}{8}$	8.135	6.755	4	$2\frac{1}{4}$	$3\frac{15}{16}$	$2\frac{3}{16}$

Nominal Size	Regular Slotted Nuts Semifinished			Heavy Slotted Nuts Semifinished				
	Width		Thickness (in.)	Width		Thickness (in.)	Slot	
	Across Flats (in.)	Across Corners (in.)		Across Flats (in.)	Across Corners (in.)		Width (in.)	Depth (in.)
$\frac{1}{4}$	$\frac{7}{16}$	0.485	$\frac{13}{64}$	$\frac{1}{2}$	0.556	$\frac{15}{64}$	$\frac{5}{64}$	$\frac{3}{32}$
$\frac{5}{16}$	$\frac{9}{16}$	0.624	$\frac{1}{4}$	$\frac{19}{32}$	0.659	$\frac{19}{64}$	$\frac{3}{32}$	$\frac{3}{32}$
$\frac{3}{8}$	$\frac{5}{8}$	0.691	$\frac{5}{16}$	$\frac{11}{16}$	0.763	$\frac{23}{64}$	$\frac{1}{8}$	$\frac{1}{8}$
$\frac{7}{16}$	$\frac{3}{4}$	0.830	$\frac{23}{64}$	$\frac{25}{32}$	0.865	$\frac{27}{64}$	$\frac{1}{8}$	$\frac{5}{32}$
$\frac{1}{2}$	$\frac{13}{16}$	0.898	$\frac{27}{64}$	$\frac{7}{8}$	0.969	$\frac{31}{64}$	$\frac{5}{32}$	$\frac{5}{32}$
$\frac{9}{16}$	$\frac{7}{8}$	0.966	$\frac{31}{64}$	$\frac{15}{16}$	1.037	$\frac{35}{64}$	$\frac{5}{32}$	$\frac{3}{16}$
$\frac{5}{8}$	1	1.104	$\frac{17}{32}$	$1\frac{1}{16}$	1.175	$\frac{39}{64}$	$\frac{3}{16}$	$\frac{7}{32}$
$\frac{3}{4}$	$1\frac{1}{8}$	1.240	$\frac{41}{64}$	$1\frac{1}{4}$	1.382	$\frac{47}{64}$	$\frac{3}{16}$	$\frac{1}{4}$
$\frac{7}{8}$	$1\frac{5}{16}$	1.447	$\frac{3}{4}$	$1\frac{7}{16}$	1.589	$\frac{55}{64}$	$\frac{3}{16}$	$\frac{1}{4}$
1	$1\frac{1}{2}$	1.653	$\frac{55}{64}$	$1\frac{5}{8}$	1.796	$\frac{63}{64}$	$\frac{1}{4}$	$\frac{9}{32}$
$1\frac{1}{8}$	$1\frac{11}{16}$	1.859	$\frac{31}{32}$	$1\frac{13}{16}$	2.002	$1\frac{7}{64}$	$\frac{1}{4}$	$\frac{11}{32}$

TABLE 51.103 (Continued)

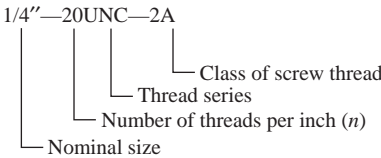
Nominal Size	Regular Slotted Nuts Semifinished			Heavy Slotted Nuts Semifinished			Slot	
	Width		Thickness (in.)	Width		Thickness (in.)		
	Across Flats (in.)	Across Corners (in.)		Across Flats (in.)	Across Corners (in.)			
	Width (in.)	Depth (in.)						
1 $\frac{1}{4}$	1 $\frac{7}{8}$	2.066	1 $\frac{1}{16}$	2	2.209	1 $\frac{7}{32}$	$\frac{5}{16}$	$\frac{3}{8}$
1 $\frac{3}{8}$	2 $\frac{1}{16}$	2.273	1 $\frac{11}{64}$	2 $\frac{3}{16}$	2.416	1 $\frac{11}{32}$	$\frac{5}{16}$	$\frac{3}{8}$
1 $\frac{1}{2}$	2 $\frac{1}{4}$	2.480	1 $\frac{9}{32}$	2 $\frac{3}{8}$	2.622	1 $\frac{15}{32}$	$\frac{3}{8}$	$\frac{7}{16}$
1 $\frac{5}{8}$	2 $\frac{7}{16}$	0.686	1 $\frac{25}{64}$	2 $\frac{9}{16}$	2.828	1 $\frac{19}{32}$	$\frac{3}{8}$	$\frac{7}{16}$
1 $\frac{3}{4}$	2 $\frac{5}{8}$	2.893	1 $\frac{1}{2}$	2 $\frac{3}{4}$	3.035	1 $\frac{23}{32}$	$\frac{7}{16}$	$\frac{1}{2}$
1 $\frac{7}{8}$	2 $\frac{13}{16}$	3.100	1 $\frac{39}{64}$	2 $\frac{15}{16}$	3.242	1 $\frac{27}{32}$	$\frac{7}{16}$	$\frac{9}{16}$
2	3	3.306	1 $\frac{23}{32}$	3 $\frac{1}{8}$	3.449	1 $\frac{31}{32}$	$\frac{7}{16}$	$\frac{9}{16}$
2 $\frac{1}{4}$	3 $\frac{3}{8}$	3.719	1 $\frac{59}{64}$	3 $\frac{1}{2}$	3.862	2 $\frac{13}{64}$	$\frac{7}{16}$	$\frac{9}{16}$
2 $\frac{1}{2}$	3 $\frac{3}{4}$	4.133	2 $\frac{9}{64}$	3 $\frac{7}{8}$	4.275	2 $\frac{29}{64}$	$\frac{9}{16}$	$\frac{11}{16}$
2 $\frac{3}{4}$	4 $\frac{1}{8}$	4.546	2 $\frac{23}{64}$	4 $\frac{1}{4}$	4.688	2 $\frac{45}{64}$	$\frac{9}{16}$	$\frac{11}{16}$
3	4 $\frac{1}{2}$	4.959	2 $\frac{37}{64}$	4 $\frac{5}{8}$	5.102	2 $\frac{61}{64}$	$\frac{5}{8}$	$\frac{3}{4}$

^aANSI standards B18.2.1 – 1981, B18.2.2 – 1972 (R1983), B18.6.3 – 1972 (R1983).

The *fine-thread series* is the former “Regular Screw Thread Series” established by the Society of Automotive Engineers (SAE). The *fine-thread series* is recommended for general use in automotive and aircraft work and where special conditions require a fine thread. The *extra-fine-thread series* is the same as the former SAE fine series and the present SAE extra-fine series. It is used particularly in aircraft and aeronautical equipment where (a) thin-walled material is to be threaded; (b) thread depth of nuts clearing ferrules, coupling flanges, and so on, must be held to a minimum; and (c) a maximum practicable number of threads is required within a given thread length.

The method of designating a screw thread is by the use of the initial letters of the thread series, preceded by the nominal size (diameter in inches or the screw number) and number of threads per inch, all in Arabic numerals, and followed by the classification designation, with or without the pitch diameter tolerances or limits of size. An example of an external thread designation and its meaning is as follows:

EXAMPLE 1



A left-hand thread must be identified by the letters LH following the class designation. If no such designation is used, the thread is assumed to be right hand.

Classes of thread are distinguished from each other by the amounts of tolerance and allowance specified in ASA B 1.1—1949.

Selection of Screws

By definition, a *screw* is a fastener that is intended to be torqued by the head. Screws are the most widely used method of assembly despite recent technical advances of adhesives, welding, and other joining techniques. Use of screws is essential in those applications that require ease of disassembly for normal maintenance and service. There is no real economy if savings made in factory installation create service problems later. There are many types of screws, and each variety will be treated separately. Material selection is generally common to all types of screws.

Material. Not all materials are suitable for the processes used in the manufacture of fasteners. Large-volume users or those with critical requirements can be very selective in their choice of materials. Low-volume users or those with noncritical applications would be wise to permit a variety of materials in a general category in order to improve availability and lower cost. For example, it is usually desirable to specify low-carbon steel or 18-8-type stainless steel¹⁴ rather than ask for a specific grade.

Low-carbon steel is widely used in the manufacture of fasteners where lowest cost is desirable and tensile strength requirements are ~50,000 psi. If corrosion is a problem, these fasteners can be plated either electrically or mechanically. Zinc or cadmium plating is used in most applications. Other finishes include nickel, chromium, copper, tin, and silver electroplating; electroless nickel and other immersion coatings; hot dip galvanizing; and phosphate coatings.

Medium-carbon steel, quenched, and tempered is widely used in applications requiring tensile strengths from 90,000 to 120,000 psi. Alloy steels are used in applications requiring tensile strengths from 115,000 to 180,000 psi, depending on the grade selected. Where better corrosion resistance is required, 300 series stainless steel can be specified. The 400 series stainless steel is used if it is necessary to have a corrosion-resistant material that can be hardened and tempered by heat treatment.

For superior corrosion resistance, materials such as brass, bronze, aluminum, or nickel are sometimes used in the manufacture of fasteners. If strength is no problem, plastics such as nylons are used in severe corrosion applications.

Drivability. When selecting a screw, thought must be given to the means of driving for assembly and disassembly as well as the head shape. Most screw heads provide a slot, a recess, or a hexagon shape as a means of driving. The slotted screw is the least preferred driving style and serves only when appearance must be combined with ease of disassembly with a common screwdriver. Only a limited amount of torque can be applied with a screwdriver. A slot can become inoperative after repeated disassembly destroys the

¹⁴ Manufacturer may use UNS—S30200, S30300, S30400, S30500 (AISI type 302, 303, 304, or 305) depending upon quantity, diameter, and manufacturing process.

edge of the wall that the blade of the screwdriver bears against. The hexagon head is preferred for the following reasons:

- Least likely to accidentally spin out (thereby marring the surface of the product)
- Lowest initial cost
- Adaptable to high-speed power drive
- Minimum worker fatigue
- Ease of assembly in difficult places
- Permits higher driving torque, especially in large sizes where strength is important
- Contains no recess to become clogged with dirt and interfere with driving
- Contains no recess to weaken the head

Unless frequent field disassembly is required, use of the unslotted hex head is preferred.

Appearance is the major disadvantage of the hex head, and this one factor is judged sufficient to eliminate it from consideration for the front or top of products.

The recessed head fastener is widely used and becomes the first choice for appearance applications. It usually costs more than a slot or a hexagon shape. There are many kinds of recesses. The Phillips and Phillips POZIDRIV are most widely used. To a lesser extent the Frearson, clutch-type, hexagonal, and fluted socket heads are used. For special applications, proprietary types of tamper-resistant heads can be selected (Figure 51.1).

The recessed head has some of the same advantages as the hex head (see preceding list). It also has improved appearance. The Phillips POZIDRIV is slowly replacing the Phillips recess. The POZIDRIV recess can be readily identified by four radial lines centered between each recess slot. These slots are a slight modification of the conventional Phillips recess. This change improves the fit between the driver and the recess, thus

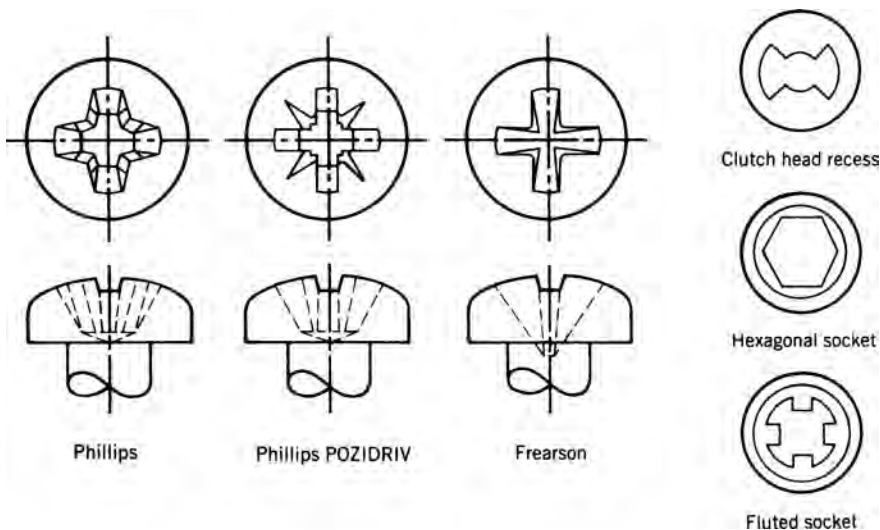


FIGURE 51.1 Recessed head fasteners.

minimizing the possibility of marring a surface from accidental spinout of the driver as well as increasing the life of the driver. The POZIDRIV design is recommended in high-production applications requiring high driving torques. The POZIDRIV recess usually sells at a high-production applications requiring high driving torques. The POZIDRIV recess usually sells at a slightly higher price than the conventional Phillips recess, but some suppliers will furnish either at the same price. The savings resulting from longer tool life will usually justify the higher initial cost.

A conventional Phillips driver could be used to install or disassemble a POZIDRIV screw. However, a POZIDRIV driver should be used with a POZIDRIV screw in order to take advantage of the many features inherent in the new design. To avoid confusion, it should be clearly understood that a POZIDRIV driver cannot be used to install or remove a conventional Phillips head screw.

A Frearson recess is a somewhat different design than a Phillips recess and has the big advantage that one driving tool can be used for all sizes whereas a Phillips may require four driving tools in the range from no. 2 (0.086-in.) to 3/8 (0.375-in.) screw size. This must be balanced against the following disadvantages:

Limited availability.

Greater penetration of the recess means thinner walls between the bottom of the recess and the outer edge of the screw, which tends to weaken the head.

The sharp point of the driver can easily scratch or otherwise mar the surface of the product if it accidentally touches.

Although one driver can be used for all sizes, for optimum results, different size drivers are recommended for installing various screw sizes, thus minimizing the one real advantage of the Frearson recess.

The hexagon and fluted socket head cap screws are only available in expensive high-strength alloy steel. Its unique small outside diameter or cylindrical head is useful on flanges, counterbored holes, or other locations where clearances are restricted. Such special applications may justify the cost of a socket head cap screw. Appreciable savings can be made in other applications by substitution of a hexagon head screw.

Despite any claims to the contrary, the dimensional accuracy of hexagon socket head cap screws is no better than that of other cold-headed products, and there is no merit in close-thread tolerances, which are advocated by some manufacturers of these products. The high prices, therefore, should be justified solely on the basis of possible space savings in using the cylindrical head.

The fluted socket is not as readily available and should only be considered in the very small sizes where a hexagon key tends to round out the socket. The fluted socket offers spline design so that the key will neither slip nor be subject to excessive wear.

Many types of special recesses are tamper resistant. In most of these designs, the recess is an unusual shape requiring a special tool for assembly and disassembly. A readily available driving tool such as a screwdriver or hexagon key would not fit the recess. The purpose of a tamper-resistant fastener is to prevent unauthorized removal of parts and equipment. Their protection is needed on any product located in public places to discourage vandalism and thievery. They may also be necessary on some consumer products as a safety measure to protect the amateur repairman from injury or to prevent him from causing serious damage to equipment. With product liability mania what it is today, the term "tamperproof" has all but disappeared. Now the fasteners are called "tamper resistant."

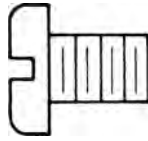


FIGURE 51.2 Pan head.

They are the same as they were under their previous name, but the new term better reflects their true capabilities. Any skilled thief with ample time and proper tools can saw, drill, blast, or otherwise disassemble any tamper-resistant fastener. Therefore, these fasteners are intended only to discourage the casual thief or amateur tinkerer and make it more difficult for a skilled professional. Whatever the choice of fastener design, it is essential that hardened material be specified. No fastener is ever truly tamperproof, but hardened steel helps. Fasteners made of soft material can be disassembled easily by sawing a slot, hammering with a chisel, or drilling a hole and using an extraction bit.

Head Shapes

The following information is equally applicable to all types of recesses as well as a slotted head. For simplification only slotted screws are shown.

The pan head is the most widely used and is intended to replace the round, binding, and truss heads in order to keep varieties to a minimum. It is preferred because it presents the best combination of appearance with adequate head height to minimize weakness due to depth of penetration of the recess (Figure 51.2).

The round head was widely used in the past (Figure 51.3). It has since been delisted as an American National Standard. Give preference to pan heads on all new designs. Figure 51.4 shows the superiority of the pan head: The high edge of the pan head at its periphery, where driving action is most effective, provides superior driver-slot engagement and reduces the tendency to chew away the metal at the edge of the slot.

The flat head is used where a flush surface is required. The countersunk section aids in centering the screw (Figure 51.5).

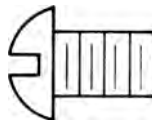


FIGURE 51.3 Round head.

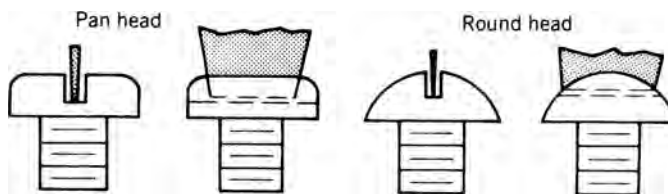
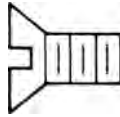
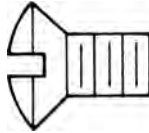


FIGURE 51.4 Drive-slot engagement.

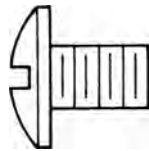
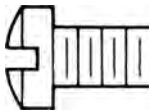
**FIGURE 51.5** Flat head.**FIGURE 51.6** Oval head.

The oval head is similar to a flat head except that instead of a flush surface it presents a low silhouette that improves the appearance (Figure 51.6).

The truss head is similar to the round head except that the head is shallower and has a larger diameter. It is used where extra bearing surface is required for extra holding power or where the clearance hole is oversized or the material is soft. It also presents a low silhouette that improves the appearance (Figure 51.7).

The binding head is similar to the pan head and is commonly used for electrical connections where an undercut is usually specified to bind and prevent the fraying of stranded wire (Figure 51.8).

The fillister head has the smallest diameter for a given shank size. It also has a deep slot that allows a higher torque to be applied during assembly. It is not as readily available or as widely used as some of the other head styles (Figure 51.9).

**FIGURE 51.7** Truss head.**FIGURE 51.8** Binding head.**FIGURE 51.9** Fillister head.

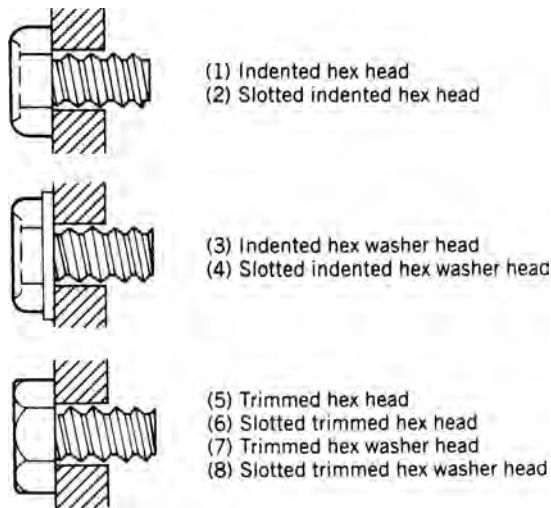


FIGURE 51.10 Hex head.

The advantages of a hex head are listed in the discussion on drivability. This type head is available in eight variations (Figure 51.10).

The indented design is lowest cost as the hex is completely cold upset in a counterbore die and possesses an identifying depression in the top surface of the head.

The trimmed design requires an extra operation to produce clean sharp corners with no indentation. Appearance is improved and there is no pocket on top to collect moisture.

The washer design has a larger bearing surface to spread the load over a wider area. The washer is an integral part of the head and also serves to protect the finish of the assembly from wrench disfigurement.

The slot is used to facilitate field service. It adds to the cost, weakens the head, and limits the amount of tightening torque that can be applied. A slot is unnecessary in high-production factory installation.

Any given location should standardize on one or possibly two of the eight variations.

Types of Screws

Machine Screws Machine screws are meant to be assembled in tapped holes, either into a product or into a nut. The screw threads of a machine screw are readily available in American National Standard Unified Inch Coarse and Fine Thread series. They are generally considered for applications where the material to be joined is too hard, too weak, too brittle, or too thick to take a tapping screw. It is also used in applications where the assembly requires a fastener made of a material that cannot be hardened enough to make its own thread, such as brass or nylon machine screws. Applications requiring freedom from dust or particles of any kind cannot use thread-cutting screws and, therefore, must be joined by machine screws or a tapping screw which forms or rolls a thread.

There are many combinations of head styles, shapes, and materials.

Self-Tapping Screws There are many different types of self-tapping screws commercially available. The following three types are capable of creating an internal thread

by being twisted into a smooth hole:

1. Thread-forming screws
2. Thread-cutting screws
3. Thread-rolling screws

The following two types create their own opening before generating the thread:

4. Self-drilling and tapping screws
5. Self-ex truding and tapping screws

1. **Thread-Forming Screws.** Thread-forming screws create an internal thread by forming or squeezing material. They rely on the pressure of the screw thread to force a mating thread into the workpiece. They are applicable in materials where large internal stresses are permissible or desirable to increase resistance to loosening. They are generally used to fasten sheet metal parts. They cannot be used to join brittle materials, such as plastics, because the stresses created in the workpiece can cause cracking. The following types of thread-forming screws are commonly used:

Types A and AB. Type AB screws have a spaced thread. This means that each thread is spaced further away from its adjacent thread than the popular machine screw series. They also have a gimlet point for ease in entering a predrilled hole. This type of screw is primarily intended to be used in sheet metal with a thickness from 0.015 in. (0.38 mm) to 0.05 in. (1.3 mm), resin-impregnated plywood, natural woods, and asbestos compositions.

Type AB screws were introduced several years ago to replace the type A screws. The type A screw is the same as the type AB except for a slightly wider spacing of the threads. Both are still available and can be used interchangeably. The big advantage of the type AB screw is that its threads are spaced exactly as the type B screws to be discussed later. In the interest of standardization it is recommended that type AB screws be used in place of either the type A or the type B series (Figure 51.11).

Type B. Type B screws have the same spacing as type AB screws. Instead of a gimlet point, they have a blunt point with incomplete threads at the point. This point makes the type B more suitable for thicker metals and blind holes. The type B screws can be used in any of the applications listed under type AB. In addition the type B screw can be used in sheet metal up to a thickness of 0.200 in. (5 mm) and in nonferrous castings (Figure 51.12).

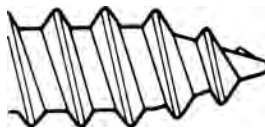


FIGURE 51.11 Type AB.

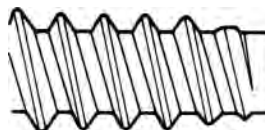


FIGURE 51.12 Type B.

Type C. Type C screws look like type B screws except that threads are spaced to be exactly the same as a machine screw thread and may be used to replace a machine screw in the field. They are recommended for general use in metal 0.030–0.100 in. (0.76–2.54 mm) thick. It should be recognized that in specific applications, involving long thread engagement or hard materials, this type of screw requires extreme driving torques.

2. Thread-Cutting Screws. Thread-cutting screws create an internal thread by actual removal of material from the internal hole. The design of the cavity to provide space for the chips and the design of the cutting edge differ with each type. They are used in place of the thread-forming type for applications in materials where disruptive internal stresses are undesirable or where excessive driving torques are encountered. The following types of thread-cutting screws are commonly used:

Type BT (Formerly Known as Type 25). Type BT screws have a spaced thread and a blunt point similar to the type B screw. In addition they have one cutting edge and a wide chip cavity. These screws are primarily intended for use in very friable plastics such as urea compositions, asbestos, and other similar compositions. In these materials, a larger space between threads is required to produce a satisfactory joint because it reduces the buildup of internal stresses that fracture brittle plastic when a closer spaced thread is used. The wide cutting slot creates a large cutting edge and permits rapid deflection of the chips to produce clean mating threads. For best results all holes should be counterbored to prevent fracturing the plastic. Use of this type screw eliminates the need to use tapped metallic inserts in plastic materials (Figure 51.13).

Type ABT. Type ABT screws are the same as type BT screws except that they have a gimlet point similar to a type AB screw. This design is not recognized as an American National Standard and should only be selected for large-volume applications (over 50,000 pieces of one size and type). It is primarily intended for use in plastic for the same reasons as listed for type BT screws (Figure 51.14).

Type D (Formerly Known as Type 1). Type D screws have threads of machine screw diameter-pitch combinations approximating unified form with a blunt point and tapered entering threads. In addition a slot is cut off center with one side on the

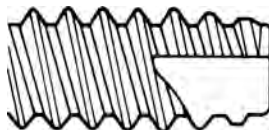


FIGURE 51.13 Type BT.



FIGURE 51.14 Type ABT.

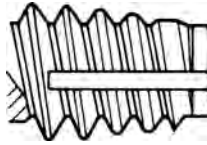


FIGURE 51.15 Type D.

center line. This radial side of the slot creates the sharp serrated cutting edge such as formed on a tap. The slot leaves a thinner section on one side of the screw that collapses and helps concentrate the pressure on the cutting edge. This screw is suitable for use in all thicknesses of metals (Figure 51.15).

Type F. Type F screws are identical to type D except that instead of one slot there are several slots cut at a slight angle to the axis of the thread. This screw is suitable for use in all thicknesses of metals and can be used interchangeably with a type D screw in many applications. However, the type F screw is superior to the type D screw for tapping into cast iron and permits the use of a smaller pilot hole (Figure 51.16).

Type D or Type F. Because in many applications these two types can be used interchangeably with the concomitant advantages of simpler inventory and increased availability, a combined specification is often issued permitting the supplier to furnish either type.

Type T (Formerly Known as Type 23). Type T screws are similar to type D and type F except that they have an acute rake angle cutting edge. The cut in the end of the screw is designed to eliminate a pocket that confines the chips. The shape of the slot is such that the chips are forced ahead of the screw as it is driven. This screw is suitable for plastics and other soft materials when a standard machine screw series thread is desired. It is used in place of type D and type F when more chip room is required because of deep penetration (Figure 51.17).

Type BF. Type BF screws are intended for use in plastics. The wide thread pitch reduces the buildup of internal stresses that fracture brittle plastics when a smaller thread pitch is used. The screw has a blunt point and tapered entering threads with several cutting edges and chip cavity (Figure 51.18).

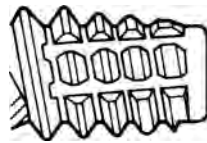


FIGURE 51.16 Type F.

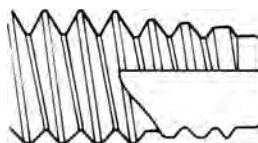


FIGURE 51.17 Type T.

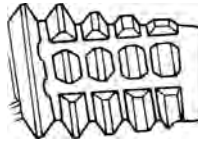


FIGURE 51.18 Type BF.

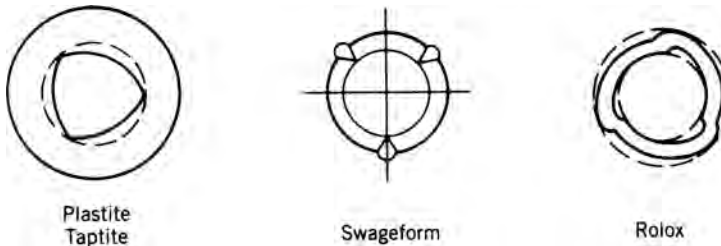


FIGURE 51.19 Thread-rolling screws.

3. **Thread-Rolling Screws.** Thread-rolling screws (see Figure 51.19) form an internal thread by flowing metal and thus do not cut through or disrupt the grain flow lines of materials as do thread-cutting screws. The screw compacts and work hardens the material, thereby forming strong, smoothly burnished internal threads. The screws have the threads of machine screw diameter–pitch combinations. This type screw is ideal for applications where chips can cause electrical shorting of equipment or jamming of delicate mechanism. Freedom from formation of chips eliminates the costly problem of cleaning the product of chips and burrs as would otherwise be required.

The ratio of driving torque to stripping torque is approximately 1:8 for a thread-rolling screw as contrasted to 1:3 for a conventional tapping screw. This higher ratio permits the driver torque release to be set well over the required driving torque and yet safely below the stripping torque. This increased ratio minimizes poor fastening due to stripped threads or inadequate seating of the screws.

Plastite is intended for use in filled or unfilled thermoplastics and some of the thermo-setting plastics. The other three types are intended for use in metals. At present, there are no data to prove the superiority of one type over another.

4. **Self-Drilling and Tapping Screws.** The self-drilling and tapping screw (Figure 51.20) drills its own hole and forms a mating thread, thus making a complete fastening in a single operation. Assembly labor is reduced by eliminating the need to predrill holes at assembly and by solving the problem of hole alignment. These screws must complete their metal-drilling function and fully penetrate the material before the screw thread can engage and begin its advancement. In order to meet this requirement, the unthreaded

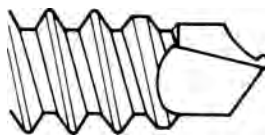


FIGURE 51.20 Self-drilling and tapping screws.

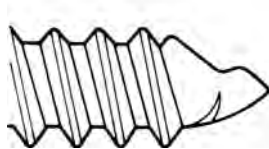


FIGURE 51.21 Self-extruding screw.

point length must be equal to or greater than the material thickness to be drilled. Therefore, there is a strict limitation on minimum and maximum material thickness that varies with screw size. There are many different styles and types of self-drilling and tapping screws to meet specific needs.

5. Self-Extruding Screws. Self-extruding screws provide their own extrusion as they are driven into an inexpensively produced punched hole. The resulting extrusion height is several times the base material thickness. This type screw is suitable for material in thicknesses up to 0.048 in. (1.2 mm). By increasing the thread engagement, these screws increase the differential between driving and stripping torque and provide greater pull-out strength. Since they do not produce chips, they are excellent for grounding sheet metal for electrical connections (Figure 51.21).

There is almost no limit to the variety of head styles, thread forms, and screw materials that are available commercially. The listing only shows representative examples. Users should attempt to keep varieties to a minimum by carefully selecting those variations that best meet the needs of their type of product.

Set Screws. Set screws are available in various combinations of head and point style as well as material and are used as locking, locating, and adjustment devices. The common head styles are slotted headless, square head, hexagonal socket, and fluted socket. The slotted headless has the lowest cost and can be used in a counterbored hole to provide a flush surface. The square head is applicable for location or adjustment of static parts where the projecting head is not objectionable. Its use should be avoided on all rotating parts. The hexagonal socket head can be used in a counterbored hole to provide a flush surface. It permits greater torque to be applied than with a slotted headless design. Fluted sockets are useful in very small diameters, that is, no. 6 (0.138 in.) and under, where hexagon keys tend to round out the socket in hexagonal socket set screws. Set screws should not be used to transmit large amounts of torque, particularly under shock torsion loads. Increased torsion loads may be carried by two set screws located 120° apart.

The following points are available with the head styles discussed: The cup point (Table 51.104) is the standard stock point for all head shapes and is recommended for all general locking purposes. Flats are recommended on round shafts when close fits are used and it is desirable to avoid interference in disassembling parts because of burrs produced by action of the cup point or when the flats are desired to increase torque transmission. When flats are not used, it is recommended that the minimum shaft diameter be not less than four times the cup diameter since otherwise the whole cup may not be in contact with

TABLE 51.104 Holding Power of Flat or Cup Point Set Screws

d (in.)	$\frac{1}{4}$	$\frac{5}{16}$	$\frac{3}{8}$	$\frac{7}{16}$	$\frac{1}{2}$	$\frac{9}{16}$	$\frac{5}{8}$	$\frac{3}{4}$	$\frac{7}{8}$	1	$1\frac{1}{8}$	$1\frac{1}{4}$
P (lb)	100	168	256	366	500	658	840	1280	1830	2500	3388	4198



FIGURE 51.22 Cup point.



FIGURE 51.23 Oval point.

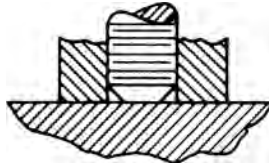


FIGURE 51.24 Flat point.

the shaft. The self-locking cup point has limited availability. It has counterclockwise knurls to prevent the screw from working loose even in poorly tapped holes (Figure 51.22).

When oval points are used, the surface it contacts should be grooved or spotted to the same general contour as the point to assure good seating. It is used where frequent adjustment is necessary without excessive deformation of the part against which it bears (Figure 51.23).

When flat points are used, it is customary to grind a flat on the shaft for better point contact. This point is preferred where wall thickness is thin and on top of plugs made of any soft material (Figure 51.24).

When the cone point is used, it is recommended that the angle of countersink be as nearly as possible the angle of screw point for the best efficiency. Cone point set screws have some application as pivot points. It is used where permanent location of parts is required. Because of penetration, it has the highest axial and torsional holding power of any point (Figure 51.25).



FIGURE 51.25 Cone point.

TABLE 51.105 Lag Screws

Diameter of screw (in.)	$\frac{1}{4}$	$\frac{5}{16}$	$\frac{3}{8}$	$\frac{7}{16}$	$\frac{1}{2}$	$\frac{5}{8}$	$\frac{3}{4}$	$\frac{7}{8}$	1
No. of threads per inch	10	9	7	7	6	5	$4\frac{1}{2}$	4	$3\frac{1}{2}$
Across flats of hexagon and square heads (in.)	$\frac{3}{8}$	$\frac{15}{32}$	$\frac{9}{16}$	$\frac{21}{32}$	$\frac{3}{4}$	$\frac{15}{16}$	$1\frac{1}{8}$	$1\frac{5}{16}$	$1\frac{1}{2}$
Thickness of hexagon and square heads (in.)	$\frac{3}{16}$	$\frac{1}{4}$	$\frac{5}{16}$	$\frac{3}{8}$	$\frac{7}{16}$	$\frac{17}{32}$	$\frac{5}{8}$	$\frac{3}{4}$	$\frac{7}{8}$

Length of Threads for Screws of All Diameters							
Length of screw (in.)	$1\frac{1}{2}$	2	$2\frac{1}{2}$	3	$3\frac{1}{2}$	4	$4\frac{1}{2}$
Length of thread (in.)	To head	$1\frac{1}{2}$	2	$2\frac{1}{4}$	$2\frac{1}{2}$	3	$3\frac{1}{2}$
Length of screw (in.)	5	$5\frac{1}{2}$	6	7	8	9	10–12
Length of thread (in.)	4	4	$4\frac{1}{2}$	5	6	6	7

The half-dog point should be considered in lieu of full-dog points when the usable length of thread is less than the nominal diameter. It is also more readily obtained than the full-dog point. It can be used in place of dowel pins and where end of thread must be protected (Figure 51.26).

Lag Screws. Lag screws (Table 51.105) are usually used in wood but also can be used in plastics and with expansion shields in masonry. A 60° gimlet point is the most readily available type. A 60° cone point, not covered in these drawings, is also available. Some suppliers refer to this item as a lag bolt (Figure 51.27).

A lag screw is normally used in wood when it is inconvenient or objectionable to use a through bolt and nut. To facilitate the insertion of the screw especially in denser types of wood, it is advisable to use a lubricant on the threads. It is important to have a pilot hole of proper size and following are some recommended hole sizes for commonly used types of wood. Hole sizes for other types of wood should be in proportion to the relative specific gravity of that wood to the ones listed in Table 51.106.

Shoulder Screws. These screws are also referred to as “stripper bolts.” They are used mainly as locators or retainers for spring strippers in punch and die operations and have



FIGURE 51.26 Half-dog point.

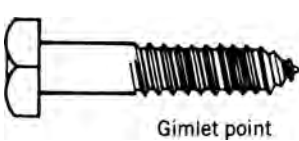


FIGURE 51.27 Lag screws.

TABLE 51.106 Recommended Diameters of Pilot Hole for Types of Wood^a

Screw Diameter (in.)	White Oak	Southern Yellow Pine, Douglas Fir	Redwood, Northern White Pine
0.250	0.160	0.150	0.100
0.312	0.210	0.195	0.132
0.375	0.260	0.250	0.180
0.438	0.320	0.290	0.228
0.500	0.375	0.340	0.280
0.625	0.485	0.437	0.375
0.750	0.600	0.540	0.480

^aPilot holes should be slightly larger than listed when lag screws of excessive lengths are to be used.

found some application as fulcrums or pivots in machine designs that involve links, levers, or other oscillating parts. Consideration should be given to the alternative use of a sleeve bearing and a bolt on the basis of both cost and good design (Figure 51.28).

Thumb Screws Thumb screws have a flattened head designed for manual turning without a driver or a wrench. They are useful in applications requiring frequent disassembly or screw adjustment (Figure 51.29).

Weld Screws Weld screws come in many different head configurations, all designed to provide one or more projections for welding the screw to a part.

Overhead projections are welded directly to the part. Underhead projections go through a pilot hole. The designs in Figures 51.30 and 51.31 are widely used.

In projection welding of carbon steel screws, care should be observed in applications, since optimum weldability is obtained when the sum, for either parent metal or screw, of one-fourth the manganese content plus the carbon content does not exceed 0.38. For good

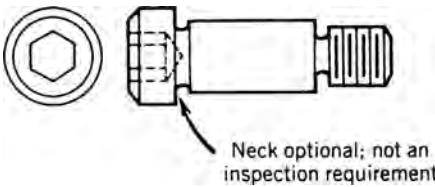


FIGURE 51.28 Shoulder screw.

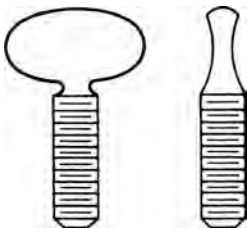


FIGURE 51.29 Thumb screws.

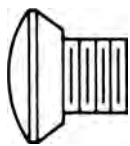


FIGURE 51.30 Single-projection weld screw.

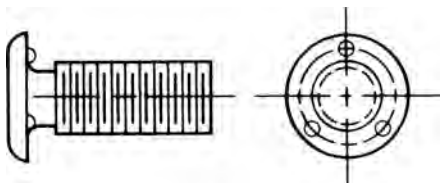


FIGURE 51.31 Underhead weld screws.

weldability with the annular ring type, the height of the weld projection should not exceed half the parent metal thickness as a rule of thumb.

Copper flash plating is provided for applications where cleanliness of the screw head is necessary in obtaining good welds.

Wood Screws Wood screws are (Table 51.107) readily available in lengths from $\frac{1}{4}$ to 5 in. for steel and from $\frac{1}{4}$ to $3\frac{1}{2}$ in. for brass. Consideration should be given to the use of type AB thread-forming screws, which are lower in cost and more efficient than wood screws for use in wood. Wood screws are made with flat, round, or oval heads.

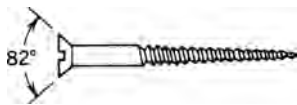
The *resistance of wood screws to withdrawal* from side grain of seasoned wood is given by the formula $P = 2850G^2D$, where P is the allowable load on the screw (lb/in. penetration of the threaded portion), G is specific gravity of oven-dry wood, and D is the diameter of the screw (in.). Wood screws should not be designed to be loaded in withdrawal from the end grain.

The *allowable safe lateral resistance* of wood screws embedded seven diameters in the side grain of seasoned wood is given by the formula $P = K D^2$, where P is the lateral resistance per screw (lb), D is the diameter (in.), and K is 4000 for oak (red and white), 3960 for Douglas fir (coast region) and southern pine, and 3240 for cypress (southern) and Douglas fir (inland region).

TABLE 51.107 American Standard Wood Screws^a

Number	0	1	2	3	4	5	6	7	8
Threads per inch	32	28	26	24	22	20	18	16	15
Diameter (in.)	0.060	0.073	0.086	0.099	0.112	0.125	0.138	0.151	0.164
Number	9	10	11	12	14	16	18	20	24
Threads per inch	14	13	12	11	10	9	8	8	7
Diameter (in.)	0.177	0.190	0.203	0.216	0.242	0.268	0.294	0.320	0.372

^aIncluded angle of flathead = 82°; see Figure 51.18.



The following rules should be observed: (a) The size of the lead hole in soft (hard) woods should be about 70% (90%) of the core or root diameter of the screw; (b) lubricants such as soap may be used without great loss in holding power; (c) long, slender screws are preferable generally, but in hardwood too slender screws may reach the limit of their tensile strength; and (d) in the screws themselves, holding power is favored by thin sharp threads, rough unpolished surface, full diameter under the head, and shallow slots.

SEMS The machine and tapping screws can be purchased with washers or lock washers as an integral part of the purchased screws. When thus joined together, the part is known as a SEMS unit. The washer is assembled on a headed screw blank before the threads are rolled. The inside diameter of the washer is of a size that will permit free rotation and yet prevent disassembly from the screw after the threads are rolled. If these screws and washers were purchased separately, there would be an initial cost savings over the pre-assembled units. However, these preassembled units reduce installation time because only one hand is needed to position them, leaving the other hand free to hold the driving tool. The time required to assemble a loose washer is eliminated. In addition, these assemblies act to minimize installation errors and inspection time because the washer is in place, correctly oriented. Also the use of a single unit, rather than two separate parts, simplifies bookkeeping, handling, inventory, and other related operations.

51.7.1 Nominal and Minimum Dressed Sizes of American Standard Lumber

Table 51.108 applies to boards, dimensional lumber, and timbers. The thicknesses apply to all widths and all widths to all thicknesses.

TABLE 51.108 Nominal and Minimum Dressed Sizes of American Standard Lumber

Item	Thicknesses			Face Widths		
	Nominal	Minimum Dressed		Nominal	Minimum Dressed	
		Dry ^a (in.)	Green (in.)		Dry ^a (in.)	Green (in.)
Boards ^b				2	1 1/2	1 9/16
				3	2 1/2	2 9/16
				4	3 1/2	3 9/16
				5	4 1/2	4 5/8
				6	5 1/2	5 5/8
	1	3/4	25/32	7	6 1/2	6 5/8
	1 1/4	1	1 1/32	8	7 1/4	7 1/2
	1 1/2	1 1/4	1 9/32	9	8 1/4	8 1/2
				10	9 1/4	9 1/2
				11	10 1/4	10 1/2
				12	11 1/4	11 1/2
				14	13 1/4	13 1/2
				16	15 1/4	15 1/2

(continued)

TABLE 51.108 (Continued)

Item	Thicknesses			Face Widths		
	Nominal	Minimum Dressed		Nominal	Minimum Dressed	
		Dry ^a (in.)	Green (in.)		Dry ^a (in.)	Green (in.)
Dimension				2	1 $\frac{1}{2}$	1 $\frac{9}{16}$
				3	2 $\frac{1}{2}$	2 $\frac{9}{16}$
				4	3 $\frac{1}{2}$	3 $\frac{9}{16}$
	2	1 $\frac{1}{2}$	1 $\frac{9}{16}$	5	4 $\frac{1}{2}$	4 $\frac{5}{8}$
	2 $\frac{1}{2}$	2	2 $\frac{1}{16}$	6	5 $\frac{1}{2}$	5 $\frac{5}{8}$
	3	2 $\frac{1}{2}$	2 $\frac{9}{16}$	8	7 $\frac{1}{4}$	7 $\frac{1}{2}$
	3 $\frac{1}{2}$	3	3 $\frac{1}{16}$	10	9 $\frac{1}{4}$	9 $\frac{1}{2}$
				12	11 $\frac{1}{4}$	11 $\frac{1}{2}$
				14	13 $\frac{1}{4}$	13 $\frac{1}{2}$
				16	15 $\frac{1}{4}$	15 $\frac{1}{2}$
				2	1 $\frac{1}{2}$	1 $\frac{9}{16}$
				3	2 $\frac{1}{2}$	2 $\frac{9}{16}$
				4	3 $\frac{1}{2}$	3 $\frac{9}{16}$
				5	4 $\frac{1}{2}$	4 $\frac{5}{8}$
Dimension	4	3 $\frac{1}{2}$	3 $\frac{9}{16}$	6	5 $\frac{1}{2}$	5 $\frac{5}{8}$
	4 $\frac{1}{2}$	4	4 $\frac{1}{16}$	8	7 $\frac{1}{4}$	7 $\frac{1}{2}$
				10	9 $\frac{1}{4}$	9 $\frac{1}{2}$
				12	11 $\frac{1}{4}$	11 $\frac{1}{2}$
				14		13 $\frac{1}{2}$
				16		15 $\frac{1}{2}$
Timbers	5 and thicker		$\frac{1}{2}$ off	5 and wider		$\frac{1}{2}$ off

^aMaximum moisture content of 19 % or less.^bBoards less than the minimum thickness for 1 in. nominal but $\frac{5}{8}$ in. or greater thickness dry ($\frac{11}{16}$ in. green) may be regarded as American Standard Lumber, but such boards shall be marked to show the size and condition of seasoning at the time of dressing. They shall also be distinguished from 1-in. boards on invoices and certificates.Source: From *American Softwood Lumber Standard*, NBS 20-70, National Bureau of Standards, Washington, DC, 1970, amended 1986 (available from Superintendent of Documents).

MEASUREMENT UNCERTAINTY

DAVID CLIPPINGER

- 52.1 Introduction
 - 52.1.1 Uncertainty as a property
 - 52.1.2 Misuse and inadequacy of significant digits
 - 52.1.3 Modern expressions of uncertainty
- 52.2 Literature
- 52.3 Evaluation of uncertainty
 - 52.3.1 ISO GUM methodology
 - 52.3.2 ASME performance test code methodology
- 52.4 Discussion
- References

52.1 INTRODUCTION

52.1.1 Uncertainty as a Property

The uncertainty of a measurement is a property of the measurement, and serves as an indication of the degree of reliability in the measurement or its usefulness in a given application (Taylor and Kuyatt, 1994). ISO (1995) rigorously defines uncertainty as “[the] parameter, associated with the result of a measurement, that characterizes the dispersion of values that could be reasonably attributed to [the particular quantity subject to measurement].”

For all but the crudest of applications, the uncertainty associated with a measurement is as important as the units of the measurement itself (i.e., kilogram, millimeter, etc.). That the units of a measurement should be reported with the measurement is beyond debate: there are numerous examples from history where a measured quantity was reported without its units, only to be mistaken by others for a value of a different dimension, often with disastrous results.

Despite its frequent omission, the uncertainty of a measurement is actually important in all situations where a measured quantity must be trusted with a certain level of

confidence. A manufacturing process that requires that a component have a mass of very close to 10 kg is one such example. Would it be acceptable to measure the mass of these objects with an inexpensive spring scale, or should a laboratory-calibrated electronic balance be used for such a measurement? The answer depends upon the definition of “very close” in the manufacturing specification, and the level of what is colloquially called the “accuracy” of the two mass measurement systems. To be more specific, the measured value must have an *uncertainty* that is less than the tolerance specified by the process. For example, when the mass is measured using the spring scale, it might be reasonable to attribute mass values between 9.73 kg and 10.27 kg to any mass that displays “10 kg” on the scale. In this case, the uncertainty of the measurement would be considered to be 0.27 kg. The acceptability of the spring scale as a mass measurement system can now be assessed quantitatively: if an error of 0.27 kg is “close enough” to 10 kg to satisfy requirements, the spring scale is perfectly satisfactory for this purpose.

Of course, the ability to make such a quantitatively based decision implies that the uncertainty of measurements made on the spring scale can be evaluated as 0.27 kg (or some other value) in the first place. Fortunately, the uncertainty of a measurement can be estimated with a high degree of confidence. Methods to do this are presented later in this chapter, after brief sections describing unacceptable and acceptable ways to report uncertainty, respectively.

52.1.2 Misuse and Inadequacy of Significant Digits

The concept of uncertainty is often first introduced to students couched in terms of *significance*, where the number of digits (or figures) in a numerical quantity is used to indicate the level of reliability of the quantity: the more digits past the decimal place, the more reliable (or precise) the measurement. It is understood that the maximum error associated with a quantity reported using significant digits is $\pm 1/2$ the right-most (smallest) digit. This is sometimes called *round-off error*.

However, significant digits are only valid when expressing the approximate value of exact quantities, and (perhaps surprisingly!) this validity does *not* extend to computations performed with these numbers (Bragg, 1974). Bragg (1974) uses computations with the value of π as an illustration, and the same will be done here. This quantity is often approximated as 3.14 (using three significant digits), which is taken to mean that the true value of π lies between 3.135 and 3.145 (i.e., ± 0.005). This is in fact the case, as $\pi = 3.141592654 \dots$

Now consider the computation π^2 . Performing the computation 3.14^2 results in the quantity 9.8596. Expressed to three significant digits, this quantity is 9.86, which would imply that the true value of π^2 should be somewhere between 9.855 and 9.865. Unfortunately it is not, (it is 9.869604401 \dots , or 9.87 to three significant digits), and while the error is arguably small (0.01 difference), the purpose of this illustration is to show that simply “carrying” significant digits through computational problems does not necessarily produce results with the same number of significant digits. It is sometimes proposed that carrying an “extra” significant digit through all computations and rounding off at the end will alleviate the problem described above. However, this too is erroneous. Such an approach is described in Hibbeler (2005), but is merely presented as the format for worked example problems, with no claims made about its validity as a computational practice. Kirkup and Frenkel (2006) list several “rules” for working with significant digits, at the same time observing that the use of significant digits represents no more than

“common sense” on the part of the experimentalist and cautioning that their use is “not a substitute for the detailed calculation of uncertainty.”

As an illustration, consider the computation π^{10} as an example. The “true” value of this quantity is 93648.047476 . . . (or 93.6×10^3 to three significant digits). However, if the four significant digit quantity 3.142 is used as approximation for π and carried through all computations, only rounding off to three significant digits at the very end, the resulting quantity is 93.8×10^3 . While this result is reasonably close to the true value, it does not meet the definition of a number with three significant digits: its quantity is not within ± 50 (0.05×10^3) of the true quantity.

On top of all this, Bragg (1974) also points out a serious limitation to the use of significant digits as a means of reporting the uncertainty of a quantity: such a method implies that the uncertainty of *every* number is simply $\pm \frac{1}{2}$ the right-most digit. Such an approach is excessively simplistic and assumes that the uncertainty of *experimentally* derived values can always be reduced to simple round-off error. While round-off error is valid when reporting the approximate value of exact numbers (such as π , e , or square roots of integers, all of which may be computed to a high degree of precision) it is not valid to consider that a *measurement* recorded to three decimal places is automatically accurate to ± 0.0005 . (If this were the case, why buy sophisticated test equipment when cheaper instruments could simply be fitted with extra-long digital displays!)

Returning to the spring scale example, it can be seen that an uncertainty of ± 0.27 kg cannot be readily expressed using significant digits alone. Instead, a more modern notation is required.

52.1.3 Modern Expressions of Uncertainty

As described above, the significant-digit method is very limited in its application and of little use to the experimentalist. Rather, the preferred methods today are the four that are contained in the ISO’s publication *Guide to the Expression of Uncertainty in Measurement* (1995). Briefly, they are as follows (paraphrased from the original, Section 7.2.2):

- (1) state the uncertainty in the text accompanying the measurement. For example, “mass=10.00 kg, with an uncertainty of 0.27 kg,”
- (2) provide the uncertainty in parentheses after the reported quantity, specifically stating that the digits used for the uncertainty are the right-most digits of the reported measurement. For example, “mass=10.00(27) kg, where the number in parentheses is the numerical value of the uncertainty corresponding to the last digits of the reported result,”
- (3) provide the uncertainty in parentheses after the reported quantity, specifically stating that the quantity in parentheses is the uncertainty of the measurement. For example, “mass=10.00(0.27) kg, where the number in parentheses is the uncertainty of the measurement,”
- (4) provide the uncertainty following a plus or minus sign that appears after the reported quantity, specifically stating that the quantity following the plus or minus is the uncertainty (and not a confidence interval). For example, “mass=(10.00 \pm 0.27) kg, where the number following the plus or minus is the uncertainty of the measurement, and not a confidence interval.”

A quick check of the engineering experimentation texts published by the leading textbook publishers in the United States seems to indicate that fourth style is the most popular for textbook use (the “not a confidence interval” clause is omitted in many textbooks). The fourth style is also used extensively in the otherwise “harmonized” ASME *Test Uncertainty* (1998). Ironically, ISO (1995) actually *discourages* this style, citing the ease with which this notation may be confused with that of a confidence interval (ISO makes special mention that the confusion may occur if the “not a confidence interval” clause is omitted).

ISO (1995) also provided further guidance for reporting uncertainties, such as specifying the coverage factor, the level of confidence used, and distribution used to obtain the coverage factor. These terms will be defined, and their application explained, in later sections of this chapter as the need appears.

52.2 LITERATURE

Uncertainties may be computed with a high degree of confidence with access to proper information. Methods for doing this may be found in several places, but the ISO (1995) *Guide to the Expression of Uncertainty in Measurement* (the “Guide” or GUM) cited earlier should be viewed as the authoritative reference for reporting uncertainty. The development of this reference was actually supported by a host of authoritative scientific and engineering organizations—BIPM, IEC, IFCC, IUPAC, IUPAP, and OIML—and is published in their name. As such it can be considered the “official” guide to those working in the fields of chemistry, physics, electrotechnology, and many others.

For engineers, ASME (1998) has harmonized their Performance Test Code *Test Uncertainty* (PTC 19.1-1998) with the ISO GUM, extending the guidance of the *Guide to the Expression of Uncertainty in Measurement* (ISO, 1995) to the Mechanical Engineering discipline as well. While “harmonized” with ISO the ASME test code does differ from the ISO guidelines in the actual methodology used to evaluate the overall uncertainty of a measurement, and has slight differences in terminology and notation. Both approaches are explained in this chapter and yield almost identical results.

NIST has adopted the ISO approach as standard practice in their Administrative Manual, Subchapter 4.09. This subchapter is included as an appendix to NIST’s *Technical Note 1297* (Taylor and Kuyatt, 1994). The NIST technical note (Taylor and Kuyatt, 1994) is particularly convenient since it is concise and (as a U.S. Government publication) accessible free of charge to anyone with an internet connection. (The ISO standard was actually published in 1993, and was corrected and re-printed in 1995, which is the version cited in this chapter. Taylor and Kuyatt (1994) reference the 1993 printing.) Organizations such as the Instrumentation, Systems, and Automation Society, NATO, and others have also adopted methods consistent with the ISO approach (Dieck, 2007).

In short, for engineers and scientists working in the United States, the ISO *Guide to the Expression of Uncertainty in Measurement* and the ASME *Test Uncertainty* (1998) are the de-facto “official” documents addressing the expression of uncertainty. Engineers or scientists working in other countries should consult their appropriate government ministry, professional society, or corporate leadership for clarification before using a standard other than the ISO *Guide* or one of its “harmonized” counterparts when evaluating measurement uncertainty.

52.3 EVALUATION OF UNCERTAINTY

52.3.1 ISO GUM Methodology

The ISO (1995) guidelines for the evaluation of uncertainty center on the evaluation of the various *standard uncertainties*, their addition into what is called *combined uncertainty*, and (where desired) the development of what is called *extended uncertainty*. The procedure described below is for cases where the measurements may be collected independently (i.e., the readings of different instruments are uncorrelated) and exhibit random errors that are approximately normally distributed. ISO (1995) should be consulted for special cases of correlated readings or asymmetrical error distributions.

52.3.1.1 Standard Uncertainties The current guidance from ISO (1995) describes that evaluation of the components of uncertainty in two methods, which it designates Types A and B. Type A refers to the “method of evaluation of uncertainty by the statistical analysis of a series of observations” (ISO, 1995). Type B refers to the “method of evaluation of uncertainty by means other than the statistical analysis of a series of observations” (ISO, 1995). Taylor and Kuyatt (1994) emphasize that these two methods are *not* the same as “random” and “systematic” components of error that are commonly described in other references on the subject (e.g., Wheeler and Ganji, 2010; also ASME, 1998). ISO’s guidance should not be taken as implying that measurement errors are no longer thought to have random and systematic components. Rather, the two types (A and B) refer to the means by which the component of the uncertainty is obtained for any given measurement, and are introduced for “convenience of discussion only” (ISO, 1995).

The key difference between types A and B uncertainties is that type A uncertainties are evaluated directly from data collected during the experiment, while type B uncertainties are evaluated from data collected beforehand or made available from some other means. Type A standard uncertainty is most commonly computed using the definition of standard deviation. Examples of sources of information from which type B uncertainty may be estimated include “previous measurement data, . . . manufacturer’s specifications, data provided in calibration and other certificates, uncertainties assigned to reference data taken from handbooks,” and from “general knowledge” of the experimental procedures and instruments used (ISO, 1995). Examples of types of uncertainty typically evaluated in a Type B manner include uncertainties due to hysteresis, readability, linearization of the sensor’s response, and resolution uncertainty, all of which are typically obtained prior to performing a measurement.

In both cases, the uncertainty so evaluated is designed u , multiple contributions to uncertainty typically indicated with subscripts (e.g., u_1 , u_2 , etc.). If it is a type A uncertainty, it is called a *Type A Standard Uncertainty*; otherwise it is called a *Type B Standard Uncertainty*. It is these standard uncertainties that are the building blocks that are used when evaluating what is called *combined uncertainty*.

As mentioned above, Type A standard uncertainty is usually the familiar statistical standard deviation, although other statistical means may be used to estimate its value (ISO, 1995). Typically, the standard uncertainty of the mean measurement is desired, in which case the standard deviation of the mean is the appropriate quantity to compute. In Equation (52.1), u is the standard uncertainty, x_i are the individual values of the measured quantity, n is the number of measurements, and \bar{x} is the

average value of the n measurements.

$$u = \sqrt{\frac{1}{n(n-1)} \sum_{i=1}^n (x_i - \bar{x})^2} \tag{52.1}$$

In contrast, Type B standard uncertainty may be obtained from a variety of different sources. It may be reported as a standard uncertainty (in which case it may be used as the value of u directly), or (more commonly) it may be obtained by dividing a reported confidence interval size by the appropriate *coverage factor*, assuming that the confidence interval is based on the standard normal distribution. For example, a manufacturer-supplied hysteresis uncertainty of ± 0.02 reported at the 95% confidence level is divided by 1.96 (the coverage factor or “z-value” for a 95% confidence, although 2 is also used as an approximation for this value) to obtain a Type B standard uncertainty of about 0.010 (to three significant digits). Taylor and Kuyatt (1994) describe situations where other techniques may be used, such as when a rectangular or triangular distribution more suitably describes the dispersion of values than the normal distribution.

A real measurement system will have multiple sources of uncertainty in both the type A and B categories. For example, a mass measurement using an electronic balance may (if taken a total of 10 times) return a mean measurement of 10.01 kg, with a standard deviation of the means (Type A standard uncertainty) of 0.022 kg, and have a manufacturer-specified (Type B) uncertainty due to linearization of the sensor response of 0.0003 kg. If the measurement device displays two digits past the decimal point, it will have a resolution uncertainty of 0.005.

It is often convenient to arrange these different standard uncertainties in a table. Two such tables are illustrated below. Table 52.1 shows a sample table for a mass measurement with an electronic balance with digital display. In this case the manufacturer’s data provides the linearization uncertainty, but no details as to its exact definition (i.e., whether the value provided is a standard uncertainty or a confidence interval). If the manufacturer cannot be contacted for clarification, one approach suggested by Wheeler and Ganji (2010) is to assume that the coverage factor is 2 for the linearization error since the manufacturer’s data was likely based on a large sample of normally-distributed data and likely reported at the 95% confidence level. Such an approach is used here. Table 52.2 shows a sample table for a power measurement in which the power consumed by a resistor is computed from its resistance and the voltage drop across it while in use. The resistance and voltage drop are both measured with a digital multimeter, but not simultaneously. This is

TABLE 52.1 Standard Uncertainties of a Mass Measurement

Name	Symbol	Source	Type	Value	Units	Coverage Factor	Standard Uncertainty
Standard deviation of the means	u_1	Experimental data	A	0.010	kg	1	0.01
Linearization error	u_2	Manufacturer’s data	B	0.0003	kg	2	0.0015
Resolution error	u_3	Inspection of display	B	0.005	kg	1.732 (assumed) ^a	0.0025

^aAssumes resolution uncertainty may be modeled with a rectangular distribution.

TABLE 52.2 Standard Uncertainties of a Power Measurement

Name	Symbol	Source	Type	Value	Units	Coverage Factor	Standard Uncertainty
Standard deviation of the means-voltage	u_{11}	Experimental data	A	0.010	V	1	0.01
Linearization uncertainty-voltage	u_{12}	Manufacture's data	B	0.0001	V	2 (assumed) ^a	0.0005
Resolution uncertainty-voltage	u_{13}	Inspection of display	B	0.005	V	1.732 (assumed) ^b	0
Standard deviation of the means-resistance	u_{21}	<i>A priori</i> experimental data	B	0.040	W	1	0.04
Linearization uncertainty-resistance	u_{22}	Manufacturer's data	B	0.0001	W	2 (assumed) ^a	0.0005
Resolution uncertainty-resistance	u_{23}	Inspection of display	B	0.050	W	1.732 (assumed) ^b	0.029

^aAssumes reported value is based on a large-sample normal distribution at 95% confidence.

^bAssumes resolution uncertainty may be modeled with a rectangular distribution.

appropriate since resistance readings cannot typically be directly obtained while a resistor is passing current; in this example the resistance readings were taken in advance, while the system was un-powered. Accordingly, the associated uncertainty is a Type B standard uncertainty (the actual classification of Type A or Type B ultimately has no bearing on the evaluation of uncertainty).

The conversion of resolution uncertainty to standard uncertainty deserves special mention. Wheeler and Ganji (2010) propose that resolution uncertainty be treated the same as a 95% confidence interval as an “arbitrary rule.” If this rule is adopted, and it is assumed the error associated with resolution uncertainty may be modeled as a normal distribution the resolution uncertainty should be divided by 2 (or 1.96) to obtain the standard uncertainty. However, resolution uncertainty (which is essentially due to round-off error of the display) is arguably best modeled as a random process with a rectangular (or uniform) probability distribution. If this model is adopted, the guidance from Taylor and Kuyatt (1994) applies, in that case the appropriate divisor to convert resolution uncertainty to a standard uncertainty is $\sqrt{3}$ (or about 1.732). Table 52.1 shows the resolution uncertainty of an electronic digital balance that displays two digits past the decimal point. Table 52.2 shows the resolution uncertainty of a multimeter that displays two digits past the decimal point when measuring voltage and one digit past the decimal when measuring resistance. In both tables resolution uncertainty is treated as if arising from a random process with a rectangular distribution.

52.3.1.2 Combined Uncertainty The individual standard uncertainties are then combined using the “root sum square” method of addition (ISO, 1995). In the simplest cases, where the contributions to the uncertainty all have the same dimension, (such as the example shown in Table 52.1) the individual uncertainties add as shown in Equation (52.2).

$$u_c = \sqrt{u_1^2 + u_2^2 + \dots} \quad (52.2)$$

In Equation (52.2), u_c is the *combined standard uncertainty*. The situation becomes more involved when the standard uncertainties have different dimensions. In these cases the experimental measurement is said to be a *result*. Numerous engineering measurements are actually results. For example, electrical resistance strain gage measurements are really millivolt measurements, which are then converted to strain with knowledge of the appropriate gage factor and excitation voltage. Hence, three independent factors—two voltages and a gage factor, all of which have uncertainties—are needed to compute strain. These different uncertainties are said to *propagate* to the uncertainty of the final result.

In such situations the individual standard uncertainties are multiplied by their appropriate *sensitivity coefficients* to compute the combined uncertainty. The sensitivity coefficients are the partial derivatives of the measurement function with respect to each of the variables that possess uncertainty, evaluated at the *operating point*, or mean value of the measured variables. To use the power analogy from Table 52.2, the appropriate expression for power is $P = V^2/R$. Therefore, the appropriate expression for the combined uncertainty of the power measurement is given by Equation (52.3), where bar overscores indicate average values of voltage (V) and resistance (R). The uncertainties u_1 , u_2 , and so on that appear in Equation (52.3) are the uncertainty values listed in Table 52.2. The minus sign that would normally appear in the denominator of the last three sensitivity coefficients has been dropped in Equation (52.3) for clarity (the minus signs ultimately drop out as all terms are squared).

$$u_c = \sqrt{\left(\frac{2\bar{V}}{\bar{R}}u_{11}\right)^2 + \left(\frac{2\bar{V}}{\bar{R}}u_{12}\right)^2 + \left(\frac{2\bar{V}}{\bar{R}}u_{13}\right)^2 + \left(\frac{\bar{V}^2}{\bar{R}}u_{21}\right)^2 + \left(\frac{\bar{V}^2}{\bar{R}}u_{22}\right)^2 + \left(\frac{\bar{V}^2}{\bar{R}}u_{23}\right)^2} \quad (52.3)$$

Note that the uncertainties with common first subscripts (the u_{1i} and u_{2i}) are dimensionally the same. It is sometimes convenient to add these using Equation (52.2) to obtain combined uncertainties u_1 , u_2 , and so on (without the double subscript), as shown in Equations (52.4a) and (52.4b).

$$u_1 = \sqrt{u_{11}^2 + u_{12}^2 + u_{13}^2} \quad (52.4a)$$

$$u_2 = \sqrt{u_{21}^2 + u_{22}^2 + u_{23}^2} \quad (52.4b)$$

The general form of Equation (52.3) is shown in Equation (52.5), where x_i represents the individual measured values that contribute to the result, X now represents an ordered n -tuple of the measured values (x_1, x_2, x_3, \dots), the u_i represent the n individual standard combined uncertainties (combined using Equation (52.2)), and bar overscores indicate average values.

$$u_c = \sqrt{\left(\frac{\partial R}{\partial x_1}\bigg|_{\bar{X}}u_1\right)^2 + \left(\frac{\partial R}{\partial x_2}\bigg|_{\bar{X}}u_2\right)^2 + \dots} \quad (52.5)$$

As a check to ensure the correct pairings of the partial derivatives with the appropriate uncertainties, it is helpful to verify that all of the squared terms inside the square root sign

of Equation (52.5) are dimensionally identical. This can be seen in Equation (52.3), where uncertainties u_{11} , u_{12} and u_{13} have the units “Volts,” and are multiplied by sensitivity coefficients that are dimensionally “Volts per Ohm” to yield “Volts squared per Ohm” or “Watts.” The remaining three uncertainties have units of “Ohms” are multiplied by sensitivity coefficients that are dimensionally “Volts squared per Ohm squared” to yield (as expected) terms that are dimensionally “Watts.”

52.3.1.3 Extended Uncertainties It is perfectly acceptable to simply report combined standard uncertainties in one of the formats shown in the section titled *Modern Expressions of Uncertainty* that appeared earlier in this chapter. However, in some situations it may be desirable to report an *extended uncertainty*. An extended uncertainty is a combined standard uncertainty multiplied by an appropriate coverage factor (ISO, 1995). Popular coverage factors are 2 (in which case the extended uncertainty represents approximately 95% of the dispersion of the values that may be reasonably attributed to the measured quantity) or 3 (where over 99% is represented). In other cases, a coverage factor between 2 and 3 may be desired. (A coverage factor of 1.65 is used if about 90% of the dispersion is desired to be represented). NIST uses a coverage factor of 2 in all cases, unless specific needs require otherwise (Taylor and Kuyatt, 1994). Examples of such situations include cases with small sample sizes where use of the “Student t ” distribution rather than the standard normal distribution is more appropriate. In these situations, the appropriate “ t -value” should be used as the coverage factor, taking care to report this (along with the number of degrees of freedom and confidence level used) in the text accompanying the measurement.

In all cases, the extended uncertainty is computed from the combined uncertainty as shown in Equation (52.6), where U is the extended uncertainty and k is the coverage factor (ISO, 1995).

$$U = ku_c \quad (52.6)$$

For the situations where a small sample size is used to compute some (or all) of the Type A uncertainties, a complication arises when selecting the appropriate number of degrees of freedom to use as the entering argument for a table of critical t -values. In these cases, the effective number of degrees of freedom v_{eff} may be estimated using the Welch–Satterthwaite formula, Equation (52.7),

$$v_{\text{eff}} = \frac{u_c^4}{\sum_{i=1}^N \frac{c_i^4 u_i^4}{v_i}} \quad (52.7)$$

where $c_i = \partial R / \partial x_i |_{\bar{x}}$ (i.e., the sensitivity coefficients), v_i are the degrees of freedom associated with the uncertainties u_i , and the number of effective degrees of freedom is less than the sum of the individual degrees of freedom (Taylor and Kuyatt, 1994), Equation (52.8). (Note that the sensitivity coefficients in Equation (52.7) are all equal to 1 if the measurement is not a result, as was the case in Table 52.1.)

$$v_{\text{eff}} \leq \sum_{i=1}^N v_i \quad (52.8)$$

A conservative approximation is to truncate (rather than round) the value of v_{eff} (i.e., to favor the lower number of degrees of freedom).

52.3.2 ASME Performance Test Code Methodology

52.3.2.1 Types of Uncertainty The approach favored by ASME (1998) and many textbooks (e.g., Figliola and Beasley, 2000 or Wheeler and Ganji, 2010) is to categorize sources of uncertainty into two types: *Systematic* and *Random*. As mentioned above, these terms are not synonymous with Types A and B uncertainties, but rather refer to the nature of the contribution to the overall uncertainty. If the source of uncertainty is expected to contribute to a scattering of the measurement values, it is considered to be a random contribution to uncertainty. If the contribution to the overall uncertainty is constant, it is considered to be systematic (Rabinovich, 1995). There may be multiple contributors to each of these types of uncertainty (Kirkup and Frenkel, 2006).

One way to distinguish between systematic and random sources of uncertainty is by determining if the overall uncertainty can be reduced by increasing the sample size. If the overall uncertainty may be reduced in this method, the uncertainty source is clearly random. Uncertainty that is always present and cannot be reduced by increasing the sample size (which effectively reduces the “scatter” in the estimate of the mean value) is considered systematic (Wheeler and Ganji, 2010). The presence of the systematic uncertainties makes increasing the sample size to huge values in an attempt to reduce uncertainty to zero not only impractical but futile (Kirkup and Frenkel, 2006). The systematic uncertainties can be thought of as forming a “floor,” below which the total uncertainty cannot be lowered due to the limitations of the measuring system.

Examples of systematic uncertainties are those due to the limitations of the calibration techniques, which merely reduce, but do not eliminate systematic uncertainty (Figliola and Beasley, 2000; also Rabinovich, 1995), uncertainties due to hysteresis and linearization of the sensor output (commonly called “linearity” uncertainty), and repeatability uncertainty. These three contributors to systematic uncertainty (and occasionally others) are sometimes combined and reported in manufacturers specifications as a sensor’s *accuracy* (Wheeler and Ganji, 2010).

ASME (1998) denotes random uncertainty with the variable P , and systematic uncertainty with the variable B , reflecting the older terms for these two components (“Precision” and “Bias” uncertainty, respectively). The ASME notation will also be used here, as will the terms random and systematic, consistent with the “harmonization” of ASME’s *Test Uncertainty* (1998) with ISO’s *Guide to the Expression of Uncertainty in Measurement* (1995).

52.3.2.2 Combined Systematic Uncertainty Even a simple measurement will typically have multiple contributions to the systematic uncertainty. These individual contributions may be combined using the root sum square method as shown in Equation (52.9),

$$B = \sqrt{b_1^2 + b_2^2 + b_3^2 + \dots} \quad (52.9)$$

where the b_i are the individual contributions to the systematic uncertainty and B is the total systematic uncertainty of the measured value. Care must be taken when combining the b_i to ensure that they are all evaluated at the same level of confidence (Wheeler and Ganji, 2010). For example, it is inconsistent to combine an uncertainty due to hysteresis that is provided at the 90% confidence level with an uncertainty due to linearity that is provided at the 95% confidence level. In the cases where the different components of systematic uncertainty are provided at different levels of

confidence, the 95% confidence level should be used (ASME, 1998). Uncertainties provided at other levels of confidence should be converted to standard uncertainties in the manner described in the ISO GUM Methodology section of this chapter, and then multiplied by 2 to obtain an approximate 95% confidence interval (ASME (1998) uses 2—not 1.96—as its 95% coverage factor).

In cases where the systematic uncertainty of a result is desired, the individual contributions to systematic uncertainty will be dimensionally different and have different “weights” in the computation of total systematic uncertainty. As was the case when combining standard uncertainties when using ISO GUM methodology, the systematic uncertainty components are multiplied by their corresponding sensitivity coefficients before combining, as shown in Equation (52.10),

$$B_R = \sqrt{\left(\frac{\partial R}{\partial x_1}\bigg|_{\bar{x}} B_1\right)^2 + \left(\frac{\partial R}{\partial x_2}\bigg|_{\bar{x}} B_2\right)^2 + \dots} \quad (52.10)$$

where B_R is the systematic uncertainty of the result, and the B_i are the overall systematic uncertainties of the different measurements that together contribute to the result, X is an ordered n -tuple of the various measured quantities, and bars indicate average values. For example, if the measurement of power dissipated by a resistor is desired to be computed from values of voltage and resistance using the relationship $R = V^2/R$, the expression for total systematic uncertainty of the power measurement is computed as shown in Equation (52.11),

$$B_{\text{Power}} = \sqrt{\left(\frac{2\bar{V}}{\bar{R}} B_V\right)^2 + \left(\frac{\bar{V}}{\bar{R}^2} B_{\text{Res}}\right)^2} \quad (52.11)$$

where bars indicate average values, B_V is the systematic uncertainty of the voltage reading and B_{Res} is the systematic uncertainty of the resistance reading. The required values of B_V and B_{Res} are obtained using Equation (52.9). As was the case when computing combined uncertainty of a result using ISO GUM methodology, each of the squared terms in Equation (52.11) can be seen to be dimensionally consistent: they all have units of power.

52.3.2.3 Combined Random Uncertainties For a direct measurement, the random uncertainty is the standard deviation of the means, shown as $S_{\bar{x}}$ in Equation (52.12). This equation is identical to Equation (52.1), but uses the notation from ASME (1998).

$$S_{\bar{x}} = \sqrt{\frac{1}{n(n-1)} \sum_{i=1}^n (x_i - \bar{x})^2} \quad (52.12)$$

In cases where the measurement is a single result, the individual standard deviations of the means are weighted by the appropriate sensitivity coefficients and combined using the root-sum-squares method, as shown in Equation (52.13),

$$S_R = \sqrt{\left(\frac{\partial R}{\partial x_1}\bigg|_{\bar{x}} S_1\right)^2 + \left(\frac{\partial R}{\partial x_2}\bigg|_{\bar{x}} S_2\right)^2 + \dots} \quad (52.13)$$

where S_R is the standard deviation of the result, the S_i are the standard deviations of the means of the individual measured parameters that contribute to the result (ASME, 1998), and \bar{X} represents the average value of an n -tuple of the measured parameters (i.e., $[\bar{x}_1, \bar{x}_2, \bar{x}_3 \dots]$). For example, to return to the example of a power measurement based on resistances and voltages, the standard deviation of the power dissipated by a resistor of average resistance \bar{R} and average voltage drop \bar{V} is computed as shown in Equation (52.14),

$$S_{\text{Power}} = \sqrt{\left(\frac{2\bar{V}}{\bar{R}}S_{\bar{V}}\right)^2 + \left(\frac{\bar{V}^2}{\bar{R}^2}S_{\bar{R}}\right)^2} \quad (52.14)$$

where bars indicate average values, $S_{\bar{V}}$ is the standard deviations of the means of the voltage (computed using Equation (52.12)) and $S_{\bar{R}}$ is the standard deviations of the means of the resistance, computed in identical manner. Again, the squared terms that appear in Equation (52.14) can all be seen to have the dimension of power.

52.3.2.4 Multiple Tests A special case exists for what ASME (1998) calls “multiple tests.” A multiple test is one where more than one set of measurements is made with the same instrument package. An important feature that must be present for this technique to be applied is a reasonable simultaneity of the measurements.

To illustrate the difference between a single result with multiple readings and a multiple test, again consider the measurement of power dissipated through a resistor. One such method to measure this power is the previously described measurement of voltage drop and resistance, and the computation of power dissipation from the relationship $R = V^2/R$. However, as mentioned earlier, the resistance measurements are typically taken in advance—while the system is in an unpowered state—and the voltage measurements are made afterward (the sequence could also be reversed, or the system could be repeatedly powered up and powered down to alternate voltage and resistance readings). However, regardless of the method chosen, no resistance measurement is ever collected at the same instant as any voltage measurement. The value of random uncertainty is then obtained from Equation (52.13). The net result is that there is one average value of voltage obtained (with an uncertainty) and one average value of resistance obtained (with its own uncertainty), which allows *one* value of average power to be computed (which has its own uncertainty).

Now consider an alternative instrumentation of the same measurement. In this case, the circuit is fitted with an ammeter and voltmeter, so that an amperage reading may be taken at the same instant as each of the voltage measurements (this is easily done with modern computer controlled data acquisition systems but could also be reasonably approximated by a trained observer provided that both measurements changed slow enough to be recorded manually with reasonable simultaneity). The data may then be arranged in a table and power computed for each (nearly) simultaneous set of measurements as shown in Table 52.3.

From Table 52.3, it can be seen that eight distinct values of power may be computed, which in turn have a mean and standard deviation. Such a test procedure provides the opportunity to consider the eight computed values of power as a sample of results, for which the mean and standard deviation of the result may be computed. In

TABLE 52.3 Simultaneous Voltage and Amperage Measurements

Reading Number	Voltage (V)	Amperage (A)	Power (W)
1	10.01	1.99	19.89
2	10.00	2.02	20.18
3	10.02	1.95	19.58
4	9.98	2.05	20.43
5	10.01	2.02	20.24
6	10.00	2.03	20.26
7	10.01	2.02	20.23
8	9.98	1.98	19.75
<i>Average power (W)</i>			20.07
<i>Standard deviation of power (W)</i>			0.29

short, since multiple individual results can be computed from multiple sets of simultaneously collected data, the results themselves can be treated much like individual measurements.

In these special cases, the standard deviation of the result is computed in a manner very similar to that used for the standard deviation of a direct measurement, Equation (52.12). Modifying the notation to reflect that the appropriate uncertainty is now that of a result, the appropriate computation is shown in Equation (52.15),

$$S_{\bar{R}} = \sqrt{\frac{1}{n(n-1)} \sum_{i=1}^N (R - \bar{R})^2} \quad (52.15)$$

where \bar{R} is the average value of the result, and $S_{\bar{R}}$ is the standard deviation of the means of the result.

From the above discussion it can be seen that ASME (1998) provides two methods to evaluate the uncertainty of a result: a method employing Equation (52.14) that is essentially identical to the ISO GUM method employing Equation (52.5) which is applicable to measurement procedures that produce a single result, and a method employing Equation (52.15) that treats multiple results as independent measurements.

52.3.2.5 Total Uncertainty The ASME (1998) method combines the systematic and random components of uncertainty to produce the *total uncertainty*. Note that since the systematic component of uncertainty was developed with the presumption of a 95% level of confidence, the random component needs to be treated in an identical manner. The value of total uncertainty is obtained from Equation (52.16),

$$U_{95} = 2 \sqrt{\left(\frac{B}{2}\right)^2 + S_{\bar{x}}^2} \quad (52.16)$$

where U_{95} is the total uncertainty expressed at the 95% level of confidence and B and $S_{\bar{x}}$ are defined in Equations (52.9) and (52.12), respectively.

In the case of a single result, the appropriate computation is given by Equation (52.17),

$$U_{95} = 2\sqrt{\left(\frac{B_R}{2}\right)^2 + S_R^2} \quad (52.17)$$

where B_R is computed using Equation (52.10) and S_R is computed using Equation (52.13). In the case of multiple results, Equation (52.17) is still used, but with S_R from Equation (52.15) taking the place of S_R .

52.4 DISCUSSION

As mentioned in the Section 52.1, uncertainty is a property of a measurement. Its value allows the engineer or scientist to quantify the expected level of dispersion in the quantity being measured. It is an important parameter for quantitatively determining the suitability of a measurement technique.

The two methods of evaluating or computing uncertainty presented here are consistent with each other while each possessing its own particular advantages.

The ISO approach has the advantage of ease of application and flexibility. All uncertainties are treated identically, allowing for simplicity in approach (the distinction between Types A and B uncertainties is only made for purposes of discussion). Indeed, combined uncertainty for any measurement can be computed with only two equations: Equations (52.1) and (52.4a). Further, the experimentalist may select an appropriate coverage factor k , or simply report the combined uncertainty u_c and allow the end user of the measurement to determine the suitability of the measurement. However, the use of the *standard uncertainty* as the building block of this approach requires that the user have at least a basic knowledge of statistics and probability distributions to properly compute this value, particularly in Type B cases where the uncertainty data may only be available in a nonstandard or ambiguous format.

The ASME approach requires that the sources of measurement uncertainty be sorted into systematic and random categories. There are also more equations involved, each applying to a specific measurement method. The coverage factor is set at 2, effectively fixing the level of confidence at 95%. One key advantage of the ASME approach is that it provides a simplified method for the computation of the random component of uncertainty in multiple-test cases. Another advantage, as noted in (Coleman and Steele, 1999), is that the ASME approach allows for the separate consideration of systematic and random components of uncertainty separately. As such, it allows the experimentalist to compare the relative contributions of each source of uncertainty, and make design or procedure decisions (such as to obtain a more reliable sensor to lower the systematic uncertainty or to proceed with the same sensor but to increase the sample size to reduce the random uncertainty).

DISCLAIMER

The views and opinions expressed in this chapter are entirely those of the author and are not to be considered official statements of policies of the U.S. Government or any of its agencies.

REFERENCES

- ASME, *Test Uncertainty: Performance Test Code 19.1-1998*, New York: The American Society of Mechanical Engineers; 1998.
- Bragg G. *Principles of Experimentation and Measurement*, New Jersey: Prentice Hall, Englewood Cliffs; 1974.
- Coleman H, Steele W. *Experimentation and Uncertainty Analysis for Engineers*, 2nd Edition, New York: John Wiley & Sons; 1999.
- Dieck R. *Measurement Uncertainty: Methods and Applications*, 4th Edition, Research Triangle Park, North Carolina: The Instrumentation, Systems, and Automation Society; 2007.
- Figliola R, Beasley D. *Theory and Design for Mechanical Measurement* 3rd Edition, New York: John Wiley and Sons; 2000.
- Hibbeler R. *Principles of Statics*, 10th Edition, New York: Prentice Hall; 2005.
- ISO, *Guide to the Expression of Uncertainty in Measurement*, Geneva: International Organization for Standardization; 1995.
- Kirkup L, Frenkel R. *An Introduction to Uncertainty in Measurement*, Cambridge: Cambridge University Press; 2006.
- Rabinovich S. *Measurement Errors: Theory and Practice*, New York: American Institute of Physics; 1995.
- Taylor B, Kuyatt C. *Guidelines for Evaluating and Expressing the Uncertainty of NIST Measurement Results*, NIST Technical Note 1297, National Institute of Standards and Technology. 1994.
- Wheeler, A, Ganji, A. *Introduction to Engineering Experimentation* 3rd Edition, New York: Prentice Hall; 2010.

53

MEASUREMENTS

E. L. HIXSON and E. A. RIPPERGER

- 53.1 Standards and accuracy
 - 53.1.1 Standards
 - 53.1.2 Accuracy and precision
 - 53.1.3 Sensitivity and resolution
 - 53.1.4 Linearity
- 53.2 Impedance concepts
- 53.3 Error analysis
 - 53.3.1 Internal estimates
 - 53.3.2 Use of normal distribution to calculate probable error in X
 - 53.3.3 External estimates
- References

53.1 STANDARDS AND ACCURACY

53.1.1 Standards

Measurement is the process by which a quantitative comparison is made between a standard and a measurand. The measurand is the particular quantity of interest—the thing that is to be quantified. The standard of comparison is of the same character as the measurand, and so far as mechanical engineering is concerned the standards are defined by law and maintained by the National Institute of Standards and Technology (NIST, formerly known as the National Bureau of Standards). The four independent standards which have been defined are length, time, mass, and temperature (Wildhack, 1961). All other standards are derived from these four. Before 1960, the standard for length was the international prototype meter, kept at Sevres, France. In 1960, the meter was redefined as 1,650,763.73 wavelengths of krypton light. Then in 1983, at the Seventeenth General Conference on Weights and Measures, a new standard was adopted: A meter is the distance traveled in a

vacuum by light in $1/299,792,458$ s (Giacomo, 1984). However, there is a copy of the international prototype meter, known as the national prototype meter, kept by NIST. Below that level, there are several bars known as national reference standards and below that there are the working standards. Interlaboratory standards in factories and laboratories are sent to NIST for comparison with the working standards. These interlaboratory standards are the ones usually available to engineers.

Standards for the other three basic quantities have also been adopted by NIST, and accurate measuring devices for those quantities should be calibrated against those standards.

The standard mass is a cylinder of platinum–iridium, the international kilogram, also kept at Sevres, France. It is the only one of the basic standards that is still established by a prototype. In the United States, the basic unit of mass is the basic prototype kilogram No. 20. Working copies of this standard are used to determine the accuracy of interlaboratory standards. Force is not one of the fundamental quantities, but in the United States the standard unit of force is the pound, defined as the gravitational attraction for a certain platinum mass at sea level and 45° latitude.

Absolute time, or the time when some event occurred in history, is not of much interest to engineers. Engineers are more likely to need to measure time intervals, that is, the time between two events. The basic unit for time measurements is the *second*. At one time, the second was defined as $1/86,400$ of the average period of rotation of the Earth on its axis, but that is not a practical standard. The period varies and the Earth is slowing up. Consequently a new standard based on the oscillations associated with a certain transition within the cesium atom was defined and adopted. That standard, the cesium clock, has now been superseded by the cesium fountain atomic clock as the primary time and frequency standard of the United States (NIST-F1 Cesium Fountain Atomic Clock). Although this cesium “clock” is the basic frequency standard, it is not generally usable by mechanical engineers. Secondary standards such as tuning forks, crystals, electronic oscillators, and so on are used, but from time to time access to time standards of a higher order of accuracy may be required. To help meet these requirements, NIST broadcasts 24 h per day, 7 days per week time and frequency information from radio stations WWV, WWVB, and WWVL located in Fort Collins, Colorado, and WWVH located in Hawaii. Other nations also broadcast timing signals. For details on the time signal broadcasts, potential users should consult NIST (NIST Time and Frequency Services, 2002).

Temperature is one of four fundamental quantities in the international measuring system. Temperature is fundamentally different in nature from length, time, and mass. It is an intensive quantity, whereas the others are extensive. Join together two bodies that have the same temperature and you will have a larger body at that same temperature. If you join two bodies which have a certain mass, you will have one body of twice the mass of the original body. Two bodies are said to be at the same temperature if they are in thermal equilibrium. The international practical temperature scale, adopted in 1990 (ITS-90) by the International Committee on Weights and Measurement is the one now in effect and the one with which engineers are primarily concerned. In this system, the kelvin (K) is the basic unit of temperature. It is $1/273.16$ of the temperature at the triple point of water, the temperature at which the solid, liquid, and vapor phases of water exist in equilibrium (Bentley, 1998). Degrees celsius ($^\circ\text{C}$) is related to kelvin by the equation

$$t = T - 273.15$$

where t is the degrees Celsius and T is the kelvin.

53.1.2 Accuracy and Precision

In measurement practice, four terms are frequently used to describe an instrument. They are accuracy, precision, sensitivity, and linearity. Accuracy, as applied to an instrument, is the closeness with which a reading approaches the true value. Since there is some error in every reading, the “true value” is never known. In the discussion of error analysis which follows, methods of estimating the “closeness” with which the determination of a measured value approaches the true value will be presented. Precision is the degree to which readings agree among themselves. If the same value is measured many times and all the measurements agree very closely, the instrument is said to have a high degree of precision. It may not, however, be a very accurate instrument. Accurate calibration is necessary for accurate measurement. Measuring instruments must, for accuracy, be from time to time compared to a standard. These will usually be laboratory or company standards which are in turn compared from time to time with a working standard at NIST. This chain can be thought of as the pedigree of the instrument, and the calibration of the instrument is said to be traceable to NIST.

53.1.3 Sensitivity and Resolution

These two terms, as applied to a measuring instrument, refer to the smallest change in the measured quantity to which the instrument responds. Obviously the accuracy of an instrument will depend to some extent on the sensitivity. If, for example, the sensitivity of a pressure transducer is 1 kPa, any particular reading of the transducer has a potential error of at least 1 kPa. If the readings expected are in the range of 100 kPa and a possible error of 1% is acceptable, then the transducer with a sensitivity of 1 kPa may be acceptable, depending on what other sources of error may be present in the measurement. A highly sensitive instrument is difficult to use. Therefore, a sensitivity significantly greater than that necessary to obtain the desired accuracy is no more desirable than one with insufficient sensitivity.

Many instruments today have digital readouts. For such instruments, the concepts of sensitivity and resolution are defined somewhat differently than they are for analog-type instruments. For example, the resolution of a digital voltmeter depends on the “bit” specification and the voltage range. The relationship between the two is expressed by the equation

$$R = \frac{V}{2^n},$$

where R is the resolution in volts, V is the voltage range, and n is the number of bits.

Thus, an 8-bit instrument on a 1-V scale would have a resolution of 1/256, or 0.004, volt. On a 10-V scale that would increase to 0.04 V. As in analog instruments, the higher the resolution, the more difficult it is to use the instrument, so if the choice is available, one should use the instrument which just gives the desired resolution and no more.

53.1.4 Linearity

The calibration curve for an instrument does not have to be a straight line. However, conversion from a scale reading to the corresponding measured value is most convenient if it can be done by multiplying by a constant rather than by referring to a nonlinear

calibration curve or by computing from an equation. Consequently instrument manufacturers generally try to produce instruments with a linear readout, and the degree to which an instrument approaches this ideal is indicated by its *linearity*. Several definitions of linearity are used in instrument specification practice (Doebelin, 2004). The so-called independent linearity is probably the most commonly used in specifications. For this definition, the data for the instrument readout versus the input are plotted and then a “best straight line” fit is made using the method of least squares. Linearity is then a measure of the maximum deviation of any of the calibration points from this straight line. This deviation can be expressed as a percentage of the actual reading or a percentage of the full-scale reading. The latter is probably the most commonly used, but it may make an instrument appear to be much more linear than it actually is. A better specification is a combination of the two. Thus, linearity equals $+A$ percent of reading or $+B$ percent of full scale, whichever is greater. Sometimes the term independent linearity is used to describe linearity limits based on actual readings. Since both are given in terms of a fixed percentage, an instrument with A percent proportional linearity is much more accurate at low reading values than an instrument with A percent independent linearity.

It should be noted that although specifications may refer to an instrument as having A percent linearity, what is really meant is A percent nonlinearity. If the linearity is specified as independent linearity, the user of the instrument should try to minimize the error in readings by selecting a scale, if that option is available, such that the actual reading is close to full scale. A reading should never be taken near the low end of a scale if it can possibly be avoided.

For instruments that use digital processing, linearity is still an issue since the analog-to-digital converter used can be nonlinear. Thus, linearity specifications are still essential.

53.2 IMPEDANCE CONCEPTS

Two basic questions which must be considered when any measurement is made are how has the measured quantity been affected by the instrument used to measure it? Is the quantity the same as it would have been had the instrument not been there? If the answers to these questions are no, the effect of the instrument is called *loading*. To characterize the loading, the concepts of *stiffness* and *input impedance* are used (Harris and Piersol, 2002). At the input of each component in a measuring system there exists a variable q_{i1} which is the one we are primarily concerned with in the transmission of information. At the same point, however, there is associated with q_{i1} another variable q_{i2} such that the product $q_{i1}q_{i2}$ has the dimensions of power and represents the rate at which energy is being withdrawn from the system. When these two quantities are identified, the generalized input impedance Z_{gi} can be defined by

$$Z_{gi} = \frac{q_{i1}}{q_{i2}} \quad (53.1)$$

if q_{i1} is an *effort variable*. The effort variable is also sometimes called the *across variable*. The quantity q_{i2} is called the *flow variable* or *through variable*. In the dynamic case, these variables can be represented in the frequency domain by their Fourier transform. Then the quantity Z is a complex number. The application of these concepts is illustrated by the example in Figure 53.1. The output of the linear network in the black box

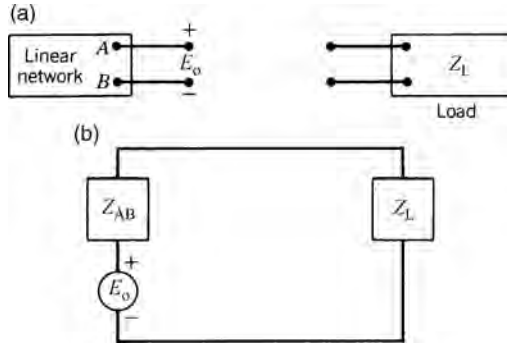


FIGURE 53.1 Application of Thévenin's theorem.

(Figure 53.1a) is the open-circuit voltage E_o until the load Z_L is attached across the terminals A–B. If Thévenin's theorem is applied after the load Z_L is attached, the system in Figure 53.1b is obtained. For that system the current is given by

$$i_m = \frac{E_o}{Z_{AB} + Z_L} \quad (53.2)$$

and the voltage E_L across Z_L is

$$E_L = i_m Z_L = \frac{E_o Z_L}{Z_{AB} + Z_L}$$

or

$$E_L = \frac{E_o}{1 + Z_{AB}/Z_L}. \quad (53.3)$$

Equations (53.2–53.7) are frequency-domain equations.

In a measurement situation, E_L would be the voltage indicated by the voltmeter, Z_L would be the input impedance of the voltmeter, and Z_{AB} would be the output impedance of the linear network. The true output voltage, E_o , has been reduced by the voltmeter, but it can be computed from the voltmeter reading if Z_{AB} and Z_L are known. From Equation (53.3) it is seen that the effect of the voltmeter on the reading is minimized by making Z_L as large as possible.

If the generalized input and output impedances Z_{gi} and Z_{go} are defined for nonelectrical systems as well as electrical systems, Equation (53.3) can be generalized to

$$q_{im} = \frac{q_{iu}}{1 + Z_{go}/Z_{gi}} \quad (53.4)$$

where q_{im} is the measured value of the effort variable and q_{iu} is the undisturbed value of the effort variable. The output impedance Z_{go} is not always defined or easy to determine; consequently Z_{gi} should be large. If it is large enough, knowing Z_{go} is unimportant.

If q_{i1} is a flow variable rather than an effort variable (current is a flow variable, voltage an effort variable), it is better to define an input admittance

$$Y_{gi} = \frac{q_{i1}}{q_{i2}} \quad (53.5)$$

rather than the generalized input impedance

$$Z_{gi} = \frac{\text{effort variable}}{\text{flow variable}}.$$

The power drain of the instrument is

$$P = q_{i1}q_{i2} = \frac{q_{i2}^2}{Y_{gi}}. \quad (53.6)$$

Hence, to minimize power drain, Y_{gi} must be large. For an electrical circuit

$$I_m = \frac{I_u}{1 + Y_o/Y_i}, \quad (53.7)$$

where I_m is the measured current, I_u is the actual current, Y_o is the output admittance of circuit, and Y_i is the input admittance of meter.

When the power drain is zero and the deflection is zero, as in structures in equilibrium, for example, when deflection is to be measured, the concepts of impedance and admittance are replaced with the concepts of *static stiffness* and *static compliance*. Consider the idealized structure in Figure 53.2.

To measure the force in member K_2 , an elastic link with a spring constant K_m is inserted in series with K_2 . This link would undergo a deformation proportional to the force in K_2 . If the link is very soft in comparison with K_1 , no force can be transmitted to K_2 . On the other hand, if the link is very stiff, it does not affect the force in K_2 but it will not provide a very good measure of the force. The measured variable is an effort variable, and in general, when it is measured, it is altered somewhat. To apply the impedance concept, a flow variable whose product with the effort variable gives power is selected. Thus,

$$\text{Flow variable} = \frac{\text{power}}{\text{effort variable}}.$$

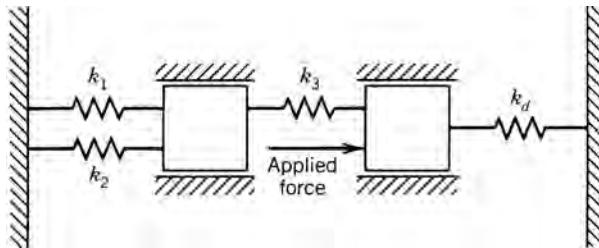


FIGURE 53.2 Idealized elastic structure.

Mechanical impedance is then defined as force divided by velocity, or

$$Z = \frac{\text{force}}{\text{velocity}},$$

where force and velocity are dynamic quantities represented by their Fourier transform and Z is a complex number. This is the equivalent of electrical impedance. However, if the static mechanical impedance is calculated for the application of a constant force, the impossible result

$$Z = \frac{\text{force}}{0} = \infty$$

is obtained.

This difficulty is overcome if energy rather than power is used in defining the variable associated with the measured variable. In that case, the static mechanical impedance becomes the *stiffness*:

$$\text{Stiffness} = S_g = \frac{\text{effort}}{\int \text{flow } dt}.$$

In structures

$$S_g = \frac{\text{effort variable}}{\text{displacement}}.$$

When these changes are made, the same formulas used for calculating the error caused by the loading of an instrument in terms of impedances can be used for structures by inserting S for Z . Thus

$$q_{im} = \frac{q_{iu}}{1 + S_{go}/S_{gi}}, \quad (53.8)$$

where q_{im} is the measured value of effort variable, q_{iu} is the undisturbed value of effort variable, S_{go} is the static output stiffness of measured system, and S_{gi} is the static stiffness of measuring system.

For an elastic force-measuring device such as a load cell S_{gi} is the spring constant K_m . As an example, consider the problem of measuring the reactive force at the end of a propped cantilever beam, as in Figure 53.3.

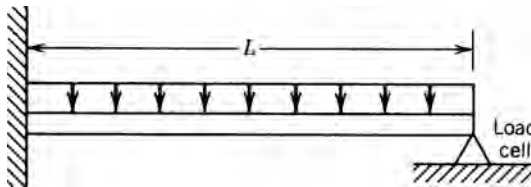


FIGURE 53.3 Measuring the reactive force at the tip.

According to Equation (53.8), the force indicated by the load cell will be

$$F_m = \frac{F_u}{1 + S_{go}/S_{gi}}$$

$$S_{gi} = K_m \quad \text{and} \quad S_{go} = \frac{3EI}{L^3}.$$

The latter is obtained by noting that the deflection at the tip of a tip-loaded cantilever is given by

$$\delta = \frac{PL^3}{3EI}.$$

The stiffness is the quantity by which the deflection must be multiplied to obtain the force producing the deflection.

For the cantilever beam

$$F_m = \frac{F_u}{1 + 3EI/K_m L^3} \quad (53.9)$$

or

$$F_u = F_m \left(1 + \frac{3EI}{K_m L^3} \right). \quad (53.10)$$

Clearly, if $K_m \gg 3EI/L^3$, the effect of the load cell on the measurement will be negligible.

To measure displacement rather than force, the concept of compliance is introduced and defined as

$$C_g = \frac{\text{flow variable}}{\int \text{effort variable } dt}.$$

Then

$$q_m = \frac{q_u}{1 + C_{go}/C_{gi}}. \quad (53.11)$$

If displacements in an elastic structure are considered, the compliance becomes the reciprocal of stiffness, or the quantity by which the force must be multiplied to obtain the displacement caused by the force. The cantilever beam in Figure 53.4 again provides a simple illustrative example.

If the deflection at the tip of this cantilever is to be measured using a dial gage with a spring constant K_m ,

$$C_{gi} = \frac{1}{K_m} \quad \text{and} \quad C_{go} = \frac{L^3}{3EI}.$$

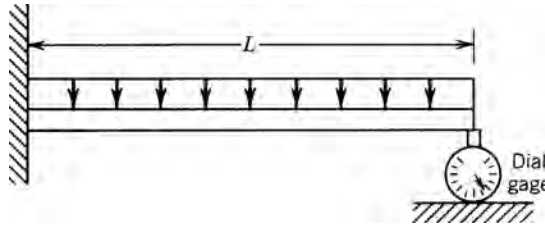


FIGURE 53.4 Measuring the tip deflection.

Thus,

$$\delta_m = \delta_u \left(1 + \frac{K_m L^3}{3EI} \right) \quad (53.12)$$

Not all interactions between a system and a measuring device lend themselves to this type of analysis. A pitot tube, for example, inserted into a flow field distorts the flow field but does not extract energy from the field. Impedance concepts cannot be used to determine how the flow field will be affected.

There are also applications in which it is not desirable for a force-measuring system to have the highest possible stiffness. A subsoil pressure gage is an example. Such a gage, if it is much stiffer than the surrounding soil, will take a disproportionate share of the total load and will consequently indicate a higher pressure than would have existed in the soil if the gage had not been there.

53.3 ERROR ANALYSIS

It may be accepted as axiomatic that there will always be errors in measured values. Thus, if a quantity X is measured, the correct value q , and X will differ by some amount e . Hence,

$$\pm(q - X) = e$$

or

$$q = X \pm e. \quad (53.13)$$

It is essential, therefore, in all measurement work that a realistic estimate of e be made. Without such an estimate the measurement of X is of no value. There are two ways of estimating the error in a measurement. The first is the external estimate, or ϵ_E , where $\epsilon = e/q$. This estimate is based on knowledge of the experiment and measuring equipment and to some extent on the internal estimate ϵ_I .

The internal estimate is based on an analysis of the data using statistical concepts.

53.3.1 Internal Estimates

If a measurement is repeated many times, the repeat values will not, in general, be the same. Engineers, it may be noted, do not usually have the luxury of repeating

measurements many times. Nevertheless the standardized means for treating results of repeated measurements are useful, even in the error analysis for a single measurement (Cook and Rabinowicz, 1963).

If some quantity is measured many times and it is assumed that the errors occur in a completely random manner, that small errors are more likely to occur than large errors, and that errors are just as likely to be positive as negative, the distribution of errors can be represented by the curve

$$F(X) = \frac{Y_0 e^{-(X-U)}}{2\sigma^2}, \quad (53.14)$$

where $F(X)$ is the number of measurements for a given value of $(X - U)$, Y_0 is the maximum height of curve or number of measurements for which $X = U$, and U is the value of X at point where maximum height of curve occurs σ determines lateral spread of the curve.

This curve is the normal, or Gaussian, frequency distribution. The area under the curve between X and δX represents the number of data points which fall between these limits and the total area under the curve denotes the total number of measurements made. If the normal distribution is defined so that the area between X and $X + \delta X$ is the probability that a data point will fall between those limits, the total area under the curve will be unity and

$$F(X) = \frac{\exp - (X - U)^2 / 2\sigma^2}{\sigma\sqrt{2\Pi}} \quad (53.15)$$

and

$$P_x = \int \frac{\exp - (X - U)^2 / 2\sigma^2}{\sigma\sqrt{2\Pi}} dx. \quad (53.16)$$

Now if U is defined as the average of all the measurements and s as the standard deviation,

$$\sigma = \left[\frac{\sum (X - U)^2}{N} \right]^{1/2}, \quad (53.17)$$

where N is the total number of measurements. Actually this definition is used as the best estimate for a universe standard deviation, that is, for a very large number of measurements. For smaller subsets of measurements the best estimate of σ is given by

$$\sigma = \left(\frac{\sum (X - U)^2}{n - 1} \right)^{1/2}, \quad (53.18)$$

where n is the number of measurements in the subset. Obviously, the difference between the two values of σ becomes negligible as n becomes very large (or as $n \rightarrow N$).

The probability curve based on these definitions is shown in Figure 53.5.

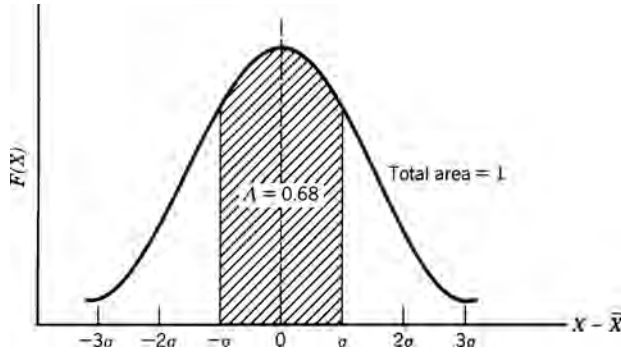


FIGURE 53.5 Probability curve.

The area under this curve between $-\sigma$ and $+\sigma$ is 0.68. Hence, 68% of the measurements can be expected to have errors that fall in the range of $\pm\sigma$. Thus, the chances are 68/32, or better than 2 to 1, that the error in a measurement will fall in this range. For the range $\pm 2\sigma$ the area is 0.95. Hence, 95% of all the measurement errors will fall in this range and the odds are about 20:1 that a reading will be within this range. The odds are about 384:1 that any given error will be in the range of $\pm 3\sigma$.

Some other definitions related to the normal distribution curve are as follows:

1. *Probable Error*: The error likely to be exceeded in half of all the measurements and not reached in the other half of the measurements. This error in Figure 53.5 is about 0.67σ .
2. *Mean Error*: The arithmetic mean of all the errors regardless of sign. This is about 0.8σ .
3. *Limit of Error*: The error that is so large it is most unlikely ever to occur. It is usually taken as 4σ .

53.3.2 Use of Normal Distribution to Calculate Probable Error in X

The foregoing statements apply strictly only if the number of measurements is very large. Suppose that n measurements have been made. That is a sample of n data points out of an infinite number. From that sample U and σ are calculated as above. How good are these numbers? To determine that, we proceed as follows. Let

$$U = F(X_1, X_2, X_3, \dots, X_n) = \frac{\sum X_i}{n} \quad (53.19)$$

$$e_u = \sum \frac{\partial F}{\partial X_i} e_{xi}, \quad (53.20)$$

where e_u is the error in U , e_{xi} is the error in X_i

$$(e_u)^2 = \sum \left(\frac{\partial F}{\partial X_i} e_{xi} \right)^2 + \sum \left(\frac{\partial F}{\partial X_i} e_{xi} \right) \left(\frac{\partial F}{\partial X_j} e_{xj} \right), \quad (53.21)$$

where $I \neq j$. If the errors e_i to e_n are independent and symmetrical, the cross-product terms will tend to disappear and

$$(e_u)^2 = \sum \left(\frac{\partial F}{\partial X_i} e_{xi} \right)^2. \quad (53.22)$$

Since $\partial F / \partial X_i = 1/n$,

$$e_u = \left[\sum \left(\frac{1}{n} \right)^2 e_{xi}^2 \right]^{1/2} \quad (53.23)$$

or

$$e_u = \left[\left(\frac{1}{n} \right)^2 \sum (e_{xi})^2 \right]^{1/2} \quad (53.24)$$

from the definition of σ

$$\sum (e_{xi})^2 = n\sigma^2 \quad (53.25)$$

and

$$e_u = \frac{\sigma}{\sqrt{n}}.$$

This equation must be corrected because the real errors in X are not known. If the number n were to approach infinity, the equation would be correct. Since n is a finite number, the corrected equation is written as

$$e_u = \frac{\sigma}{(n-1)^{1/2}} \quad (53.26)$$

and

$$q = U \pm \frac{\sigma}{(n-1)^{1/2}}. \quad (53.27)$$

This says that if one reading is likely to differ from the true value by an amount σ , then the average of 10 readings will be in error by only $\sigma/3$ and the average of 100 readings will be in error by $\sigma/10$. To reduce the error by a factor of 2, the number of readings must be increased by a factor of 4.

53.3.3 External Estimates

In almost all experiments several steps are involved in making a measurement. It may be assumed that in each measurement there will be some error, and if the measuring devices

are adequately calibrated, errors are as likely to be positive as negative. The worst condition insofar as accuracy of the experiment is concerned would be for all errors to have the same sign. In that case, assuming the errors are all much less than 1, the resultant error will be the sum of the individual errors, that is

$$\epsilon_E = \epsilon_1 + \epsilon_2 + \epsilon_3 + \cdots \quad (53.28)$$

It would be very unusual for all errors to have the same sign. Likewise, it would be very unusual for the errors to be distributed in such a way that

$$\epsilon_E = 0.$$

A general method follows for treating problems that involve a combination of errors to determine what error is to be expected as a result of the combination.

Suppose that

$$V = F(a, b, c, d, e, \cdots, x, y, z), \quad (53.29)$$

where a, b, c, \cdots, x, y, z represent quantities which must be individually measured to determine V . Then

$$\delta V = \sum \left(\frac{\partial F}{\partial n} \right) \delta n$$

and

$$\epsilon_E = \sum \left(\frac{\partial F}{\partial n} \right) e_n. \quad (53.30)$$

The sum of the squares of the error contributions is given by

$$e_E^2 = \left[\sum \left(\frac{\partial F}{\partial n} \right) e_n \right]^2. \quad (53.31)$$

Now, as in the discussion of internal errors, assume that errors e_n are independent and symmetrical. This justifies taking the sum of the cross products as zero:

$$\sum \left(\frac{\partial F}{\partial n} \right) \left(\frac{\partial F}{\partial m} \right) e^n e^m = 0 \quad n \neq m. \quad (53.32)$$

Hence,

$$(\epsilon_E)^2 = \sum \left(\frac{\partial F}{\partial n} \right)^2 e_n^2$$

or

$$e_E = \left[\sum \left(\frac{\partial F}{\partial n} \right)^2 e_n^2 \right]^{1/2}. \quad (53.33)$$

This is the *most probable value* of e_E . It is much less than the worst case:

$$\epsilon_e = [|\epsilon_a| + |\epsilon_b| + |\epsilon_c| \cdots + |\epsilon_z|]. \quad (53.34)$$

As an application, the determination of g , the local acceleration of gravity, by use of a simple pendulum will be considered:

$$g = \frac{4\Pi^2 L}{T^2}, \quad (53.35)$$

where L is the length of pendulum and T is the period of pendulum.

If an experiment is performed to determine g , the length L and the period T would be measured. To determine how the accuracy of g will be influenced by errors in measuring L and T write

$$\frac{\partial g}{\partial L} = \frac{4\Pi^2}{T^2} \quad \text{and} \quad \frac{\partial g}{\partial T} = \frac{-8\Pi^2 L}{T^3}. \quad (53.36)$$

The error in g is the variation in g written as follows:

$$\delta g = \left(\frac{\partial g}{\partial L} \right) \Delta L + \left(\frac{\partial g}{\partial T} \right) \Delta T \quad (53.37)$$

or

$$\delta g = \left(\frac{4\Pi^2}{T^2} \right) \Delta L - \left(\frac{8\Pi^2 L}{T^3} \right) \Delta T. \quad (53.38)$$

It is always better to write the errors in terms of percentages. Consequently Equation (53.38) is rewritten as

$$\delta g = \frac{(4\Pi^2 L/T^2) \Delta L}{L} - \frac{2(4\Pi^2 L/T^2) \Delta T}{T} \quad (53.39)$$

or

$$\frac{\delta g}{g} = \frac{\Delta L}{L} - \frac{2\Delta T}{T} \quad (53.40)$$

Then

$$e_g = [e_L^2 + (2e_T)^2]^{1/2}, \quad (53.41)$$

where e_g is the *most probable error* in the measured value of g . That is,

$$g = \frac{4\pi^2 L}{T^2} \pm e_g, \quad (53.42)$$

where L and T are the measured values. Note that even though a positive error in T causes a negative error in the calculated value of g , the contribution of the error in T to the most probable error is taken as positive. Note also that an error in T contributes four times as much to the most probable error as an error in L contributes. It is fundamental in measurements of this type that those quantities which appear in the functional relationship raised to some power greater than unity contribute more heavily to the most probable error than other quantities and must, therefore, be measured with greater care.

The determination of the most probable error is simple and straightforward. The question is how are the errors, such as $\Delta L/L$ and $\Delta T/T$, determined. If the measurements could be repeated often enough, the statistical methods discussed in the internal error evaluation could be used to arrive at a value. Even in that case it would be necessary to choose some representative error such as the standard deviation or the mean error. Unfortunately, as was noted previously, in engineering experiments it usually is not possible to repeat measurements enough times to make statistical treatments meaningful. Engineers engaged in making measurements will have to use what knowledge they have of the measuring instruments and the conditions under which the measurements are made to make a reasonable estimate of the accuracy of each measurement. When all of this has been done and a most probable error has been calculated, it should be remembered that the result is not the actual error in the quantity being determined but is, rather, the engineer's best estimate of the magnitude of the uncertainty in the final result (Kline and McClintock, 1953; Taylor and Kuyatt, 1994).

Consider again the problem of determining g . Suppose that the length L of the pendulum has been determined by means of a meter stick with 1-mm calibration marks and the error in the calibration is considered negligible in comparison with other errors. Suppose the value of L is determined to be 91.7 cm. Since the calibration marks are 1 mm apart, it can be assumed that ΔL is no greater than 0.5 mm. Hence the maximum

$$\frac{\Delta L}{L} = 5.5 \times 10^{-4}.$$

Suppose T is determined with the pendulum swinging in a vacuum with an arc of $\pm 5^\circ$ using a stopwatch that has an inherent accuracy of one part in 10,000. (If the arc is greater than $\pm 5^\circ$, a nonisochronous swing error enters the picture.) This means that the error in the watch reading will be no more than 10^{-4} s. However, errors are introduced in the period determination by human error in starting and stopping the watch as the pendulum passes a selected point in the arc. This error can be minimized by selecting the highest point in the arc because the pendulum has zero velocity at that point and timing a large number of swings so as to spread the error out over that number of swings. Human reaction time may vary from as low as 0.2 s to as high as 0.7 s. A value of 0.5 s will be assumed. Thus the estimated maximum error in starting and stopping the watch will be 1 s (± 0.5 s at the start and ± 0.5 s at the stop). A total of 100 swings will be timed. Thus, the estimated maximum error in the period will be 1/100 s. If the period is determined to be 1.92 s, the estimated maximum error will be $0.01/1.92 = 0.005$. Compared to this, the

error in the period due to the inherent inaccuracy of the watch is negligible. The nominal value of g calculated from the measured values of L and T is 982.03 cm/s^2 . The most probable error (Equation 53.29) is

$$\left[4(0.005)^2 + (5.5 \times 10^{-4})^2\right]^{1/2} = 0.01. \quad (53.43)$$

The uncertainty in the value of g is then $\pm 9.82 \text{ cm/s}^2$, or in other words the value of g will be somewhere between 972.21 and 991.85 cm/s^2 .

Often it is necessary for the engineer to determine in advance how accurately the measurements must be made in order to achieve a given accuracy in the final calculated result. For example, in the pendulum problem it may be noted that the contribution of the error in T to the most probable error is more than 300 times the contribution of the error in the length measurement. This suggests, of course, that the uncertainty in the value of g could be greatly reduced if the error in T could be reduced. Two possibilities for doing this might be (1) find a way to do the timing that does not involve human reaction time or (2) if that is not possible, increase the number of cycles timed. If the latter alternative is selected and other factors remain the same, the error in T timed over 200 swings is $1/200$ or 0.005 , second. As a percentage the error is $0.005/1.92 = 0.0026$. The most probable error in g then becomes

$$e_g = [4 \times (2.6 \times 10^{-3})^2 + (5.5 \times 10^{-4})^2]^{1/2} = 0.005. \quad (53.44)$$

This is approximately half of the most probable error in the result obtained by timing just 100 swings. With this new value of e_g the uncertainty in the value of g becomes $\pm 4.91 \text{ cm/s}^2$ and g then can be said to be somewhere between 977.12 and 986.94 cm/s^2 . The procedure for reducing this uncertainty still further is now self-evident.

Clearly, the value of this type of error analysis depends on the skill and objectivity of the engineer in estimating the errors in the individual measurements. Such skills are acquired only by practice and careful attention to all the details of the measurements.

REFERENCES

- Wildhack WA. NBS source of American standards. *ISA Journal* 1961;8(2).
- Giacomo P. News from the IBPM. *Metrologia* 1984;20(1):171.
- NIST Time and Frequency Services, NIST Special Publication 432;2002.
- Bentley RE. editor. *Handbook of Temperature Measurement*. CSIRO Springer; 1998.
- Doebelin EA. *Measurement Systems—Application and Design*. 5th ed. New York: McGraw Hill; 2004. p 85–91.
- Harris CM, Piersol AG. Mechanical impedance. In: *Shock and Vibration Handbook*, Chap. 10. 5th ed. New York: McGraw-Hill; 2002. p 10.1–10.14.
- Cook NH, Rabinowicz E. *Physical Measurement and Analysis*. Reading (MA): Addison Wesley; 1963. p 29–68.
- Kline SJ, McClintock FA. Describing uncertainties in single sample experiments. *Mechanical Engineering* 1953;75(3):3–8.
- Taylor BN, Kuyatt CE. Guidelines for Evaluating and Expressing the Uncertainty of NIST Measurement Results, NIST Technical Note 1297;1994.



HANDBOOK OF MEASUREMENT

IN SCIENCE AND ENGINEERING

Volume 3



EDITED BY

MYER KUTZ



HANDBOOK OF MEASUREMENT IN SCIENCE AND ENGINEERING

Volume 3

Edited by

MYER KUTZ

WILEY

Copyright © 2016 by John Wiley & Sons, Inc. All rights reserved

Published by John Wiley & Sons, Inc., Hoboken, New Jersey
Published simultaneously in Canada

No part of this publication may be reproduced, stored in a retrieval system, or transmitted in any form or by any means, electronic, mechanical, photocopying, recording, scanning, or otherwise, except as permitted under Section 107 or 108 of the 1976 United States Copyright Act, without either the prior written permission of the Publisher, or authorization through payment of the appropriate per-copy fee to the Copyright Clearance Center, Inc., 222 Rosewood Drive, Danvers, MA 01923, (978) 750-8400, fax (978) 750-4470, or on the web at www.copyright.com. Requests to the Publisher for permission should be addressed to the Permissions Department, John Wiley & Sons, Inc., 111 River Street, Hoboken, NJ 07030, (201) 748-6011, fax (201) 748-6008, or online at <http://www.wiley.com/go/permissions>.

Limit of Liability/Disclaimer of Warranty: While the publisher and author have used their best efforts in preparing this book, they make no representations or warranties with respect to the accuracy or completeness of the contents of this book and specifically disclaim any implied warranties of merchantability or fitness for a particular purpose. No warranty may be created or extended by sales representatives or written sales materials. The advice and strategies contained herein may not be suitable for your situation. You should consult with a professional where appropriate. Neither the publisher nor author shall be liable for any loss of profit or any other commercial damages, including but not limited to special, incidental, consequential, or other damages.

For general information on our other products and services or for technical support, please contact our Customer Care Department within the United States at (800) 762-2974, outside the United States at (317) 572-3993 or fax (317) 572-4002.

Wiley also publishes its books in a variety of electronic formats. Some content that appears in print may not be available in electronic formats. For more information about Wiley products, visit our web site at www.wiley.com.

Library of Congress Cataloging-in-Publication Data:

Handbook of Measurement in Science and Engineering / Myer Kutz, editor.
volumes cm

Includes bibliographical references and index.

ISBN 978-0-470-40477-5 (volume 1) – ISBN 978-1-118-38464-0 (volume 2) – ISBN 978-1-118-38463-3 (set) – ISBN 978-1-118-64724-0 (volume 3) 1. Structural analysis (Engineering) 2. Dynamic testing.
3. Fault location (Engineering) 4. Strains and stresses–Measurement. I. Kutz, Myer.
TA645.H367 2012
620'.0044–dc23

2012011739

Set in 10.5/13.5pt Times by SPi Global, Pondicherry, India

Printed in the United States of America

10 9 8 7 6 5 4 3 2 1

To Mark Berger, teacher

CONTENTS

VOLUME 3

LIST OF CONTRIBUTORS	xxi
PREFACE	xxv
PART VII PHYSICS AND ELECTRICAL ENGINEERING	1943
54 Laser Measurement Techniques	1945
<i>Cecil S. Joseph, Gargi Sharma, Thomas M. Goyette, and Robert H. Giles</i>	
54.1 Introduction, 1945	
54.1.1 History and Development of the MASER, 1945	
54.1.2 Basic Laser Physics, 1946	
54.1.3 Laser Beam Characteristics, 1951	
54.1.4 Example: CO ₂ Laser Pumped Far-Infrared Gas Laser Systems, 1956	
54.1.5 Heterodyned Detection, 1959	
54.1.6 Transformation of Multimode Laser Beams from THz Quantum Cascade Lasers, 1962	
54.1.7 Suggested Reading, 1965	
54.2 Laser Measurements: Laser-Based Inverse Synthetic Aperture Radar Systems, 1965	
54.2.1 ISAR Theory, 1966	
54.2.2 DFT in Radar Imaging, 1967	

54.2.3	Signal Processing Considerations: Sampling Theory, 1970	
54.2.4	Measurement Calibration, 1971	
54.2.5	Example Terahertz Compact Radar Range, 1972	
54.2.6	Suggested Reading, 1974	
54.3	Laser Imaging Techniques, 1974	
54.3.1	Imaging System Measurement Parameters, 1975	
54.3.2	Terahertz Polarized Reflection Imaging of Nonmelanoma Skin Cancers, 1981	
54.3.3	Confocal Imaging, 1985	
54.3.4	Optical Coherence Tomography, 1987	
54.3.5	Femtosecond Laser Imaging, 1990	
54.3.6	Laser Raman Spectroscopy, 1996	
54.3.7	Suggested Reading, 1997	
	References, 1997	
55	Magnetic Force Images Using Capacitive Coupling Effect	2001
	<i>Byung I. Kim</i>	
55.1	Introduction, 2001	
55.2	Experiment, 2004	
55.2.1	Principle, 2004	
55.2.2	Instrumentation, 2004	
55.2.3	Approach, 2005	
55.3	Results and Discussion, 2006	
55.3.1	Separation of Topographic Features from Magnetic Force Images Using Capacitive Coupling Effect, 2007	
55.3.2	Effects of Long-Range Tip–Sample Interaction on Magnetic Force Imaging: A Comparative Study Between Bimorph-Driven System and Electrostatic Force Modulation, 2012	
55.4	Conclusion, 2020	
	References, 2021	
56	Scanning Tunneling Microscopy	2025
	<i>Kwok-Wai Ng</i>	
56.1	Introduction, 2025	
56.2	Theory of Operation, 2026	
56.3	Measurement of the Tunnel Current, 2030	
56.4	The Scanner, 2032	
56.5	Operating Mode, 2035	

- 56.6 Coarse Approach Mechanism, 2036
- 56.7 Summary, 2041
- References, 2042

57 Measurement of Light and Color 2043

John D. Bullough

- 57.1 Introduction, 2043
- 57.2 Lighting Terminology, 2043
 - 57.2.1 Fundamental Light and Color Terms, 2043
 - 57.2.2 Terms Describing the Amount and Distribution of Light, 2047
 - 57.2.3 Terms Describing Lighting Technologies and Performance, 2048
 - 57.2.4 Common Quantities Used in Lighting Specification, 2052
- 57.3 Basic Principles of Photometry and Colorimetry, 2056
 - 57.3.1 Photometry, 2056
 - 57.3.2 Colorimetry, 2063
- 57.4 Instrumentation, 2072
 - 57.4.1 Illuminance Meters, 2072
 - 57.4.2 Luminance Meters, 2072
 - 57.4.3 Spectroradiometers, 2074
- References, 2074

58 The Detection and Measurement of Ionizing Radiation 2075

Clair J. Sullivan

- 58.1 Introduction, 2075
- 58.2 Common Interactions of Ionizing Radiation, 2076
 - 58.2.1 Radiation Interactions, 2076
- 58.3 The Measurement of Charge, 2077
 - 58.3.1 Counting Statistics, 2078
 - 58.3.2 The Two Measurement Modalities, 2080
- 58.4 Major Types of Detectors, 2081
 - 58.4.1 Gas Detectors, 2081
 - 58.4.2 Ionization Chambers, 2086
 - 58.4.3 Proportional Counters, 2090
 - 58.4.4 GM Detectors, 2092
 - 58.4.5 Scintillators, 2092
 - 58.4.6 Readout of Scintillation Light, 2094
 - 58.4.7 Semiconductors, 2096

- 58.5 Neutron Detection, 2100
 - 58.5.1 Thermal Neutron Detection, 2102
 - 58.5.2 Fast Neutron Detection, 2104
- 58.6 Concluding Remarks, 2106
- References, 2106

59 Measuring Time and Comparing Clocks 2109

Judah Levine

- 59.1 Introduction, 2109
- 59.2 A Generic Clock, 2109
- 59.3 Characterizing the Stability of Clocks and Oscillators, 2110
 - 59.3.1 Worst-Case Analysis, 2111
 - 59.3.2 Statistical Analysis and the Allan Variance, 2113
 - 59.3.3 Limitations of the Statistics, 2116
- 59.4 Characteristics of Different Types of Oscillators, 2117
- 59.5 Comparing Clocks and Oscillators, 2119
- 59.6 Noise Models, 2121
 - 59.6.1 White Phase Noise, 2121
 - 59.6.2 White Frequency Noise, 2122
 - 59.6.3 Long-Period Effects: Frequency Aging, 2123
 - 59.6.4 Flicker Noise, 2124
- 59.7 Measuring Tools and Methods, 2126
- 59.8 Measurement Strategies, 2129
- 59.9 The Kalman Estimator, 2133
- 59.10 Transmitting Time and Frequency Information, 2135
 - 59.10.1 Modeling the Delay, 2136
 - 59.10.2 The Common-View Method, 2137
 - 59.10.3 The “Melting-Pot” Version of Common View, 2138
 - 59.10.4 Two-Way Methods, 2139
 - 59.10.5 The Two-Color Method, 2139
- 59.11 Examples of the Measurement Strategies, 2141
 - 59.11.1 The Navigation Satellites of the GPS, 2141
 - 59.11.2 The One-Way Method of Time Transfer: Modeling the Delay, 2144
 - 59.11.3 The Common-View Method, 2145
 - 59.11.4 Two-Way Time Protocols, 2147
- 59.12 The Polling Interval: How Often Should I Calibrate a Clock?, 2152
- 59.13 Error Detection, 2155
- 59.14 Cost–Benefit Analysis, 2156
- 59.15 The National Time Scale, 2157

- 59.16 Traceability, 2158
- 59.17 Summary, 2159
- 59.18 Bibliography, 2160
- References, 2160

60 Laboratory-Based Gravity Measurement 2163

Charles D. Hoyle, Jr.

- 60.1 Introduction, 2163
- 60.2 Motivation for Laboratory-Scale Tests of Gravitational Physics, 2164
- 60.3 Parameterization, 2165
- 60.4 Current Status of Laboratory-Scale Gravitational Measurements, 2166
 - 60.4.1 Tests of the ISL, 2166
 - 60.4.2 WEP Tests, 2167
 - 60.4.3 Measurements of G , 2167
- 60.5 Torsion Pendulum Experiments, 2167
 - 60.5.1 General Principles and Sensitivity, 2168
 - 60.5.2 Fundamental Limitations, 2168
 - 60.5.3 ISL Experiments, 2171
 - 60.5.4 Future ISL Tests, 2172
 - 60.5.5 WEP Tests, 2176
 - 60.5.6 Measurements of G , 2176
- 60.6 Microoscillators and Submicron Tests of Gravity, 2177
 - 60.6.1 Microcantilevers, 2177
 - 60.6.2 Very Short-Range ISL Tests, 2177
- 60.7 Atomic and Nuclear Physics Techniques, 2178
- Acknowledgements, 2178
- References, 2178

61 Cryogenic Measurements 2181

Ray Radebaugh

- 61.1 Introduction, 2181
- 61.2 Temperature, 2182
 - 61.2.1 ITS-90 Temperature Scale and Primary Standards, 2182
 - 61.2.2 Commercial Thermometers, 2183
 - 61.2.3 Thermometer Use and Comparisons, 2193
 - 61.2.4 Dynamic Temperature Measurements, 2199
- 61.3 Strain, 2201
 - 61.3.1 Metal Alloy Strain Gages, 2202
 - 61.3.2 Temperature Effects, 2203

61.3.3	Magnetic Field Effects,	2204
61.3.4	Measurement System,	2205
61.3.5	Dynamic Measurements,	2205
61.4	Pressure,	2205
61.4.1	Capacitance Pressure Sensors,	2206
61.4.2	Variable Reluctance Pressure Sensors,	2206
61.4.3	Piezoresistive Pressure Sensors,	2208
61.4.4	Piezoelectric Pressure Sensors,	2210
61.5	Flow,	2211
61.5.1	Positive Displacement Flowmeter (Volume Flow),	2212
61.5.2	Angular Momentum Flowmeter (Mass Flow),	2212
61.5.3	Turbine Flowmeter (Volume Flow),	2213
61.5.4	Differential Pressure Flowmeter,	2213
61.5.5	Thermal or Calorimetric (Mass Flow),	2216
61.5.6	Hot-Wire Anemometer (Mass Flow),	2217
61.6	Liquid Level,	2218
61.7	Magnetic Field,	2219
61.8	Conclusions,	2220
	References,	2220
62	Temperature-Dependent Fluorescence Measurements	2225
	<i>James E. Parks, Michael R. Cates, Stephen W. Allison, David L. Beshears, M. Al Akerman, and Matthew B. Scudiere</i>	
62.1	Introduction,	2225
62.2	Advantages of Phosphor Thermometry,	2227
62.3	Theory and Background,	2227
62.4	Laboratory Calibration of TP Systems,	2235
62.5	History of Phosphor Thermometry,	2238
62.6	Representative Measurement Applications,	2239
62.6.1	Permanent Magnet Rotor Measurement,	2239
62.6.2	Turbine Engine Component Measurement,	2240
62.7	Two-Dimensional and Time-Dependent Temperature Measurement,	2241
62.8	Conclusion,	2243
	References,	2243
63	Voltage and Current Transducers for Power Systems	2245
	<i>Carlo Muscas and Nicola Locci</i>	
63.1	Introduction,	2245
63.2	Characterization of Voltage and Current Transducers,	2247

63.3	Instrument Transformers, 2248	
63.3.1	Theoretical Fundamentals and Characteristics, 2248	
63.3.2	Instrument Transformers for Protective Purposes, 2252	
63.3.3	Instrument Transformers under Nonsinusoidal Conditions, 2253	
63.3.4	Capacitive Voltage Transformer, 2254	
63.4	Transducers Based on Passive Components, 2255	
63.4.1	Shunts, 2255	
63.4.2	Voltage Dividers, 2256	
63.4.3	Isolation Amplifiers, 2257	
63.5	Hall-Effect and Zero-Flux Transducers, 2258	
63.5.1	The Hall Effect, 2258	
63.5.2	Open-Loop Hall-Effect Transducers, 2259	
63.5.3	Closed-Loop Hall-Effect Transducers, 2259	
63.5.4	Zero-Flux Transducers, 2262	
63.6	Air-Core Current Transducers: Rogowski Coils, 2262	
63.7	Optical Current and Voltage Transducers, 2267	
63.7.1	Optical Current Transducers, 2268	
63.7.2	Optical Voltage Transducer, 2271	
63.7.3	Applications of OCTs and OVTs, 2272	
	References and Further Reading, 2273	

64 Electric Power and Energy Measurement **2275**

Alessandro Ferrero and Marco Faifer

64.1	Introduction, 2275
64.2	Power and Energy in Electric Circuits, 2276
64.2.1	DC Circuits, 2276
64.2.2	AC Circuits, 2277
64.3	Measurement Methods, 2282
64.3.1	DC Conditions, 2282
64.3.2	AC Conditions, 2285
64.4	Wattmeters, 2288
64.4.1	Architecture, 2288
64.4.2	Signal Processing, 2289
64.5	Transducers, 2290
64.5.1	Current Transformers, 2291
64.5.2	Hall-Effect Sensors, 2296
64.5.3	Rogowski Coils, 2297
64.5.4	Voltage Transformers, 2299
64.5.5	Electronic Transformers, 2302
64.6	Power Quality Measurements, 2303
	References, 2305

PART VIII CHEMISTRY 2307**65 An Overview of Chemometrics for the Engineering and Measurement Sciences 2309***Brad Swarbrick and Frank Westad*

- 65.1 Introduction: The Past and Present of Chemometrics, 2309
- 65.2 Representative Data, 2311
 - 65.2.1 A Suggested Workflow for Developing Chemometric Models, 2313
 - 65.2.2 Accuracy and Precision, 2313
 - 65.2.3 Summary of Representative Data Principles, 2316
- 65.3 Exploratory Data Analysis, 2317
 - 65.3.1 Univariate and Multivariate Analysis, 2317
 - 65.3.2 Cluster Analysis, 2318
 - 65.3.3 Principal Component Analysis, 2323
- 65.4 Multivariate Regression, 2352
 - 65.4.1 General Principles of Univariate and Multivariate Regression, 2352
 - 65.4.2 Multiple Linear Regression, 2354
 - 65.4.3 Principal Component Regression, 2355
 - 65.4.4 Partial Least Squares Regression, 2356
- 65.5 Multivariate Classification, 2369
 - 65.5.1 Linear Discriminant Analysis, 2370
 - 65.5.2 Soft Independent Modeling of Class Analogy, 2372
 - 65.5.3 Partial Least Squares Discriminant Analysis, 2381
 - 65.5.4 Support Vector Machine Classification, 2383
- 65.6 Techniques for Validating Chemometric Models, 2385
 - 65.6.1 Test Set Validation, 2386
 - 65.6.2 Cross Validation, 2388
- 65.7 An Introduction to MSPC, 2389
 - 65.7.1 Multivariate Projection, 2389
 - 65.7.2 Hotelling's T^2 Control Chart, 2390
 - 65.7.3 Q -Residuals, 2391
 - 65.7.4 Influence Plot, 2391
 - 65.7.5 Continuous versus Batch Monitoring, 2392
 - 65.7.6 Implementing MSPC in Practice, 2394
- 65.8 Terminology, 2397
- 65.9 Chapter Summary, 2401
- References, 2404

66 Liquid Chromatography **2409**

*Zhao Li, Sandya Beeram, Cong Bi, Ellis Kaufmann, Ryan Matsuda,
Maria Podariu, Elliott Rodriguez, Xiwei Zheng, and David S. Hage*

- 66.1 Introduction, 2409
- 66.2 Support Materials in LC, 2412
- 66.3 Role of the Mobile Phase in LC, 2413
- 66.4 Adsorption Chromatography, 2414
- 66.5 Partition Chromatography, 2415
- 66.6 Ion-Exchange Chromatography, 2417
- 66.7 Size-Exclusion Chromatography, 2419
- 66.8 Affinity Chromatography, 2421
- 66.9 Detectors for Liquid Chromatography, 2423
- 66.10 Other Components of LC Systems, 2426
- Acknowledgements, 2427
- References, 2427

67 Mass Spectroscopy Measurements of Nitrotyrosine-Containing Proteins **2431**

Xianquan Zhan and Dominic M. Desiderio

- 67.1 Introduction, 2431
 - 67.1.1 Formation, Chemical Properties, and Related Nomenclature of Tyrosine Nitration, 2431
 - 67.1.2 Biological Roles of Tyrosine Nitration in a Protein, 2432
 - 67.1.3 Challenge and Strategies to Identify a Nitroprotein with Mass Spectrometry, 2432
 - 67.1.4 Biological Significance Measurement of Nitroproteins, 2434
- 67.2 Mass Spectrometric Characteristics of Nitropeptides, 2434
 - 67.2.1 MALDI-MS Spectral Characteristics of a Nitropeptide, 2434
 - 67.2.2 ESI-MS Spectral Characteristics of a Nitropeptide, 2437
 - 67.2.3 Optimum Collision Energy for Ion Fragmentation and Detection Sensitivity for a Nitropeptide, 2438
 - 67.2.4 MS/MS Spectral Characteristics of a Nitropeptide under Different Ion-Fragmentation Models, 2440
- 67.3 MS Measurement of *In Vitro* Synthetic Nitroproteins, 2443
 - 67.3.1 Importance of Measurement of *In Vitro* Synthetic Nitroproteins, 2443
 - 67.3.2 Commonly Used *In Vitro* Nitroproteins and Their Preparation, 2443
 - 67.3.3 Methods Used to Measure *In Vitro* Synthetic Nitroproteins, 2444

67.4	MS Measurement of <i>In Vivo</i> Nitroproteins,	2446
67.4.1	Importance of Isolation and Enrichment of <i>In Vivo</i> Nitroprotein/Nitropeptide Prior to MS Analysis,	2446
67.4.2	Methods Used to Isolate and Enrich <i>In Vivo</i> Nitroproteins/Nitropeptides,	2446
67.5	MS Measurement of <i>In Vivo</i> Nitroproteins in Different Pathological Conditions,	2449
67.6	Biological Function Measurement of Nitroproteins,	2456
67.6.1	Literature Data-Based Rationalization of Biological Functions,	2457
67.6.2	Protein Domain and Motif Analyses,	2459
67.6.3	Systems Pathway Analysis,	2459
67.6.4	Structural Biology Analysis,	2460
67.7	Pitfalls of Nitroprotein Measurement,	2462
67.8	Conclusions,	2463
	Nomenclature,	2464
	Acknowledgments,	2465
	References,	2465

68 Fluorescence Spectroscopy

2475

Yevgen Povrozin and Beniamino Barbieri

68.1	Observables Measured in Fluorescence,	2476
68.2	The Perrin–Jablonski Diagram,	2476
68.3	Instrumentation,	2479
68.3.1	Light Source,	2480
68.3.2	Monochromator,	2480
68.3.3	Light Detectors,	2481
68.3.4	Instrumentation for Steady-State Fluorescence: Analog and Photon Counting,	2483
68.3.5	The Measurement of Decay Times: Frequency-Domain and Time-Domain Techniques,	2484
68.4	Fluorophores,	2486
68.5	Measurements,	2487
68.5.1	Excitation Spectrum,	2487
68.5.2	Emission Spectrum,	2488
68.5.3	Decay Times of Fluorescence,	2490
68.5.4	Quantum Yield,	2492
68.5.5	Anisotropy and Polarization,	2492
68.6	Conclusions,	2498
	References,	2498
	Further Reading,	2498

69 X-Ray Absorption Spectroscopy 2499*Grant Bunker*

- 69.1 Introduction, 2499
- 69.2 Basic Physics of X-Rays, 2499
 - 69.2.1 Units, 2500
 - 69.2.2 X-Ray Photons and Their Properties, 2500
 - 69.2.3 X-Ray Scattering and Diffraction, 2501
 - 69.2.4 X-Ray Absorption, 2502
 - 69.2.5 Cross Sections and Absorption Edges, 2503
- 69.3 Experimental Requirements, 2505
- 69.4 Measurement Modes, 2507
- 69.5 Sources, 2507
 - 69.5.1 Laboratory Sources, 2507
 - 69.5.2 Synchrotron Radiation Sources, 2508
 - 69.5.3 Bend Magnet Radiation, 2509
 - 69.5.4 Insertion Devices: Wigglers and Undulators, 2509
- 69.6 Beamlines, 2512
 - 69.6.1 Instrument Control and Scanning Modes, 2512
 - 69.6.2 Double-Crystal Monochromators, 2513
 - 69.6.3 Focusing Conditions, 2514
 - 69.6.4 X-Ray Lenses and Mirrors, 2515
 - 69.6.5 Harmonics, 2516
- 69.7 Detectors, 2518
 - 69.7.1 Ionization Chambers and PIN Diodes, 2519
 - 69.7.2 Solid-State Detectors, SDDs, and APDs, 2520
- 69.8 Sample Preparation and Detection Modes, 2521
 - 69.8.1 Transmission Mode, 2521
 - 69.8.2 Fluorescence Mode, 2521
 - 69.8.3 HALO, 2522
 - 69.8.4 Sample Geometry and Background Rejection, 2523
 - 69.8.5 Oriented Samples, 2525
- 69.9 Absolute Measurements, 2526
- References, 2526

70 Nuclear Magnetic Resonance (NMR) Spectroscopy 2529*Kenneth R. Metz*

- 70.1 Introduction, 2529
- 70.2 Historical Review, 2530
- 70.3 Basic Principles of Spin Magnetization, 2531
- 70.4 Exciting the NMR Signal, 2534

70.5	Detecting the NMR Signal, 2538
70.6	Computing the NMR Spectrum, 2540
70.7	NMR Instrumentation, 2542
70.8	The Basic Pulsed FTNMR Experiment, 2550
70.9	Characteristics of NMR Spectra, 2551
70.9.1	The Chemical Shift, 2552
70.9.2	Spin–Spin Coupling, 2557
70.10	NMR Relaxation Effects, 2563
70.10.1	Spin–Lattice Relaxation, 2563
70.10.2	Spin–Spin Relaxation, 2565
70.10.3	Quantitative Analysis by NMR, 2568
70.11	Dynamic Phenomena in NMR, 2568
70.12	Multidimensional NMR, 2573
70.13	Conclusion, 2580
	References, 2580

71	Near-Infrared Spectroscopy and Its Role in Scientific and Engineering Applications	2583
-----------	---	-------------

Brad Swarbrick

71.1	Introduction to Near-Infrared Spectroscopy and Historical Perspectives, 2583
71.1.1	A Brief Overview of Near-Infrared Spectroscopy and Its Usage, 2583
71.1.2	A Short History of NIR, 2585
71.2	The Theory Behind NIR Spectroscopy, 2588
71.2.1	IR Radiation, 2588
71.2.2	The Mechanism of Interaction of NIR Radiation with Matter, 2588
71.2.3	Absorbance Spectra, 2591
71.3	Instrumentation for NIR Spectroscopy, 2595
71.3.1	General Configuration of Instrumentation, 2595
71.3.2	Filter-Based Instruments, 2597
71.3.3	Holographic Grating-Based Instruments, 2598
71.3.4	Stationary Spectrographic Instruments, 2600
71.3.5	Fourier Transform Instruments, 2601
71.3.6	Acoustooptical Tunable Filter Instruments, 2603
71.3.7	Microelectromechanical Spectrometers, 2604
71.3.8	Linear Variable Filter Instruments, 2605
71.3.9	A Brief Overview of Detectors Used for NIR Spectroscopy, 2606
71.3.10	Summary, 2608

71.4	Modes of Spectral Collection and Sample Preparation in Nir Spectroscopy, 2609
71.4.1	Transmission Mode, 2609
71.4.2	Diffuse Reflectance, 2611
71.4.3	Sample Preparation, 2613
71.4.4	Fiber Optic Probes, 2617
71.4.5	Summary of Sampling Methods, 2619
71.5	Preprocessing of Nir Spectra for Chemometric Analysis, 2620
71.5.1	Preprocessing of NIR Spectra, 2621
71.5.2	Minimizing Additive Effects, 2621
71.5.3	Minimizing Multiplicative Effects, 2627
71.5.4	Preprocessing Summary, 2633
71.6	A Brief Overview of Applications of NIR Spectroscopy, 2633
71.6.1	Agricultural Applications, 2634
71.6.2	Pharmaceutical/Biopharmaceutical Applications, 2636
71.6.3	Applications in the Petrochemical and Refining Sectors, 2644
71.6.4	Applications in the Food and Beverage Industries, 2646
71.7	Summary and Future Perspectives, 2647
71.8	Terminology, 2648
	References, 2652

72 Nanomaterials Properties 2657

Paul J. Simmonds

72.1	Introduction, 2657
72.2	The Rise of Nanomaterials, 2660
72.3	Nanomaterial Properties Resulting from High Surface-Area-to-Volume Ratio, 2661
72.3.1	The Importance of Surfaces in Nanomaterials, 2661
72.3.2	Electrostatic and Van der Waals Forces, 2662
72.3.3	Color, 2663
72.3.4	Melting Point, 2663
72.3.5	Magnetism, 2664
72.3.6	Hydrophobicity and Surface Energetics, 2664
72.3.7	Nanofluidics, 2666
72.3.8	Nanoporosity, 2668
72.3.9	Nanomembranes, 2669
72.3.10	Nanocatalysis, 2670
72.3.11	Further Increasing the SAV Ratio, 2671
72.3.12	Nanopillars, 2672
72.3.13	Nanomaterial Functionalization, 2673
72.3.14	Other Applications for High SAV Ratio Nanomaterials, 2674

72.4	Nanomaterial Properties Resulting from Quantum Confinement,	2674
72.4.1	Quantum Well Nanostructures,	2677
72.4.2	Quantum Wire Nanostructures,	2682
72.4.3	Quantum Dot Nanostructures,	2691
72.5	Conclusions,	2695
	References,	2695

73 Chemical Sensing **2707**

W. Rudolf Seitz

73.1	Introduction,	2707
73.2	Electrical Methods,	2709
73.2.1	Potentiometry,	2709
73.2.2	Voltammetry,	2713
73.2.3	Chemiresistors,	2715
73.2.4	Field Effect Transistors,	2716
73.3	Optical Methods,	2717
73.3.1	In situ Optical Measurements,	2717
73.3.2	Raman Spectroscopy,	2719
73.3.3	Indicator-Based Optical Sensors,	2721
73.4	Mass Sensors,	2722
73.5	Sensor Arrays (Electronic Nose),	2724
	References,	2724

INDEX **2727**

LIST OF CONTRIBUTORS

M. Al Akerman, Emco-Williams Inc, Knoxville, TN, USA

Stephen W. Allison, Emco-Williams Inc, Knoxville, TN, USA

Beniamino Barbieri, ISS, Champaign, IL, USA

David L. Beshears, Emco-Williams Inc, Knoxville, TN, USA

Sandya Beeram, Department of Chemistry, University of Nebraska, Lincoln, NE, USA

Cong Bi, Department of Chemistry, University of Nebraska, Lincoln, NE, USA

John D. Bullough, Lighting Research Center, Rensselaer Polytechnic Institute, Troy, NY, USA

Grant Bunker, Department of Physics, Illinois Institute of Technology, Chicago, IL, USA

Michael R. Cates, Emco-Williams Inc, Knoxville, TN, USA

Dominic M. Desiderio, The Charles B. Stout Neuroscience Mass Spectrometry Laboratory, Department of Neurology, College of Medicine, University of Tennessee Health Science Center, Memphis, TN, USA

Marco Faifer, Dipartimento di Elettronica, Informazione e Bioingegneria, Politecnico di Milano, Milano, Italy

Alessandro Ferrero, Dipartimento di Elettronica, Informazione e Bioingegneria, Politecnico di Milano, Milano, Italy

Robert H. Giles, Biomedical Terahertz Technology Center and Submillimeter-Wave Technology Laboratory, Department of Physics and Applied Physics, University of Massachusetts Lowell, Lowell, MA, USA

Thomas M. Goyette, Submillimeter-Wave Technology Laboratory, Department of Physics and Applied Physics, University of Massachusetts Lowell, Lowell, MA, USA

David S. Hage, Department of Chemistry, University of Nebraska, Lincoln, NE, USA

Charles D. Hoyle, Jr., Department of Physics and Astronomy, Humboldt State University, Arcata, CA, USA

Cecil S. Joseph, Biomedical Terahertz Technology Center, Department of Physics and Applied Physics, University of Massachusetts Lowell, Lowell, MA, USA

Ellis Kaufmann, Department of Chemistry, University of Nebraska, Lincoln, NE, USA

Byung I. Kim, Department of Physics, Boise State University, Boise, ID, USA

Judah Levine, Time and Frequency Division and JILA, NIST and the University of Colorado, Boulder, CO, USA

Zhao Li, Department of Chemistry, University of Nebraska, Lincoln, NE, USA

Nicola Locci, Department of Electrical and Electronic Engineering, University of Cagliari, Cagliari, Italy

Ryan Matsuda, Department of Chemistry, University of Nebraska, Lincoln, NE, USA

Kenneth R. Metz, Chemistry Department, Merkert Chemistry Center, Boston College, Chestnut Hill, MA, USA

Carlo Muscas, Department of Electrical and Electronic Engineering, University of Cagliari, Cagliari, Italy

Kwok-Wai Ng, Department of Physics and Astronomy, University of Kentucky, Lexington, KY, USA

James E. Parks, Department of Physics, University of Tennessee, Knoxville, TN, USA

Maria Podariu, Department of Chemistry, University of Nebraska, Lincoln, NE, USA

Yevgen Povrozin, ISS, Champaign, IL, USA

Ray Radebaugh, Applied Chemicals and Materials Division, National Institute of Standards and Technology, Boulder, CO, USA

Elliott Rodriguez, Department of Chemistry, University of Nebraska, Lincoln, NE, USA

Matthew B. Scudiere, Emco-Williams Inc, Knoxville, TN, USA

W. Rudolf Seitz, Department of Chemistry, University of New Hampshire, Durham, NH, USA

Gargi Sharma, Biomedical Terahertz Technology Center, Department of Physics and Applied Physics, University of Massachusetts Lowell, Lowell, MA, USA

Paul J. Simmonds, Departments of Physics & Materials Science and Engineering, Boise State University, Boise, ID, USA

Clair J. Sullivan, Department of Nuclear, Plasma, and Radiological Engineering, University of Illinois at Urbana-Champaign, Urbana, IL, USA

Brad Swarbrick, Quality by Design Consultancy, Sydney, New South Wales, Australia

Frank Westad, CAMO Software AS, Oslo, Norway

Xianquan Zhan, Key Laboratory of Cancer Proteomics of Chinese Ministry of Health, Xiangya Hospital, Central South University, Changsha, P. R. China

Xiwei Zheng, Department of Chemistry, University of Nebraska, Lincoln, NE, USA

PREFACE

The idea for the *Handbook of Measurement in Science and Engineering* came from a Wiley book first published over 30 years ago. It was *Fundamentals of Temperature, Pressure and Flow Measurements*, written by a sole author, Robert P. Benedict, who also wrote Wiley books on gas dynamics and pipe flow. Bob was a pleasant, unassuming, and smart man. I was the Wiley editor for professional-level books in mechanical engineering when Bob was writing such books, so I knew him as a colleague. I recall meeting him in the Wiley offices at a time when he seemed to be having some medical problems, which he was reluctant to talk about. Recently, I discovered a book published in 1972 by a London firm, Pickering & Inglis, which specializes in religion. This book was *Journey Away from God*, an intriguing title. The author's name was Robert P. Benedict. I do not know whether the two Benedicts are in fact the same person, although Amazon seems to think so. (See the Robert P. Benedict page.) In any case, I do not recall Bob's mentioning the book when we had an occasion to talk.

The moral of this story, if there is one, is that the men and women who contributed the chapters in this handbook are real people, who have real-world concerns, in addition to the expertise required to write about technology. They have families, jobs, careers, and all manner of cares about the minutia of daily life to deal with. And that they have been able to find the time and energy to write these chapters is remarkable. I salute them. I have spent a lot of time in my life writing and editing books. I wrote my first Wiley book somewhat earlier than Bob Benedict wrote his. When Wiley published *Temperature Control* in 1967, I was in my mid-twenties and was a practicing engineer, working on temperature control of the Apollo inertial guidance system at the MIT Instrumentation Lab, where I had done my bachelor's thesis. One of the coauthors of my book was to have been a Tufts Mechanical Engineering Professor by the name of John Sununu (yes, that John Sununu), but he and the other coauthor dropped out of the project before the contract was signed. So I wrote the short book myself.

Bob Benedict's measurement book, the third edition of which is still in print, surfaced several years ago, during a discussion I was having with one of my Wiley editors at the time, Bob Argentieri, about possible projects we could collaborate on. It turned out that no one had attempted to update Benedict's book. I have not been a practicing engineer for some time, so I was not in a position to do an update as a single author—or even with a collaborator or two. Most of my career life has been in scientific and technical publishing, however, and for over a decade I have conceived of, and edited, numerous handbooks for several publishers. (I also write fiction, but that is another story.) So, it was natural for me to think about using Benedict's book as the kernel of a much larger and broader reference work dealing with engineering measurements. The idea, formed during that discussion, that I might edit a contributed handbook on engineering measurements took hold, and with the affable and expert guidance of my other Wiley editor at the time, George Telecki, the volume you are holding in your hands, or reading on an electronic device, came into being.

Like many such large reference works, this handbook went through several iterations before the final table of contents was set, although the general plan for arrangement of chapters has been the same throughout the project. The initial print version of the handbook was divided into two volumes. The chapters were arranged essentially by engineering discipline. The first volume contains 30 chapters related to five engineering disciplines, which are divided into three parts:

Part I, "Civil and Environmental Engineering," which contains seven chapters, all but one of them dealing with measurement and testing techniques for structural health monitoring, GIS and computer mapping, highway bridges, environmental engineering, hydrology, and mobile source emissions (the exception being the chapter on traffic congestion management, which describes the deployment of certain measurements)

Part II, "Mechanical and Biomedical Engineering," which contains 16 chapters, all of them dealing with techniques for measuring dimensions, surfaces, mass properties, force, resistive strain, vibration, acoustics, temperature, pressure, velocity, flow, heat flux, heat transfer for nonboiling two-phase flow, solar energy, wind energy, human movement, and physiological flow

Part III, "Industrial Engineering," which contains seven chapters dealing with statistical quality control, evaluating and selecting technology-based projects, manufacturing systems evaluation, measuring performance of chemical process equipment, industrial energy efficiency, industrial waste auditing, and organizational performance measurement

The second volume contains 23 chapters divided into three parts:

Part IV, "Materials Properties and Testing," which contains 15 chapters dealing with measurement of viscosity, tribology, corrosion, surface properties, and thermal

conductivity of engineering materials; properties of metals, alloys, polymers, and particulate composite materials; nondestructive inspection; and testing of metallic materials, ceramics, plastics, and plastics processing

Part V, “Instrumentation,” which contains five chapters covering electronic equipment used for measurements

Part VI, “Measurement Standards,” which contains three chapters covering units and standards, measurement uncertainty, and error analysis

This new, third volume of the handbook expands the range of the handbook to cover measurements in physics, electrical engineering, and chemistry. This volume contains 20 chapters divided into two major parts:

Part VII, “Physics and Electrical Engineering,” which contains 11 chapters, covering laser-based measurement systems, scanning probe microscopy, scanning tunneling microscopy, photometry, detection and measurement of ionizing radiation, time measurement systems, laboratory-based gravity measurement, cryogenic measurements, temperature-dependent fluorescence measurements, measurement of electrical quantities, and electrical power measurement

Part VIII, “Chemistry” which contains nine chapters, covering chemometrics/chemical metrology, liquid chromatography, mass spectrometry measurements, basic principles of fluorescence spectroscopy, X-ray absorption spectroscopy, NMR spectroscopy, near-infrared spectroscopy, nanomaterials properties, and chemical sensing

Thanks to Brett Kurzman, my new editor for this volume, and Kari Capone, Allison, McGinniss, and Alex Castro for shepherding the manuscript toward production and to the stalwarts Kristen Parrish and Shirley Thomas for bringing this handbook volume home. Thanks, also, to my wife Arlene, who helps me with everything else.

MYER KUTZ

*Delmar, NY
October 2015*

PART VII

PHYSICS AND ELECTRICAL ENGINEERING

LASER MEASUREMENT TECHNIQUES

CECIL S. JOSEPH¹, GARGI SHARMA¹, THOMAS M. GOYETTE²,
AND ROBERT H. GILES^{1,2}

¹*Biomedical Terahertz Technology Center, Department of Physics and Applied Physics,
University of Massachusetts Lowell, Lowell, MA, USA*

²*Submillimeter-Wave Technology Laboratory, Department of Physics and Applied Physics,
University of Massachusetts Lowell, Lowell, MA, USA*

54.1 INTRODUCTION

54.1.1 History and Development of the MASER

Lasers have enabled us to investigate the structure of atoms and molecules as well as accurately measure fundamental constants and observe natural phenomenology over a broad range of frequencies. But the applications of laser measurement technologies have advanced rapidly, for the instrumentation is little more than half a century in the making, and many laser-based measurement systems are now commercially available.

Albert Einstein first suggested the concept of stimulated emitted radiation in 1916. Indicating that a photon could cause the energy transition of an atom from an upper level to a lower level, Einstein proposed that the atom would emit a photon with the same energy as the photon initiating the energy transition. It was not until 1928 that Ladenburg observed stimulated emission.

As a faculty member at Columbia University, Charles H. Townes applied the concept to stimulating molecules in a resonant cavity to generate microwave radiation. His insight enabled postdoctoral fellow Herbert Zeiger and graduate student James P. Gordon to build a working maser by 1953. In 1954, Prokhorov and Basov of Moscow's Lebedev Physical Institute published the complete details for establishing

stimulated microwave emission. Prokhorov, Basov, and Townes shared the 1964 Nobel Prize for their research.

54.1.1.1 From Maser to Laser: Extending the Operable Region The science of stimulated emissions at wavelengths shorter than the microwave regime is very different. Though many physicists were now in pursuit of the technology, it was Charles Townes and Bell Lab's researcher Arthur L. Schawlow who first published the requirements for generating visible radiation in 1959. Detailing the parameters such as cavity structures, spontaneous emission ratios, and transition energy levels for generating visible radiation, Townes and Schawlow filed a patent for their development of "optical masers." However, Gordon Gould's graduate work at Columbia University predated Townes and Schawlow's patent and publication so after extended litigation Gould was awarded patents on the optical pumping techniques and his cavity designs using Brewster windows and a number of applications of what he referred to as "lasers."

While most researchers were building gas lasers at this point, Theodore H. Maiman was investigating the energy levels of ruby crystals. By the mid-1960s he demonstrated the first solid-state laser using a rod of synthetic ruby, thereby expanding the possibilities of light sources and the type of devices available. Through continued innovations by the research community, the performance characteristics of lasers were optimized, and these devices became the primary source for a rapidly growing market of measurement technologies.

54.1.2 Basic Laser Physics

Though there exist a large variety of laser devices that cover the electromagnetic spectrum from the microwave to the ultraviolet (higher frequency, i.e., X-ray lasers are also under development), the core elements of any laser device are:

1. A laser medium
2. A pump process
3. Optical feedback elements

Different approaches to these basic elements and varied combinations thereof lead to the output frequency range of the device. Put simply, the pump process injects energy into the laser media and via optical feedback techniques, and part of this energy is recovered from the media in the form of coherent, stimulated photons, which make up a laser beam. Here we discuss the basic requirements of these elements and how they combine to form a laser.

54.1.2.1 Stimulated Emission and Atomic Rate Equations A laser medium is made up of a collection of atoms/molecules in a gas, liquid, or solid phase. When the medium absorbs energy (e.g., by heating), some of the atoms transition to higher

energy levels in the quantum mechanical energy structure. A fraction of these atoms spontaneously lose energy by emitting a photon and transition back to a lower energy state. In general, the spontaneous decay rate of any state is proportional to the number of atoms in that state. So if $N_i(t)$ is the instantaneous population of an energy level, E_i , the spontaneous decay rate is given by $dN_i/dt_{\text{spontaneous}} = -\gamma_i N_i$, where γ_i is the spontaneous decay rate of the energy level E_i with $\gamma_i = 1/\tau_i$ and τ_i being the lifetime of the state. (Note that both radiative and nonradiative transitions are allowed and implicitly included. Emitted photons are radiative transitions, while dissipation of energy via lattice phonons is nonradiative transitions.)

It is also possible to stimulate absorption from a lower energy level to a higher energy level by providing a photon that corresponds to the energy difference between the levels (i.e., $\Delta E = h\nu$, where h is the Planck constant and ν is the photon frequency). This stimulates both absorption and emission at the applied signal frequency. The primary difference between stimulated transitions and spontaneous transitions is that stimulated transitions are caused by an applied signal and as such the photons emitted are coherent with that signal. Spontaneous transitions, on the other hand, are radiated out by atoms driven independently of each other and as such are incoherent.

Consider a two-energy-level system, where level E_i has a population N_i and E_j has a population N_j . Moreover, let $E_j > E_i$, corresponding to an energy difference, ΔE_{ji} . The spontaneous decay rate of level E_j to E_i is given by γ_{ji} . If we now apply a signal to this system that corresponds to the energy difference, that is, $\Delta E_{ji} = h\nu_{ji}$, then we stimulate absorption from E_i to E_j , and we stimulate emission from E_j to E_i . If $n(t)$ is the photon density of the incident signal, then the change in population of the higher energy state is given by

$$\begin{aligned} \frac{dN_j(t)}{dt} = & \text{stimulated absorption from } E_i - \text{stimulated emission from } E_j \\ & - \text{spontaneous emission from } E_j, \end{aligned}$$

that is,

$$\frac{dN_j(t)}{dt} = Kn(t)N_i(t) - Kn(t)N_j(t) - \gamma_{ji}N_j(t)$$

where K is a constant that measures the strength of the stimulated response.

Thus the rate at which atoms make *stimulated* transitions is proportional to the population difference of the two energy levels and the applied signal intensity. Each absorbed photon attenuates the applied signal, and each emitted photon amplifies it.

Thus the change in photon density of the applied signal due to stimulated transitions can be represented as $dn/dt = -K\Delta N_{ij}n(t)$, where ΔN_{ij} is the population difference between the lower and upper energy state. When there are more atoms in the lower energy state, $\Delta N_{ij} > 0$, and the applied signal is attenuated. If, however, $\Delta N_{ij} < 0$, the applied

signal is amplified; this condition that requires a larger number of atoms in the higher energy state for signal amplification is one of the basic requirements of a working laser and is referred to as *population inversion*.

Before discussing the means to achieving population inversion in laser media, it is important to consider the difference between spontaneously emitted photons and stimulated photons. Stimulated transitions are caused by an applied signal and as such are coherent with the applied signal, unlike spontaneous emission, which is incoherent.

54.1.2.2 Population Inversion and Laser Amplification Population inversion is an essential condition for laser amplification, that is, there must be a larger number of atoms in the upper excited state than the lower in order for the applied signal to be amplified. According to the Boltzmann principle, the relative populations of any two energy levels are given by $N_2/N_1 = e^{-\Delta E/kT}$ where $\Delta E = E_2 - E_1$ and $E_2 > E_1$. Therefore in order to create population inversion, we need to “pump” atoms into the upper excited state since the population difference is always attenuating at equilibrium for a two-level system. Typically other upper energy levels are used to “feed” the upper lasing level, thereby creating inversion.

As an example, consider the lasing levels and population inversion in a ruby laser. Ruby is essentially sapphire (Al_2O_3) doped with chromium. The Cr^{3+} ions replace a fraction of the Al^{3+} ions in the sapphire lattice. These Cr^{3+} ions in the lattice have energy levels in the red at approximately 694 nm. A representative energy-level diagram is shown in Figure 54.1.

In a typical ruby laser, atoms are optically pumped using a xenon arc lamp from the ground state to higher energy levels. Atoms in these levels relax rapidly down to highly metastable “R” levels via nonradiative transitions. With sufficient pumping, it is

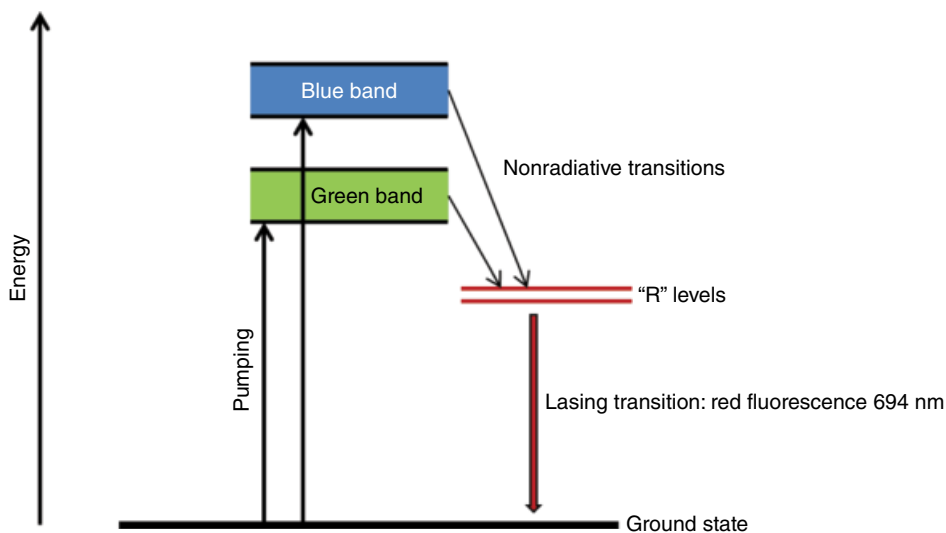


FIGURE 54.1 Energy-level transitions in ruby.

possible to create population inversion between the metastable (long lifetime) R levels and the ground state. This transition can then be amplified.

The reason inversion occurs without violating the Boltzmann principle is that through sufficient pumping more than half of the atoms transfer from the ground state to the higher energy states. These states have very short lifetimes and relax with close to 100% quantum efficiency down to the R levels. Since the R levels have a relatively long lifetime (≈ 4.3 ms), they are being fed with atoms at a significantly faster rate than atoms decaying (spontaneously) from them, leading to population inversion [1].

54.1.2.3 Laser Cavity or Oscillator In order to amplify stimulated emission as photons pass through laser media, one must generate population inversion using a pump process. With the condition of inversion met, any signal at the lasing (inversion) frequency that passes through the gain media is slightly amplified. If the media are then enclosed in an oscillator cavity that forces this signal to repeatedly pass back and forth through the gain media, then the signal strength is amplified in each pass as long as inversion exists.

For example, suppose a laser medium pumped to maintain population inversion is enclosed in an optical cavity of length L as shown in Figure 54.2. Initially there is always some spontaneous emission at the laser frequency. However, if one of these spontaneously emitted photons is aligned with the optical axis of the cavity, then while it travels through the medium it causes stimulated emission along the same axis. When this radiation reflects from either end of the cavity, it retraces its path through the gain media and is further amplified by stimulated emission. This process continues and the stimulated transition at the laser frequency in a direction aligned with the optical axis of the cavity is continuously amplified until the stimulated emissions negate population inversion.

In general the amplification process will continue until the pump process can maintain population inversion. That is the point at which the net gain in any round trip through the medium is balanced by the net loss; this is the steady-state condition. In order to extract a portion of the stimulated signal, the mirrors at either end of the cavity are generally partially transmitting. For the system shown in Figure 54.2, r_1 and r_2 are

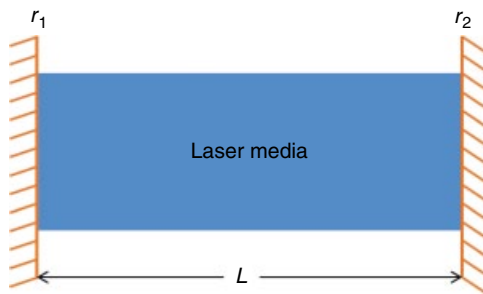


FIGURE 54.2 A laser medium enclosed in an optical cavity of length L .

the Fresnel reflection coefficients of the mirror surfaces. Then, the previously discussed steady state requires that the amplitude gain in one round trip through the length L cavity be equal to 1, that is,

$$r_1 r_2 e^{2\alpha L} = 1,$$

where α is the gain of the laser medium.

Note that r_1 , r_2 , and L are characteristics of the optical cavity; thus the steady-state amplitude condition yields the laser gain coefficient, α , that must be maintained in that specific cavity:

$$\alpha = \frac{1}{2L} \ln \left(\frac{1}{r_1 r_2} \right) = \frac{1}{4L} \ln \left(\frac{1}{R_1 R_2} \right)$$

where R_1 and R_2 are the power reflectivity of the two mirrors.

The laser gain coefficient depends on the population inversion in the laser media. The greater the inversion, the larger the gain. Assuming that the lasing transition has a Lorentzian line shape, then the *threshold inversion density* (ΔN_t) required for lasing action in a cavity is given by

$$\Delta N_t = \frac{\pi \Delta \omega}{\lambda^2 \gamma L} \ln \left(\frac{1}{R_1 R_2} \right)$$

where $\Delta \omega$ is the transition linewidth, λ is the transition wavelength, and γ is the transition decay rate [1].

Steady state also introduces a round-trip phase condition for the oscillator. The steady-state frequencies must correspond to standing waves in the oscillator. For the simple cavity shown in Figure 54.2, this implies that the length L must be an integral number of half wavelengths for sustained oscillations, that is,

$$L = \frac{m\lambda}{2}; m = \text{integer}$$

which yields discrete axial mode frequencies given by

$$\omega_m = m2\pi \left(\frac{c}{2L} \right)$$

For most lasers, several axial mode frequencies exist within the transition linewidth and m is a large integer.

Finally, we also need to consider the wave nature of light and its effects on the transverse spatial properties of the laser beam. The laser beam oscillating in the cavity has a spatial extent along the transverse axis and, as it oscillates, will spread due to diffraction. After the beam size exceeds the cavity reflectors, the lost signal can be accounted for as lowered reflectivity or increased round-trip propagation loss. However,

the pattern still needs to be self-consistent over one round trip, that is, the beam parameters must also match for a sustainable laser mode. Therefore, as the reflections on the ends introduce changes to the transverse mode profile of the beam, sustained oscillations depend upon the geometry of the cavity. The laser resonator system discussed so far is essentially two flat mirrors at the ends of the cavity; however, several other cavity designs are possible, such as curved mirrors, which focus the radiation in the transverse direction (see Suggested Readings at the end of this section).

In general, for any given laser cavity, one can find a set of discrete transverse eigenmodes that can propagate self-consistently in that cavity. Thus, for any working laser, one requires a pumping process that is capable of sustaining population inversion above a cavity defined threshold in the lasing media. The laser cavity or resonator will also define the axial frequencies that are capable of oscillations, and the cavity's geometry will then dictate which transverse modes are capable of maintaining sustained oscillations in the system.

54.1.3 Laser Beam Characteristics

When extracting the output beam from a laser cavity for an application, it is the frequency and amplitude characteristics of the output radiation that are of practical relevance. Moreover, it is also important to understand the propagation of laser beams through space and optical elements. We start by describing the frequency and amplitude characteristics of the output radiation and then discuss the solution of the wave equation and explore the fundamental mode solution. Finally we cover a basic introduction to the ABCD law for beam propagation.

54.1.3.1 Laser Frequency and Amplitude Characteristics Ideally, the output of a laser is essentially single frequency, amplitude stabilized, and highly directional. In practice however, the cavity and laser setup introduce slight variations from the ideal output. As previously described, the laser cavity defines the allowed axial modes of oscillation that ultimately set the laser frequency. In general, real lasers can operate at several axial modes, still within the atomic linewidth of the lasing transition. Several steps can be taken to improve frequency stability in real laser systems, and the ultimate limit on frequency stability is set by the spontaneous emission in the gain media. Amplitude stability is generally achieved by the gain balancing mechanism (round-trip gain=1). In practical systems, slight amplitude instability is introduced by pumping mechanisms and thermal cavity effects. Also, laser beams are temporally and spatially coherent.

54.1.3.2 Fundamental Mode Several laser modes can be supported by the lasing cavity, provided the field component, u , of the laser beam satisfies the scalar wave equation [2]

$$\nabla^2 u + k^2 u = 0, \quad \text{where } k = 2\pi/\lambda \text{ is the propagation constant.}$$

Assuming that the function varies slowly along the propagation axis, the second derivative along this axis $\left(\frac{\partial^2}{\partial z^2}\psi\right)$ can be neglected, and we get

$$\frac{\partial^2}{\partial x^2}\psi + \frac{\partial^2}{\partial y^2}\psi - 2ik\frac{\partial}{\partial z}\psi = 0$$

One solution to the above equation is a Gaussian beam profile:

$$\psi = \exp\left\{-i\left(P + \frac{k}{2q}r^2\right)\right\}$$

where $r^2 = x^2 + y^2$ and P and q are functions of z .

The parameter q is called the complex beam parameter and it can be related to real beam parameters R and w using the following relation:

$$\frac{1}{q} = \frac{1}{R} - i\frac{\lambda}{\pi w^2}$$

Both R and w are functions of position on the propagation axis z . $R(z)$ represents the radius of curvature of the wavefront at z , and $w(z)$ represents the “beam radius.” Figure 54.3 shows the intensity distribution as a function of radial distance r from the propagation axis. Note that intensity is proportional to the square of the field amplitude.

The intensity of the laser is typically concentrated near the propagation axis. As shown in Figure 54.3, $w(z)$ measures the decrease in field strength with distance from

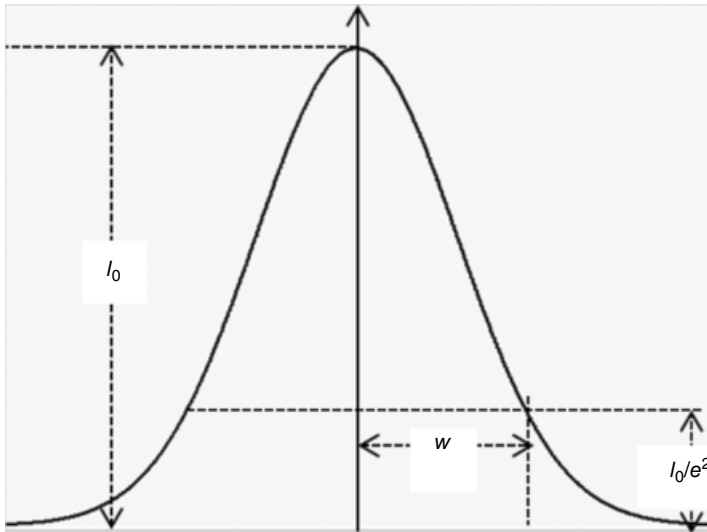


FIGURE 54.3 Intensity distribution of a Gaussian beam.

the propagation axis. In general, the “beam radius” corresponds to $w(z)$, which is defined as the radius at which the beam intensity drops to $1/e^2$ of its peak value. The beam diameter is $2w$.

The Gaussian beam contracts to a minimum value along the z -axis; the waist at this point is referred to as the “beam waist,” w_0 . Noting that at the minimum waist the wavefront is a plane wave, which indicates that $R = \infty$ at this point, thus the complex beam parameter q at this point (denoted q_0) becomes

$$q_0 = i \frac{\pi w_0^2}{\lambda}$$

The size and curvature of a wavefront as it propagates along the z -axis are described by [2]

$$w^2(z) = w_0^2 \left[1 + \left(\frac{\lambda z}{\pi w_0^2} \right)^2 \right]$$

and

$$R(z) = z \left[1 + \left(\frac{\pi w_0^2}{\lambda z} \right)^2 \right]$$

From Figure 54.4 and the equations above we can see that the far-field diffraction angle, θ , is given by

$$\theta = \frac{\lambda}{\pi w_0}$$

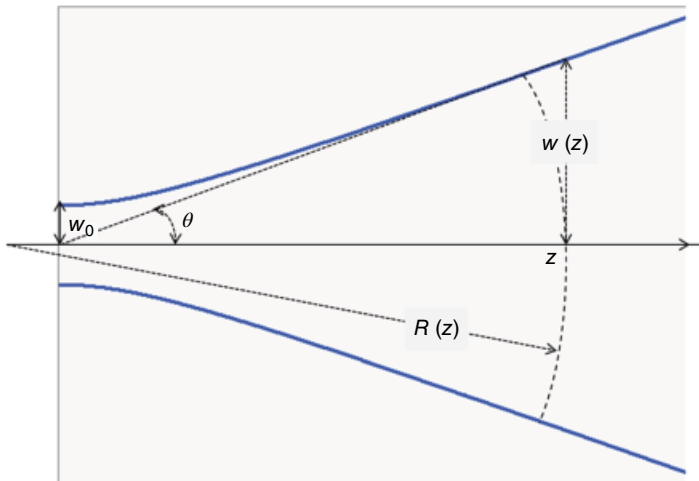


FIGURE 54.4 The contour of a Gaussian beam along the z -axis.

The solution to the wave equation can be expressed as [2]

$$u(r, z) = \frac{w_0}{w} \exp \left\{ -i(kz - \Phi) - r^2 \left(\frac{1}{w^2} + \frac{ik}{2R} \right) \right\}$$

where

$$\Phi = \tan^{-1} \left(\frac{\lambda z}{\pi w_0^2} \right)$$

This solution is not the only solution to the scalar wave equation; however, it is the desired output mode for most laser applications and is referred to as the “*fundamental mode*.” More complicated solutions to the wave equation exist and are generally referred to as “higher-order” modes. Solving the equation in Cartesian coordinates for the higher-order modes generally yields a combination of Gaussian and Hermite functions [2]. Solving equation in cylindrical coordinates (for systems with cylindrical symmetry) yields generalized Laguerre polynomials for solutions [2].

54.1.3.3 Beam Propagation The path of a ray can be characterized by its radial distance from the optical axis (r) and by its slope with respect to the optical axis (θ), as described in Figure 54.5a. The path of this ray through an optical element is then dependent on the optical properties of the element and the input beam parameters. The parameters of the output beam can then be determined by using the ABCD matrix of the optical element:

$$\begin{bmatrix} r' \\ \theta' \end{bmatrix} = \begin{bmatrix} A & B \\ C & D \end{bmatrix} \begin{bmatrix} r \\ \theta \end{bmatrix}$$

where r and θ are parameters describing the input beam and r' and θ' describe the output beam. The ABCD matrix describes the optical properties of the element and is also known as the ray transfer matrix.

As the beam travels through the optical elements, the ray transfer matrices can be stacked to determine the output beam parameters. In Figure 54.5b, an input beam

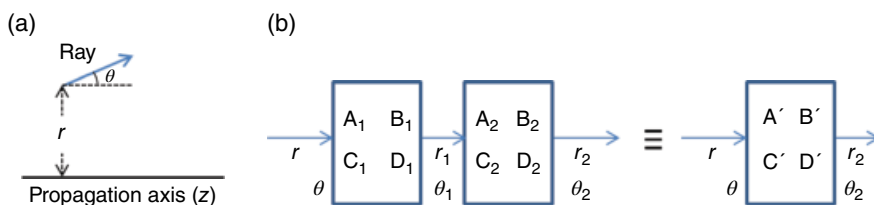


FIGURE 54.5 (a) Shows the beam parameters (r and θ) for a paraxial ray and (b) demonstrates stacking of ABCD matrices.

(parameters r and θ) travels through two optical elements sequentially. The stacked ABCD matrix for this system is given by

$$\begin{bmatrix} A' & B' \\ C' & D' \end{bmatrix} = \begin{bmatrix} A_2 & B_2 \\ C_2 & D_2 \end{bmatrix} \begin{bmatrix} A_1 & B_1 \\ C_1 & D_1 \end{bmatrix}$$

The same process can be generalized for N optical elements; if the input beam passes through N elements in sequence, going from element 1 to element N , and if the corresponding ray transfer matrices for each element are labeled $[T_1]$, $[T_2]$, ..., $[T_N]$, then the ray transfer matrix for the entire system $[T_{\text{tot}}]$ can be evaluated as

$$[T_{\text{tot}}] = [T_N][T_{N-1}] \cdots [T_2][T_1]$$

Table 54.1 gives the ABCD matrices of several common optical elements.

While the discussion has focused on ray optics, ray transfer matrices can also be used to map the effect of optical elements on laser beams. For a Gaussian laser beam, the complex q parameter is defined as

$$\frac{1}{q} = \frac{1}{R} - i \frac{\lambda}{\pi w^2}$$

where λ is the laser wavelength and $R(z)$ and $w(z)$ are the radius of curvature and beam radius at a point z on the optical axis.

If q_1 is the complex beam parameter for a Gaussian beam entering an optical system characterized by ray transfer matrix $\begin{bmatrix} A & B \\ C & D \end{bmatrix}$, then the output beam parameter q_2 is given by

$$q_2 = \frac{Aq_1 + B}{Cq_1 + D}$$

This is called the ABCD law for Gaussian beams.

TABLE 54.1 ABCD Matrices of Common Optical Elements^a

Propagation through free space by a distance d	$\begin{pmatrix} 1 & d \\ 0 & 1 \end{pmatrix}$
Propagation through a medium of refractive index n by a distance d	$\begin{pmatrix} 1 & d/n \\ 0 & 1 \end{pmatrix}$
Reflection at a mirror of the radius of curvature R for normal incidence	$\begin{pmatrix} 1 & 0 \\ -2/R & 1 \end{pmatrix}$
Thin lens of focal length “ f ”	$\begin{pmatrix} 1 & 0 \\ -1/f & 1 \end{pmatrix}$
Dielectric interface	$\begin{pmatrix} 1 & 0 \\ 0 & n_1/n_2 \end{pmatrix}$

^aRef. 2.

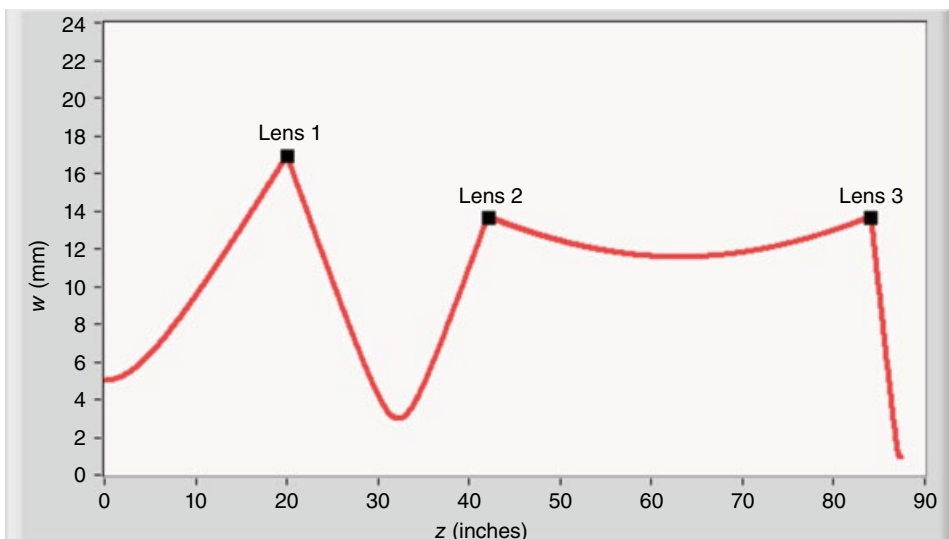


FIGURE 54.6 Tracking the beam diameter of a 500 μm Gaussian laser beam with propagation along the optical axis through three lenses using the ABCD law.

Using the ABCD law the size of a Gaussian beam propagating through an optical system can be mapped as a function of distance along the propagation axis. Figure 54.6 depicts the beam radius as function of propagation distance for a beam of wavelength 500 μm . The beam travels along the axis and interacts with three lenses of varying focal lengths placed at specific locations along the optical path. ABCD matrices for both the propagation distances and lenses determine the beam's profile along the propagation axis.

As shown in Figure 54.6, the beam diameter varies considerably with position. This variation can be controlled using focusing elements placed along the optical axis. For example, in Figure 54.6, lens 2 is used to collimate the Gaussian beam as it propagates through free space. Another aspect of laser optics that the ABCD law yields is the required optic size. As Figure 54.6 indicates, the beam size varies considerably with location on the optical axis; this defines the size of optical elements (mirrors, lenses) that are required to effectively redirect or reshape the beam. For a Gaussian beam of beam radius $w(z)$, a good rule of thumb for determining required optic size is that the diameter of the optic should be larger than $4w$ at that point on the axis. Smaller optics induce distortions on the beam profile.

54.1.4 Example: CO₂ Laser Pumped Far-Infrared Gas Laser Systems

As described in Section 54.1.1, lasers require a cavity/oscillator, a gain medium, and a pumping process. This pump process supplies the energy required to create inversion. This can be achieved using a flashlamp broadband source (as with the previously discussed ruby laser), an electrical voltage, or another laser as well. A specific example of this technique is far-infrared (FIR) gas lasers that can be “pumped” using CO₂ gas

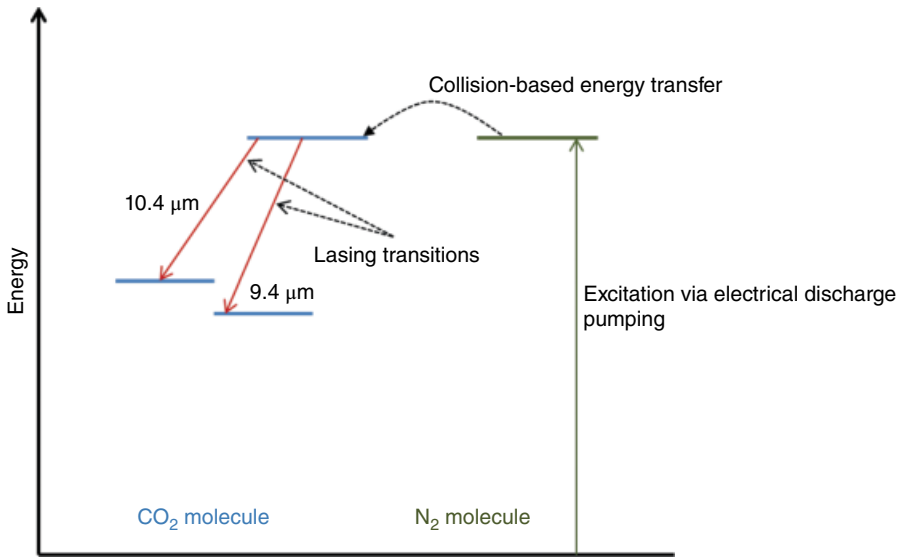


FIGURE 54.7 Energy-level diagram for CO₂ gas laser.

lasers. By tuning the CO₂ laser's frequency for an appropriate selection of gain media (gas), single-frequency laser systems can be adjusted to cover the terahertz (FIR) portion of the electromagnetic spectrum. This example describes the setup of such a system, along with the relevant measurement techniques.

A CO₂ laser is a gas laser that lases at approximately 10 μm. The lasing medium is a mix of CO₂, nitrogen, and helium. The lasing transitions are in the CO₂ molecule. Figure 54.7 shows the energy-level diagram for the relevant lasing levels of the CO₂ molecules. Typical CO₂ lasers are pumped via a process called gas-discharge pumping, wherein a current is arced through the gas-discharge tube containing the gas mixture. Collisions between electrons in the discharge and nitrogen molecules excite the nitrogen molecules to a vibrational excitation mode that closely corresponds to the desired vibrational mode in the CO₂ molecule as shown in Figure 54.7. Collisions between the nitrogen and CO₂ molecules excite the CO₂ molecules to this excited metastable state (asymmetric stretch vibrational mode), and population inversion is obtained. The lower lasing level of a CO₂ laser is actually two-level bands: one corresponding to a transition of 10.4 μm (symmetric stretch) and the other corresponding to 9.4 μm (bending mode).

After the stimulated emission to the lower CO₂ levels, they are “deexcited” by collisions with helium molecules, which act as a buffer gas. Thus, the gas mix is of critical importance for achieving high laser output power with standard CO₂ lasers. The nitrogen is needed to effectively excite the upper lasing levels of the CO₂ molecule, and the helium is needed as a buffer gas to deexcite the CO₂ molecule from the lower lasing levels.

The output frequency of the CO₂ laser can be tuned within a certain frequency bandwidth. As pointed out in Section 54.1.2.3 any laser cavity has several axial modes; the

TABLE 54.2 Far-Infrared Laser Transitions Pumper by CO₂ Lasers^a

Gas	Frequency (GHz)	CO ₂ Pump Line
HCOOH formic	247.08	9.62P28
HCOOH formic	381.34	9.17R40
HCOOH formic	403.72	9.17R40
HCOOH formic	561.72	9.23R28
HCOOH formic	584.39	9.23R28
HCOOH formic	673.99	10.53P14
HCOOH formic	692.95	9.27R20
HCOOH formic	740.23	9.60P26
HCOOH formic	832.99	9.20R34
HCOOH formic	991.78	9.37R04
CH ₂ F ₂ difluoromethane	586.17	9.23R28
CH ₂ F ₂ difluoromethane	1110.32	9.26R22
CH ₂ F ₂ difluoromethane	1267.08	9.35R06
CH ₂ F ₂ difluoromethane	1272.12	9.21R32
CH ₂ F ₂ difluoromethane	1397.12	9.20R34
CH ₂ F ₂ difluoromethane	1546.08	9.26R22
CH ₂ F ₂ difluoromethane	1626.6	9.21R32
CH ₂ F ₂ difluoromethane	1891.27	9.47P10
CH ₂ F ₂ difluoromethane	2237.3	9.57P22
CH ₂ F ₂ difluoromethane	2447.97	9.26R22

^aAdapted from various sources.

spacing between these axial modes depends upon the cavity length. Typical CO₂ lasers that are used for research also utilize a blazed grating as one of the reflectors in the cavity. Appropriate orientation of the grating end and cavity length allows for research-grade CO₂ lasers to lase specific transitions within a certain frequency bandwidth [3]. Figure 54.1 depicts the vibrational energy levels of the CO₂ molecule, and for each vibrational level there are several rotational levels. The J quantum number labels the rotational state of the molecule. When transitions take place between energy levels, certain quantum mechanical selection rules apply. These selection rules limit the change in the rotational quantum state (ΔJ) to span -1 to $+1$, though for many vibrational transitions $\Delta J=0$ is forbidden. To accurately describe which transition is lasing in a CO₂ laser, we need to define which vibrational transition occurred (10.4 or $9.4\mu\text{m}$), the change in J , and the final value of J . Knowing all of these quantities determines which exact rovibrational transition occurred (and hence the wavelength of the emitted photon). For CO₂ laser transitions the convention is to indicate which wavelength (9 or 10), followed by ΔJ (P for -1 and Q for $+1$), followed by the final J value, for example, $10P(20)$, is a $10.4\mu\text{m}$ transition from a $J=21$ to $J=20$ rovibrational state in the CO₂ molecule.

The CO₂ laser can be used as a pump laser for FIR gas lasers. As shown in Table 54.2, the output frequency of CO₂ lasers can be tuned to efficiently pump lasing transitions

in a gas laser. Selecting an appropriate laser gas media and CO₂ laser transition allows one to lase at a range of frequencies in the FIR. Table 54.2 lists some of these gases, the lasing frequency (FIR), and the CO₂ pump frequency. Notice that pumping different transitions in the same gas molecule leads to different FIR laser frequencies. Also note that sometimes the same pump laser frequency can excite multiple lasing transitions in the same gas. In such situations, the exact frequency of the FIR laser is determined by the relative gain of the lasing transitions and construction of the cavity. Reference textbooks [4] list measured transitions and the appropriate CO₂ pump frequencies.

54.1.5 Heterodyned Detection

Laser measurement systems focus on studying the interaction of laser radiation with different materials. For several applications it is the laser beam that has interacted with the sample (e.g., transmitted through or reflected from) that is of interest for the measurement. Detecting the laser beam requires a detector that is sensitive to light at the laser frequency. Depending on the frequency range of the laser, several detection options with varying sensitivities are available. The basic principle of the detector is that it produces a measurable output for a given amount of laser input.

One way to broadly classify detection techniques is coherent versus incoherent detection. For incoherent detection the output signal is dependent on the intensity of laser beam. Coherent detection schemes allow for the determination of phase information in the laser measurement as well as field strength. This section discusses heterodyned detection, which is a technique commonly used in telecommunications, astronomy, and FIR measurement systems for coherent detection.

The basic principle of heterodyned detection requires a detector that has a nonlinear output with input electric field amplitude. An example is a diode mixer, the output of which is proportional to the square of the electric field amplitude.

Consider a detector where the output current is proportional to the square of the input amplitude:

$$I \propto E^2$$

We have two input signals, a transmit signal given by

$$E_t = E_t \cos(\omega_t t + \phi_t)$$

where E_t is the amplitude of the transmit beam, ω_t is the frequency, and ϕ_t is the phase and a Local Oscillator (LO) signal given by

$$E_{LO} = E_{LO} \cos(\omega_{LO} t + \phi_{LO})$$

where E_{LO} is the amplitude of the LO, ω_{LO} is the frequency, and ϕ_{LO} is the phase.

If these two signals are simultaneously incident on the detector, the output is then proportional to the square of the input amplitude if the input amplitude is the superposition of these two signals:

$$I \propto [E_t \cos(\omega_t t + \phi_t) + E_{LO} \cos(\omega_{LO} t + \phi_{LO})]^2$$

$$\rightarrow I \propto E_t^2 \cos^2(\omega_t t + \phi_t) + E_{LO}^2 \cos^2(\omega_{LO} t + \phi_{LO}) + 2E_t E_{LO} \cos(\omega_t t + \phi_t) \cos(\omega_{LO} t + \phi_{LO})$$

Using trigonometric identities and rearranging the terms yield

$$I$$

$$\propto \underbrace{\frac{1}{2}(E_t^2 + E_{LO}^2)}_{\text{DC term}} + \underbrace{\left[\frac{E_t^2}{2} \cos 2(\omega_t t + \phi_t) + \frac{E_{LO}^2}{2} \cos 2(\omega_{LO} t + \phi_{LO}) + E_t E_{LO} \cos((\omega_t + \omega_{LO})t + (\phi_t + \phi_{LO})) \right]}_{\text{High frequency terms}}$$

$$+ \underbrace{E_t E_{LO} \cos((\omega_t - \omega_{LO})t + (\phi_t - \phi_{LO}))}_{\text{Intermediate frequency term}}$$

If at this point the signal is passed through a low-pass filter and the DC term is subtracted, the output of the detector is at the difference frequency also called the Intermediate Frequency (IF):

$$I \propto E_t E_{LO} \cos(\omega_{IF} t + \Delta\phi)$$

Thus looking for signal at the IF yields information corresponding to the amplitude and phase of the laser beam. Figure 54.8 shows the schematic for coherent single-frequency detection of the radiation reflected back from some target. The transmit laser and the LO are both FIR laser beams such as those generated by the FIR laser systems described in the previous sections. The reference diode and receiver diode are both nonlinear mixers such as Schottky diodes. The reason for a heterodyned approach is that FIR frequencies are too high for electronics to respond; thus a heterodyned system is used to down-convert the frequency of the laser signal.

In Figure 54.8, the purpose of the measurement is to measure the response (reflection) from the target at the transmit frequency (ω_t). The LO frequency is offset from the transmit frequency by an amount called the IF. A series of beam splitters (an optical device used for partially reflecting and transmitting beams, discussed later) is used to combine the transmit and LO signals at the two detectors. Following the beam path in Figure 54.8 shows that the transmit signal output from the laser source is overlapped with the LO at Beam Splitter 4 (BS4) and sent to the reference detector, while the transmit laser signal reflected from the target is combined with the same LO at BS5 and sent to the receiver diode. The reference diode is used to measure the amplitude

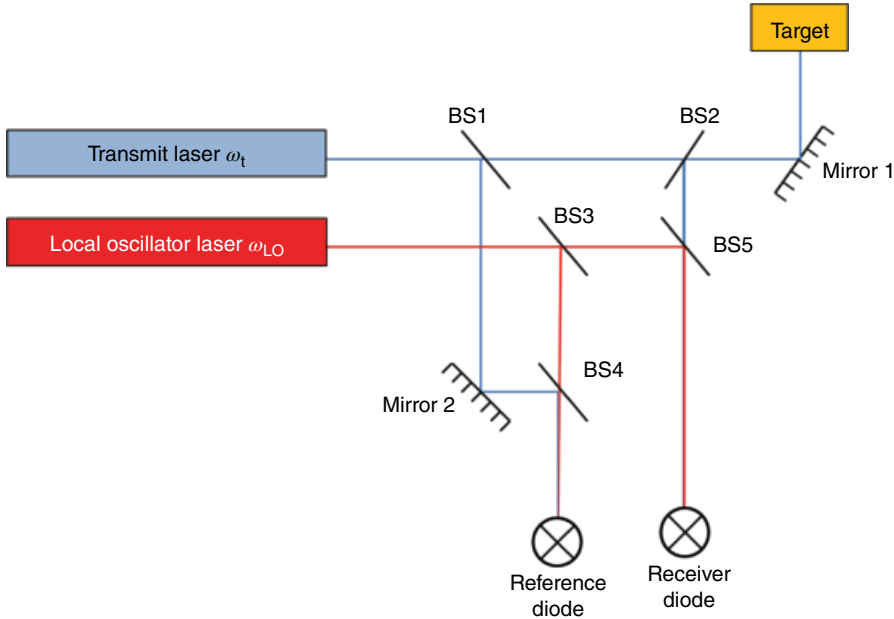


FIGURE 54.8 Schematic of heterodyned detection using a transmit laser and local oscillator.

and phase of the signal that has not been modified by the target, while the receiver diode measures the signal that has been modified by the sample. Comparing these two signals allows one to compute the change in amplitude and phase of the transmitted beam caused by the target.

A Schottky diode can be used as the required nonlinear mixer for heterodyned detection. Further details can be found in *Microwave Engineering* by Pozar [5]. The small signal current–voltage (IV) relationship of a Schottky diode is given as

$$I(V) = I_s (e^{\alpha V} - 1)$$

where $\alpha = q/nkT$, where q is electron charge, k is Boltzmann's constant, T is temperature and n is the ideality factor for the diode. I_s is the diode saturation current [5].

Now if the voltage across the diode is a DC bias voltage (V_{DC}) and a small AC voltage (v),

$$V = V_{DC} + v$$

Expanding the diode current as a Taylor series about V_{DC} and keeping the first three terms yield

$$I(V) = I_{DC} + v(\alpha(I_{DC} + I_s)) + \frac{1}{2}v^2(\alpha^2(I_{DC} + I_s)) + \dots$$

This is termed the *small signal approximation* for diode current. Note that the output is proportional to the square of the input AC voltage as is required for a nonlinear mixer in a heterodyned receiver.

In later sections we discuss the Noise Equivalent Power (NEP) for incoherent detection schemes. For incoherent detection setups, which are sensitive to the intensity of the laser signal, the NEP is proportional to the square root of the dwell time, and detector NEPs are quoted as $\text{W/Hz}^{1/2}$. For coherent detection, such as the heterodyne system described earlier, the NEP is proportional to the dwell time and measured in units of W/Hz . Thus coherent detection offers better noise Figures for the same dwell time per point when compared with incoherent detection. Mixers can also be used to generate radiation; if instead of the transmit and LO as inputs we provided a transmit frequency and IF input, then the output of the diode mixer will include frequencies of $\text{transmit} \pm \text{IF}$. If the IF is varied over a range that the mixer can respond, the output frequency of the mixer is the lasing frequency modulated over the varying bandwidth. This process is called “frequency chirping.”

54.1.6 Transformation of Multimode Laser Beams from THz Quantum Cascade Lasers

54.1.6.1 Quantum Cascade Lasers Quantum Cascade Lasers (QCLs) are semiconductor lasers based on intersubband transitions within the conduction band in quantum wells. This differs from semiconductor diode lasers where the lasing transition is interband (between conduction and valence bands in the semiconductor). The intersubband energy levels responsible for lasing transitions in QCLs are based on layer thicknesses in semiconductor materials and can be tailored, using appropriate growth and fabrication techniques, over a wide frequency range.

The schematic energy diagram of a QCL is shown in Figure 54.9. Barriers and wells (within the conduction band) are created by combining semiconductor materials with different bandgaps. Applying a bias voltage across the structure allows electrons to tunnel through the structure. The laser structure essentially consists of several periods; within each period there is an active region and an injector. The lasing transition occurs in the active region (level 3 to level 2 in Fig. 54.9), and then after relaxation processes (from level 2 to level 1 in Fig. 54.9), the electron tunnels across the barrier and is “injected” into the upper lasing levels of the next active region, where the process repeats. As the number of periods can be quite large, a single electron is capable of generating several photons as it cascades down the energy levels.

Faist et al. demonstrated the first QCL in 1994 with an emission frequency of $4.2\ \mu\text{m}$ based on a superlattice structure [7]. Since then QCLs have been designed and fabricated to span the near- to mid-infrared spectrum quite effectively. FIR or terahertz QCLs are currently under development, and while some have been demonstrated, there are significant challenges to lowering the operating frequency.

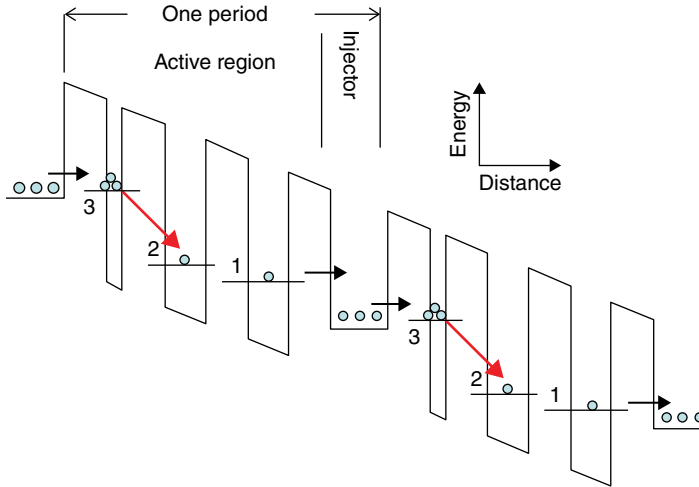


FIGURE 54.9 Conduction band energy diagram of two periods of a QCL. Source: Danylov [6]. Reproduced with permission from Andriy Danylov.

54.1.6.2 Transforming Multimode Laser Beams into Gaussian Beams As discussed in Section 54.1.2.3, laser oscillators are capable of supporting several discrete eigenmodes depending on the cavity's geometry. For several applications of laser beams (specifically imaging applications as discussed in later sections), the Gaussian output mode (TEM_{00}) is desired. Terahertz QCLs suffer from multimode output beams that diverge rapidly. A dielectric tube can be attached to the output face of the waveguide in order to improve the beam profile of the THz QCL. The method is equally relevant for other lasers with multimode, divergent output beams.

The basic idea is to attach a hollow dielectric tube to the laser output face (Fig. 54.10). Since the beam propagates in the hollow core, losses are minimal. The lossy dielectric material assists in cleaning up the mode profile as higher-order modes are more severely attenuated in the waveguide. The modes that propagate in these circular cylindrical structures are of three basic types: transverse circular magnetic (TM_{0m}), transverse circular electric (TE_{0m}), and nontransverse hybrid (EH_{nm}). If the dielectric waveguide has a radius “ a ,” which is significantly larger than the free space wavelength of the laser beam (λ), then in making this approximation the EH_{1m} modes are transverse and linearly polarized. The far-field pattern of the EH_{11} mode essentially resembles a Gaussian profile. The attenuation constant for a given mode is proportional to λ^2/a^3 ; thus larger-diameter waveguides experience lower propagation loss.

For a THz QCL operating at 2.960 THz, University of Massachusetts (UMass) Lowell researchers selected a hollow pyrex tube of inner diameter 1.8 mm with a tube length of 43 mm [8]. Thus λ/a is negligible ($\lambda=101.4\mu\text{m}$) and the EH_{1m} mode is transverse and linearly polarized. They found that the TE_{01} mode is significantly less lossy than the TM_{01} mode; however it exhibited slightly higher loss than the TE_{01} mode. However, as both the TE and TM modes are not linearly polarized like the QCL, the

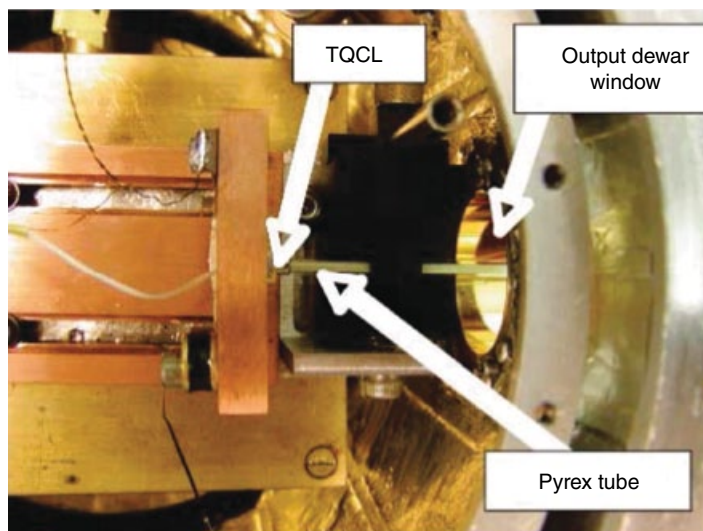


FIGURE 54.10 Photograph of terahertz QCL with dielectric tube attached at the output. Source: Danylov [8]. Reproduced with permission from the Optical Society.

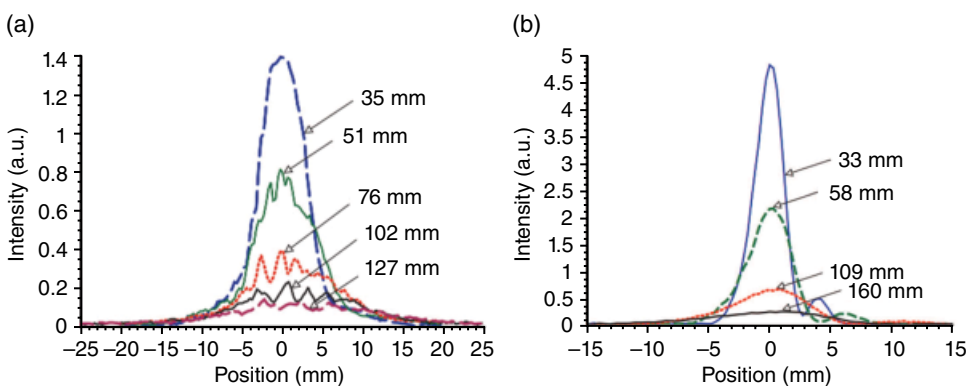


FIGURE 54.11 (a) THz QCL beam profile as a function of distance from laser output end. (b) Beam profile of the same laser as a function of distance after the dielectric tube was inserted. Notice the higher-order modes are essentially replaced by a Gaussian mode [8].

EH_{11} linearly polarized mode is dominant within the waveguide. This mode couples efficiently (98%) to the TEM_{00} (Gaussian) mode in free space, provided that the beam waist to tube radius ratio is 0.6435.

As can be seen in Figure 54.11, Danylov et al. measured the output beam profile of a THz QCL as a function of beam propagation distance both without and with a hollow dielectric waveguide (Fig. 54.11a and b, respectively). They were able to reduce the beam divergence and propagate a Gaussian mode in free space. The primary disadvantage to this technique is the loss in beam power due to the attenuation losses within the waveguide and coupling losses into the waveguide.

54.1.7 Suggested Reading

A very good standard reference textbook for understanding laser fundamentals is *Lasers* by Anthony E. Siegman, published by University Science Books. A standard textbook for optics is *Principles of Optics* by Born and Wolf, published by Cambridge University Press. A more basic optics textbook is *Optics* by Eugene Hecht, published by Pearson Education Limited. Another good reference textbook for laser optics is *Quantum Electronics* by Amnon Yariv, published by John Wiley & Sons.

54.2 LASER MEASUREMENTS: LASER-BASED INVERSE SYNTHETIC APERTURE RADAR SYSTEMS

Round-the-clock detection and location of man-made structures in all weather conditions have long been of interest. Active illumination *Radio Detection and Ranging* (RADAR) systems have provided this capability through the coherent signal processing of propagating pulses transmitted and received using directional antennas. Since radio waves propagate at the speed of light c , the time delay (t_{delay}) between the transmitted and received pulses provides a measure of the round-trip distance to the object detected where the object's range r from the radar can be expressed as $r = c t_{\text{delay}}/2$. Using large aperture antennas to direct the frequency-chirped narrow beam pulses illuminating the scene, range (distance) and angular information may be acquired to form two-dimensional (2D) or three-dimensional (3D) imagery. The resolution of the imagery is inversely proportional to the bandwidth of the frequency chirp.

To produce azimuth resolution in the imagery, in 1951 Carl Wiley proposed coherent pulse-to-pulse processing of the backscattered signal captured from the scene while moving the radar. Since this technique artificially increased the antenna aperture, the process of forming these images was referred to as Synthetic Aperture Radar (SAR). Given the same imagery could be formed by pulse-to-pulse processing of backscattered signals from a rotating object with stationary radar, the popularity of collecting high-resolution radar imagery in controlled environments through Inverse Synthetic Aperture Radar (ISAR) also grew, and a large number of turntable radar facilities were established.

By the late 1970s, radar scientist Dr. Jerry Waldman [9], working at MIT Lincoln Laboratories in Bedford, Massachusetts, recognized the potential to acquire radar ISAR imagery in a laboratory setting by using very-high-frequency radar beams and scale models of objects of interest. Drawing from Maxwell's equations the use of millimeter-wave (wavelengths near, but longer than 1 mm) and submillimeter-wave radar beams (wavelengths shorter than 1 mm), the proportional size of the target and millimeter wave beam would reproduce a full-size ISAR turntable measurement.

Waldman et al. constructed an optically pumped submillimeter-wave laser and "compact" radar range and demonstrated the concept of calibrated signature similitude.

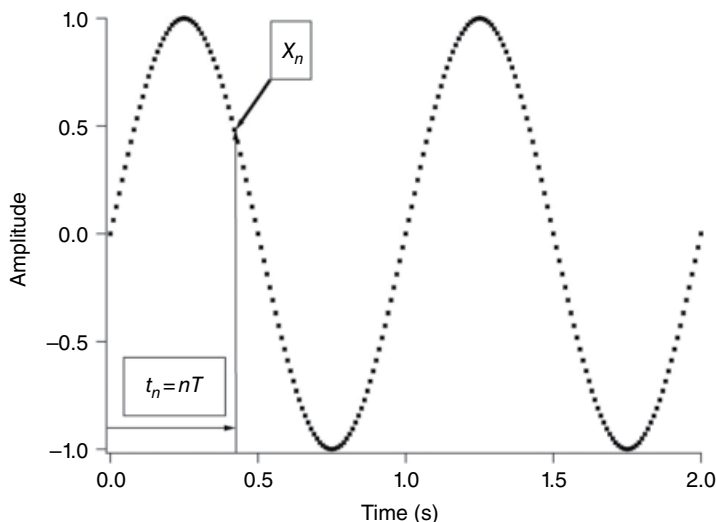


FIGURE 54.12 Example of discretely sampled data as a function of time.

The first imagery was published in the fall of 1979. Today, the UMass Lowell Submillimeter-Wave Technology Laboratory continues to produce high-quality and high-accuracy radar signatures of objects of interest over every radar band and has received numerous grants from DARPA, NASA, DoD, DHS, and the NSF.

54.2.1 ISAR Theory

54.2.1.1 Digital Fourier Transform The primary mathematical tool that is used to analyze radar data into imagery is the Fourier Transform (FT). Consider the discrete sampling of data in Figure 54.12. The data in this Figure can be represented by the array $\{x_n\}$ where $n=0$ to $N-1$. In general the discrete values of this array are composed of complex numbers and therefore

$$x_n = A_n e^{i\phi_n}$$

By using the Discrete Fourier Transform (DFT), the digitized time-varying signal $\{x_n\}$ can be represented as a summation of N uniformly spaced sinusoidal phasors such that

$$x_n = \frac{1}{N} \sum_{k=0}^{N-1} X_k e^{i\omega_k t_n}$$

where

$$t_n = nT$$

$$\omega_k \equiv \frac{2\pi k}{NT}$$

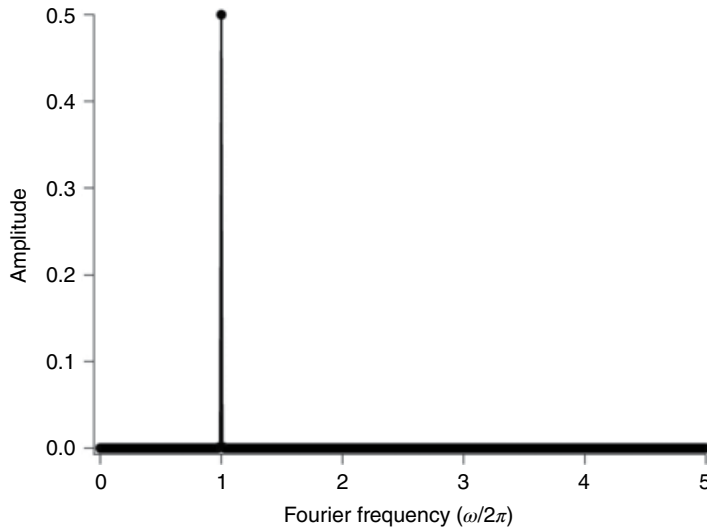


FIGURE 54.13 Discrete Fourier transform of a 1 Hz sine wave as a function of Fourier frequency.

In general the coefficients X_k are complex numbers as well such that

$$X_k = A_k e^{i\phi_k}$$

The values of these Fourier coefficients are given by

$$X_k = \sum_{n=0}^{N-1} x_n e^{-i\omega_k t_n}$$

Figure 54.13 shows a sample calculation for the data given in Figure 54.12. In this case the data represents a 1 Hz sine wave, and when a DFT is applied to the data array, only the Fourier coefficient at $f = \omega/2\pi = 1$ Hz is nonzero. In this example the result of the DFT is to represent the spectral content of the original time-domain signal in the frequency-domain space.

54.2.2 DFT in Radar Imaging

The Fourier analysis described in the previous section can be optimized for use in radar analysis and image formation. Figure 54.14 shows the general geometry for a single isolated object illuminated by a radar beam. In general, objects are more complex. However, any complex object can be viewed as a linear sum of individual scattering centers, each of which can be considered individually. Therefore, the geometry of Figure 54.14 is still the fundamental approach to the formation of images.

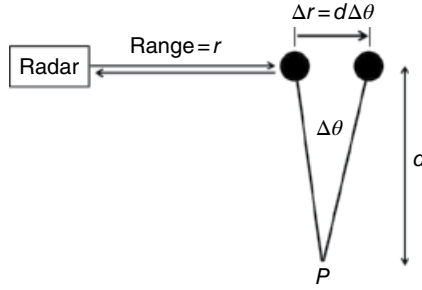


FIGURE 54.14 Range shift of a target rotated through an angle $\Delta\theta$.

In this Figure the object that is in the radar beam has a *total phase* due to its distance and the wavelength of the radar of

$$\varphi(r) = \frac{2\pi}{\lambda} [2r] = \frac{4\pi r}{\lambda}$$

Note the fact that there is an extra factor of 2 due to the round-trip path of the radar. If the target is located a distance d from the center of rotation and the target is rotated through an angle $\Delta\theta$, this equation becomes

$$\Delta\varphi(r) = \frac{4\pi(r_2 - r_1)}{\lambda} = \frac{4\pi}{\lambda} d \cdot \Delta\theta$$

In terms of a discrete sampling of points, the phase is given by

$$\Delta\varphi_n = \frac{4\pi(r_2 - r_1)}{\lambda} = \frac{4\pi}{\lambda} d \cdot \theta_n$$

where

$$\theta_n = n \cdot \Delta\theta_0$$

Therefore the discretely sampled signal in the time domain can be represented by

$$x_n = A_n e^{i \left[\frac{4\pi}{\lambda} d \cdot \theta_n \right]}$$

The DFT coefficients can then be written as

$$X_k = A \sum_{n=0}^{N-1} e^{i \left\{ \left[\frac{4\pi}{\lambda} d \cdot \theta_n \right] - \omega_k t_n \right\}}$$

It can be shown that the value of X_k rapidly approaches 0 except for the case where

$$\left[\frac{4\pi}{\lambda} d \cdot \theta_n \right] - \omega_k t_n = 0$$

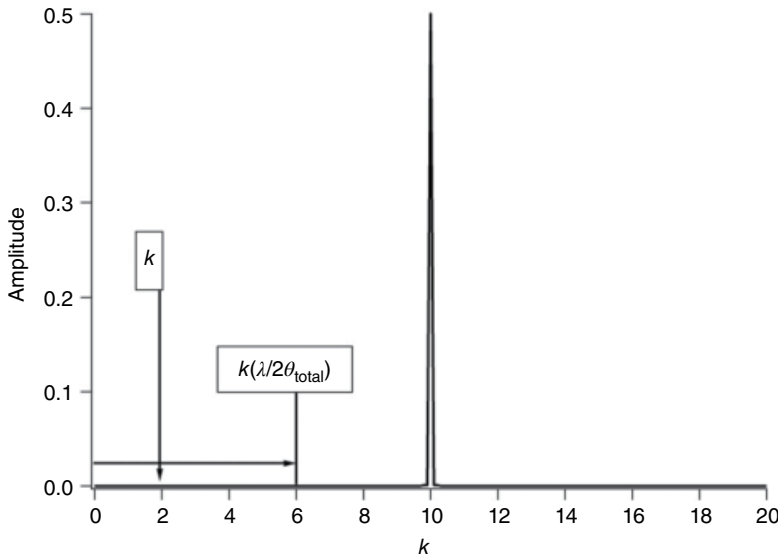


FIGURE 54.15 Fourier transform of an arbitrary signal from a target rotated through the radar beam. Each individual value of k represents another increment of the cross-range resolution.

Solving for d gives the result

$$d = k \left[\frac{\lambda}{2\theta_{\text{total}}} \right]$$

The physical interpretation of the value of k that is the integer vector from the DFT is that of equally spaced distances. An example is given in Figure 54.15. In this Figure the results of the DFT are displayed where the amplitude of the Fourier components are plotted versus the integer index k . The time-domain signal that has been transformed represents an object slowly turned through the radar beam in a manner to that shown in Figure 54.14. Complex amplitude and phase information is collected at uniform increments of θ . The resultant DFT yields a nonzero Fourier amplitude only when the constraint in the previous equation is satisfied. The quantity $[\lambda/2\theta_{\text{total}}]$ is known as the cross-range resolution. The example of Figure 54.15 shows that the target is located 10 resolution bins from the center of rotation.

The same analysis can be used for measuring the range seen in Figure 54.14 directly by calculating the phase change as a function of changing frequency. Following the same formalism given earlier it can be shown that for a series of complex measurements taken at uniformly spaced frequencies, the down-range position will be given by $r = k[c/2B]$. In this equation c is the speed of light and B is the total bandwidth that the radar frequency is changed by. The formalism given above can be extended to 2D and 3D measurements in order to form radar imagery.

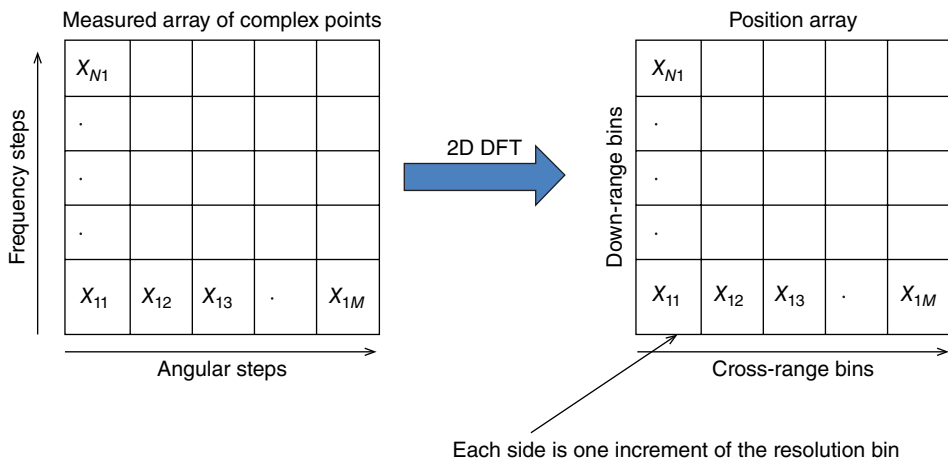


FIGURE 54.16 An example of the analysis of measured complex voltages as a function of frequency and angle transformed into positions by use of a Fourier transform.

Figure 54.16 shows a representation of a 2D data collection and DFT varying the angle and the frequency in a controlled manner. Once the DFT is performed, that index of the resulting array has the meaning of the down-range and cross-range resolutions that are given earlier. The frequency and angular information are thereby converted directly into a 2D array where the position within that array has the meaning of k times the resolution size.

54.2.3 Signal Processing Considerations: Sampling Theory

Theoretically in the course of using the FT, one is unfolding continuous periodic signals with infinite extent. Since practical measurements involve discrete sampling over limited bandwidths and displacements, ambiguities and spurious responses can form as a result of the sampling process. These spectral discontinuities at the end of the measurement interval cause spurious responses in the image. Multiplicative weighting functions (referred to as windowing, W) can be applied to smoothly taper the signal to zero at the ends of the measurement interval and reduce the effect of these discontinuities. The windowing can be applied directly to discrete sequenced signals within FFT in the form of

$$X_k = \sum_n x_n W_n e^{\frac{-i2\pi kn}{N}}$$

for n and $k=0$ to $N-1$.

By executing the FFT on the measured sequence of N backscattered signals, x_n , as a product with the windowing function, W_n , the computed signal values (i.e., resolution cells), X_k , will still represent the scattered signal at uniformly spaced locations along the range direction, but with reduced discontinuities.

54.2.4 Measurement Calibration

The electric field of an electromagnetic wave traveling in the z -direction is given by the equation

$$\vec{E} = E_{0x} \cos(\omega t - kz + \alpha_x) \hat{x} + E_{0y} \cos(\omega t - kz + \alpha_y) \hat{y}$$

where ω represents the angular frequency, t is time, k is the wavenumber, and α is an absolute phase constant. This equation can be rewritten in matrix form such that

$$\vec{E} = \begin{bmatrix} E_x \\ E_y \end{bmatrix}$$

where the components in the x and y directions are given by E_x and E_y . Radar systems normally represent the x -axis of the electric field as an electromagnetic wave whose electric field is in the horizontal plane. Similarly the y -axis is represented in the vertical plane. Thus we relabel the x - and y -axis as H and V to represent the polarization of the radar beam, H for horizontal and V for vertical. The received electric field for a radar system can be written in matrix form as

$$\begin{bmatrix} E_H^r \\ E_V^r \end{bmatrix} = \frac{e^{-ikR}}{2\sqrt{\pi}R} \begin{bmatrix} S_{HH} & S_{HV} \\ S_{VH} & S_{VV} \end{bmatrix} \begin{bmatrix} E_H^t \\ E_V^t \end{bmatrix}$$

where the transmitted signal E^t is multiplied by a scattering matrix S .

Since radar systems are composed of transmitter and receiver optics and electronics, it is necessary to calibrate the response of the system both in polarization and in frequency. The method for this calibration is described in detail by Chen et al. [10] and is briefly given here and in DeMartinis et al. [11]. Consider the measurement matrix S^m given in the equation below. In this equation the measured values S_{HH}^m , S_{HV}^m , S_{VH}^m , and S_{VV}^m are complex numbers and represent the measured amplitude and phase of the back-reflected signal of a target illuminated by a radar beam and received through the radar's optics and electronics. The target itself has reflected the radiation according to the values given in the S matrix such that the ideal radar return is represented by S_{HH} , S_{HV} , S_{VH} , and S_{VV} . The values of S undergo distortions that are represented by the R , T , and I matrices. These matrices are the receiver, transmitter, and isolation distortion matrices, respectively:

$$\begin{bmatrix} S_{HH}^m & S_{HV}^m \\ S_{VH}^m & S_{VV}^m \end{bmatrix} = \begin{bmatrix} I_{HH} & I_{HV} \\ I_{VH} & I_{VV} \end{bmatrix} + \begin{bmatrix} R_{HH} & R_{HV} \\ R_{VH} & R_{VV} \end{bmatrix} \begin{bmatrix} S_{HH} & S_{HV} \\ S_{VH} & S_{VV} \end{bmatrix} \begin{bmatrix} T_{HH} & T_{HV} \\ T_{VH} & T_{VV} \end{bmatrix}$$

The equation given above can be solved for the S matrix by inverting the transmit and receive distortion matrices:

$$[S] = [R]^{-1} ([S^m - [I]]) [T]^{-1}$$

Chen et al. [10] have shown that the R, T, and I matrices can be calculated by scanning a series of known objects with the radar system and solving a series of equations. The known objects are a flat plate, a dihedron with the seam in the horizontal plane, and a dihedron with the seam oriented at an angle q to the horizontal plane. The scattering matrices of the ideal objects are given by

$$S_{\text{plate}} = \begin{bmatrix} -1 & 0 \\ 0 & -1 \end{bmatrix}$$

$$S_{\text{dihedron}, 0^\circ} = \begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix}$$

$$S_{\text{dihedron}, \theta^\circ} = \begin{bmatrix} -\cos 2\theta & \sin 2\theta \\ \sin 2\theta & \cos 2\theta \end{bmatrix}$$

This formalism is used to accurately calculate the undistorted scattering matrix, S, from the measured scattering matrix, S^m .

54.2.5 Example Terahertz Compact Radar Range

Researchers at the UMass Lowell have acquired 2D and 3D ISAR imagery using two CO₂ lasers to optically pumped FIR lasers (Goyette et al. [12, 13]). Configured with a difference frequency of 1.9 GHz between the two 1.56 THz lasers, one laser served as the transmitter, while the second was used as an LO for the heterodyne receiver. To establish a tunable frequency, the laser transmitter was mixed with a microwave sweeper (10–18 GHz) using a Schottky diode. Producing two sideband frequencies that can be swept, the receiver's electronics were designed to only detect the lower sideband. The swept frequency system had sufficient bandwidth to achieve a range resolution of 0.625" (Fig. 54.17).

Using single-frequency terahertz laser data over a 5×5 solid angle measured in azimuth (side to side) and elevation (up and down), UMass Lowell scientists achieved a resolution on the order of a millimeter for 2D cross-range-only imagery. Figure 54.18 shows the complex amplitude and phase data taken across the angular extent in azimuth and elevation. This data has been analyzed in a manner similar to what has been described in the previous sections whereby the data, spaced incrementally in the two angular directions, is Fourier transformed in order to calculate the image. Figure 54.19 shows the result of this DFT. The image of the 1/16th scale model truck is easily recognized once the DFT is plotted in a manner similar to that shown in Figure 54.16. In this case the two axes are variations in the viewing angle producing an image in azimuth cross-range and elevation cross-range.

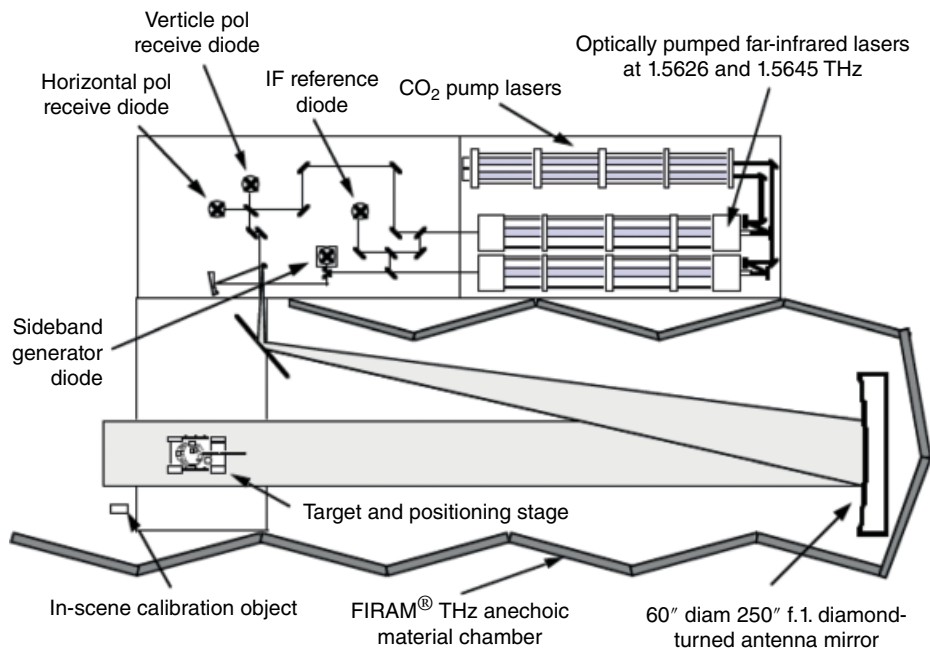


FIGURE 54.17 A 1.56 THz laser-based compact radar range. Source: Goyette et al. [12]. Reproduced with permission from SPIE.

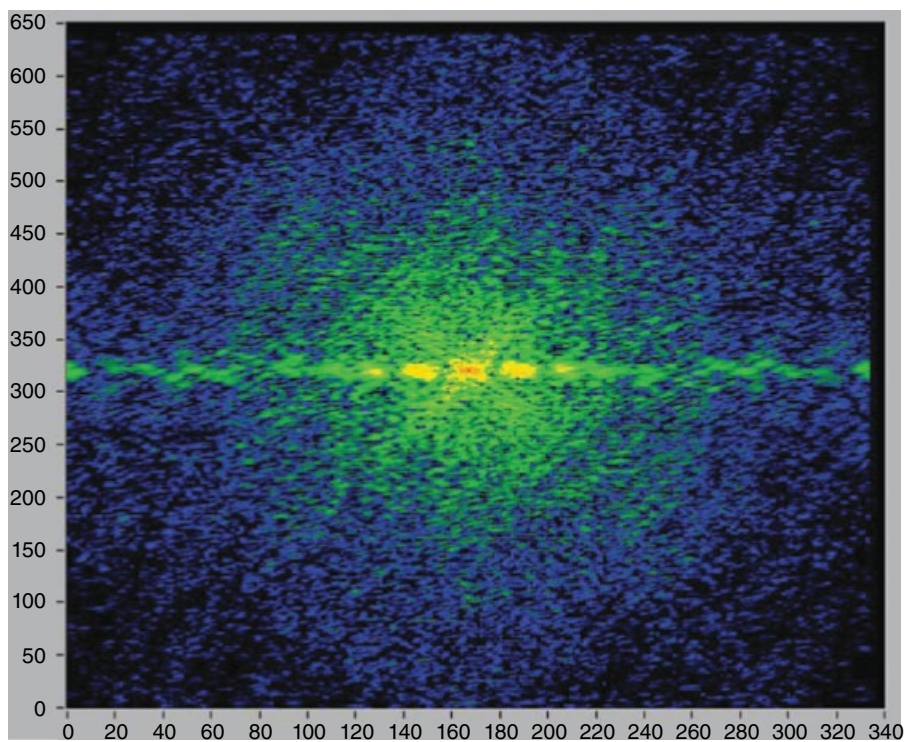


FIGURE 54.18 Amplitude plot of the complex radar return from a truck measured through a 5-degree \times 5-degree solid viewing angle at 1.56 THz.

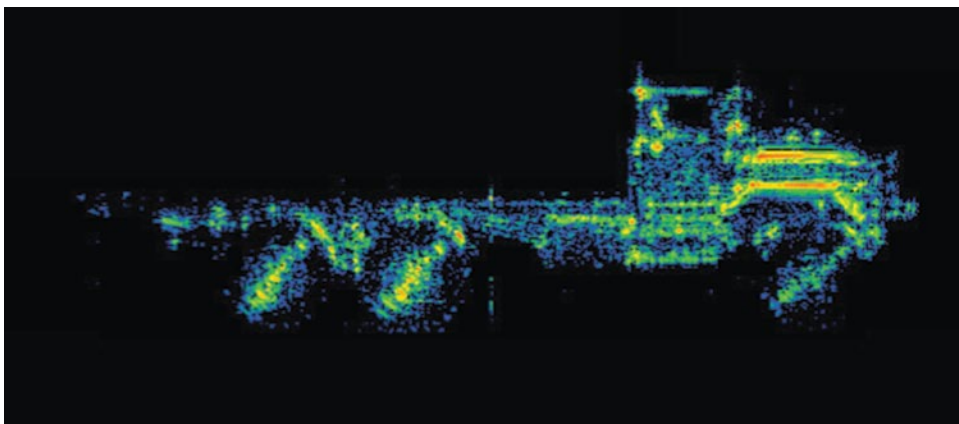


FIGURE 54.19 DFT of the complex radar return from a truck measured through a 5-degree \times 5-degree solid viewing angle, forming a side view image of the target at 1.56 THz.

54.2.6 Suggested Reading

Radar Cross Section by Knott, Shaeffer, and Tuley published by Artech House, Inc. provides a good introduction to radar measurement systems, while *High Resolution Radar Cross-Section Imaging* by Dean L. Mensa also published by Artech House Radar Library also provides an excellent reference for advanced users.

54.3 LASER IMAGING TECHNIQUES

The time required to evolve scientific discoveries into practical products has always been the challenge for any technical community. While teams of researchers with great insight and persistence establish the foundation of breakthrough science, significant engineering is also necessary for maturing technologies with these discoveries to finally create products for new applications.

The time cycle from discovery to product appears to be diminishing as global communication rapidly moves ideas from research to industry internationally. But development of source technologies, like the tunable CO₂ optically pumped FIR lasers, stretched across half a century as many sought to turn this laboratory instrument into a precision product for terahertz (THz) frequency spectroscopic imaging applications. Now with evidence of applications in the fields of forensics, security, medicine, and communications, a number of new THz source/detection technologies offer the promise of commercialization.

From the gamma to microwave frequency regime, lasers in general have always provided a significant platform for materials science. Whether using broadband or narrowband tunable sources, lasers are capable of facilitating precise amplitude,

phase-stable, polarization-sensitive measurements for characterizing the properties of a never-ending list of novel materials.

A laser is essentially a monochromatic coherent light source. The interaction of any material with a laser source is determined by the refractive index of that material, which is a function of frequency. Laser imaging typically examines light, which is reflected from, or transmitted through, a sample illuminated by a laser source. Correlating the collected signal with sample geometry yields the image. In general, when laser light is incident on a material, several processes occur: reflection at the interface (specular and diffuse), transmission through the interface, absorption and scattering within the sample, and in some cases excitation and emission of light (fluorescence). All these processes are determined by the interaction of light with materials and can be used to generate images of the samples that highlight features that are of interest.

In previous sections ISAR imaging was introduced, wherein the entire sample is illuminated by a laser beam and the image is reconstructed by changing the orientation of the sample with respect to the beam. In this section two more conventional approaches to laser-based imaging are presented: point scanning, wherein a laser beam is focused onto a moving sample, and camera imaging, wherein the beam illuminates the entire sample and the response is collected by a camera. When designing laser imaging systems, several factors are considered; these include imaging resolution, system dynamic range, and data acquisition rates. Specific examples of these parameters are discussed.

54.3.1 Imaging System Measurement Parameters

54.3.1.1 Detectors and NEP Apart from a laser source, laser imaging systems require a detector to measure the signal remitted by the sample. In Section 54.1.5, heterodyned detection was discussed. This was an example of a coherent detection scheme; other detection schemes used in imaging applications employ incoherent detection wherein the detector output is proportional to the intensity of the laser beam incident on it.

For example, a liquid helium-cooled bolometer can be used as an incoherent detector for terahertz frequency applications; it is sensitive to the intensity of the terahertz beam. The bolometer operates on the following principle: the silicon semiconductor is placed in a cold bath, which is in contact with the liquid helium ($T \approx 4.2\text{ K}$). The crystal is also connected to an absorber that heats up when terahertz radiation is incident. The amount of heat generated in the absorber is proportional to the intensity of the terahertz beam. As the absorber is in contact with the silicon, the silicon also absorbs heat; this creates electron-hole pairs in the silicon and reduces the resistivity. Thus monitoring the resistivity of the silicon measures the intensity of the terahertz beam. Modulating the incident intensity at a specific frequency allows for tracking changes in beam intensity. This is accomplished by optically chopping the terahertz beam.

The highest modulation/chop frequency is determined by the response time of the detector. The response time of the detector is a measure of how fast a detector can respond to incident signal. It is generally specified by quoting the time constant of the detector (τ). The time constant is the amount of time taken after the detector is exposed to the signal for the detector's output to rise to $(1 - 1/e)$ of its final output value. The modulation frequency can be related to the detector time constant as $f_M = 1/2\pi\tau'$ where f_M is the frequency corresponding to half the peak detector output [14]. For accurate measurements, modulation frequencies are lower than f_M .

For a silicon bolometer, the response time is also dependent on the thermal contact between the silicon and the cold bath; higher thermal conductivity implies that the detector cools down faster. Generally the trade-off is between response time and sensitivity; collecting photons for longer time periods increases the response, while cooling faster decreases the response and provides a faster response time. At terahertz frequencies, the frequency response curve of a bolometer is fairly flat; thus the detector can be used over a range of frequencies.

Also a critical characteristic for an imaging system is the Signal-to-Noise Ratio (SNR). This depends on two factors: the maximum power received at the detector and the NEP of the detector. The NEP of the detector is the signal power required for the SNR to be 1 over a 1 Hz bandwidth [14]. Thus, the smaller the NEP, the better the detector. As discussed previously, the NEP of a coherent detection scheme is measured in W/Hz, while the NEP of an incoherent detector is measured in W/Hz^{1/2}. Thus using the same detector in coherent as opposed to incoherent detection schemes yields higher SNRs. Another system characteristic often confused with SNR is dynamic range. The dynamic range of the system is the difference between the noise floor of the system and the saturation point of the detector.

54.3.1.2 Beam Waist Measurements There are several techniques to measure the beam waist at a point. Discussed here are two common techniques: the knife-edge method and profiling the beam expansion.

For the knife-edge method, a straight edge is moved across the beam perpendicular to the propagation axis, and the transmitted signal is collected by a detector as shown in Figure 54.20. If the beam intensity profile is Gaussian, then the intensity incident on the detector ($I(x)$) as the knife-edge is scanned across the beam is given by

$$I(x) = \frac{I_0}{2} \left[1 + \operatorname{erf} \left(2\sqrt{\ln(2)} \left(\frac{x - x_c}{d} \right) \right) \right]$$

where I_0 is the total intensity, x_c is the position of the intensity peak, and d is the Full Width at Half Maximum (FWHM) of the Gaussian beam [14]. The Gaussian beam waist (w) can be calculated from the FWHM using the following relation:

$$w = 0.8493218 * \text{FWHM}$$

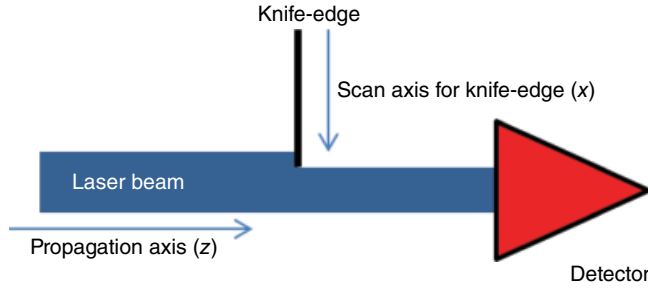


FIGURE 54.20 Knife-edge scan technique for laser beam propagating along the z -axis; the transmitted fraction of the beam is collected as a function of the knife-edge position perpendicular to the beam (x).

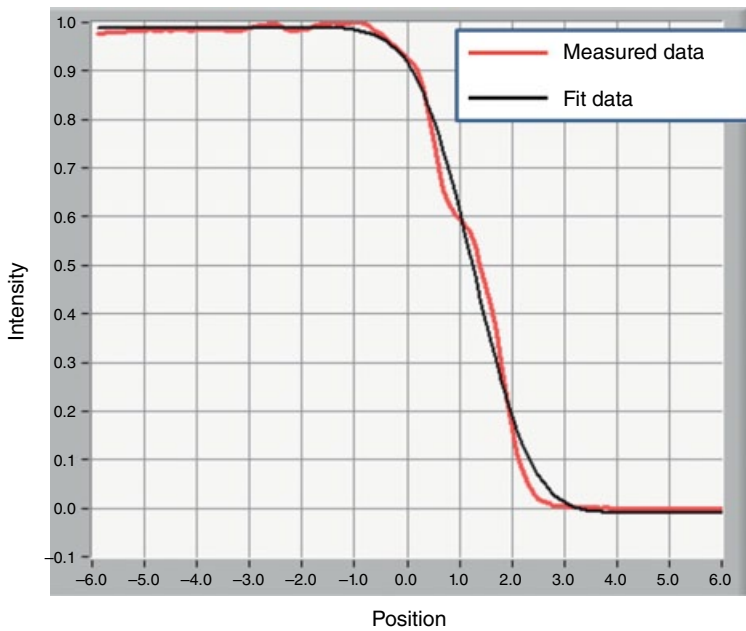


FIGURE 54.21 Knife-edge scan and best fit curve of a 1.4-mm FWHM 2.5THz beam. The black line is the best fit curve.

Figure 54.21 shows the knife-edge scan of a 2.5THz laser beam and the best fit curve. Using this one can extract the FWHM and beam waist at that point along the propagation axis. Other commonly used expressions for beam size specify the 10/90 and 20/80 points of the knife-edge scan. If x_{10} and x_{90} are the positions of 10% intensity measured and 90% intensity measured by a knife-edge measurement [15, 16], then, for a Gaussian beam profile, the waist can be calculated as

$$w \approx 0.7803(|x_{10} - x_{90}|)$$

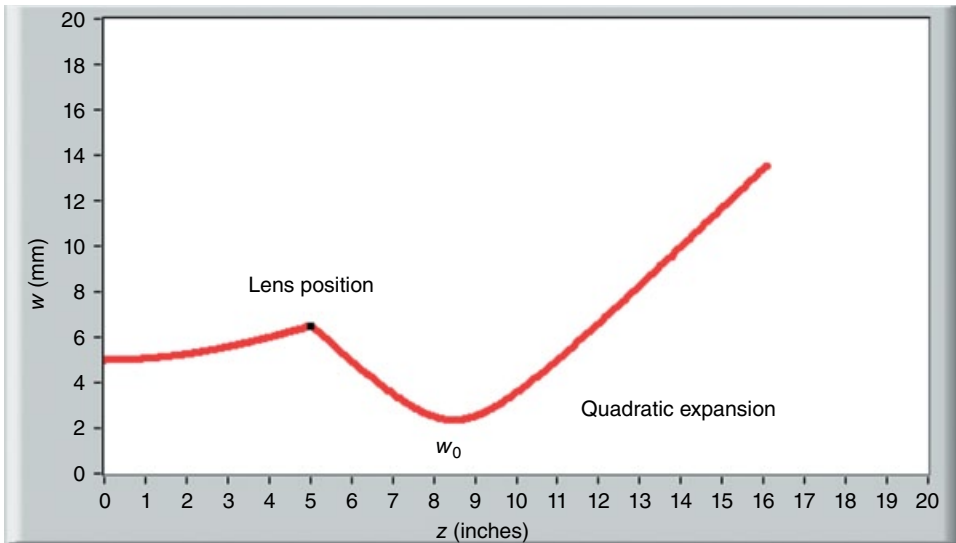


FIGURE 54.22 The quadratic expansion of a 513 μm laser beam after it is brought to focus (w_0) by a lens.

Another technique used to determine the beam waist and position of a Gaussian beam is to profile the beam expansion with an aperture. As the beam expands along the propagation axis, the waist as a function of axial distance (z) is given by

$$w^2(z) = w_0^2 \left[1 + \left(\frac{\lambda z}{\pi w_0^2} \right)^2 \right]$$

where w_0 is the beam waist at the focus (i.e., the radius of curvature at that point is infinity).

As shown in Figure 54.22, after the beam is brought to a focus, it expands quadratically in free space. Thus, using an aperture to measure the beam and fitting a Gaussian beam profile at various points along the z -axis allow one to map the quadratic expansion. Fitting the measured beam profile allows one to determine w_0 and its axial location, which is the focal plane.

54.3.1.3 Data Acquisition Data acquisition for imaging systems has two primary aspects: adequate sampling of the imaging target (scanning resolution) and the acquisition speed.

Scanning resolution essentially translates to sampling of the target with the laser beam. According to the Nyquist sampling limit, the signal must be sampled at a rate twice that of its frequency. What this implies for spatial imaging is that to generate a well-resolved image, at least two data points must be collected within the imaging beam's FWHM. This means if the beam FWHM is 1 mm, the scanning resolution

should be at most 0.5 mm in order to generate a well-resolved image. Collecting more information (higher scanning resolution) does not yield significantly more information.

Another aspect of imaging is the acquisition speed. This is determined by two primary interrelated factors: scanning speed and the detector's acquisition speed. The scanning speed determines how long it takes to move the sample across the beam or alternatively move the beam across the sample. While this is sometimes mechanically limited, as in the case of optomechanical scanning systems that direct the beam across the target area, it is ultimately limited by how fast the detector is capable of responding and the requisite signal averaging required.

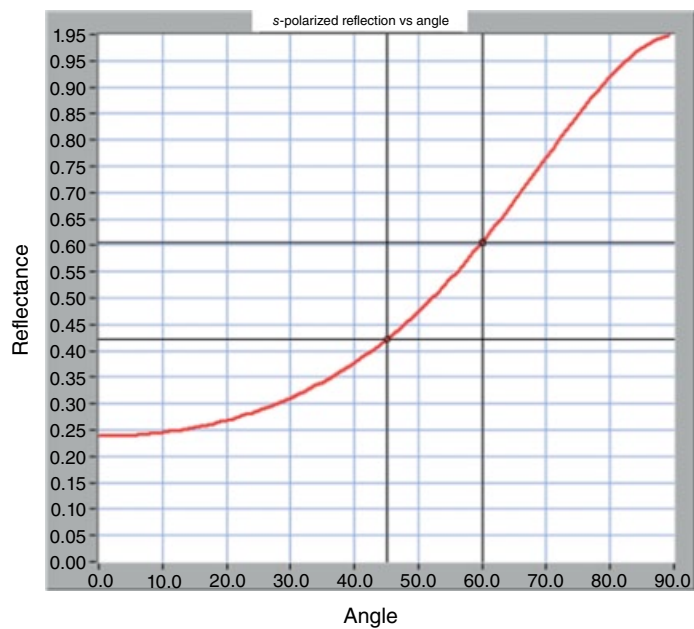
When considering the speed of the detector's response and its relation to determining system acquisition speed, it is useful to consider a specific example. Let us consider the liquid helium-cooled silicon bolometer. As discussed previously, the detector is modulated at a frequency that is related to its response time, and any higher speed modulation will not yield the requisite response. Another factor that needs to be considered is the "dwell time" per data point. The detector receives the desired signal and the background noise; since background is random, it can be eliminated by time-averaging the received signal, for example, with a lock-in amplifier averaging the signal at the detector modulation frequency. Increasing the averaging time will lower the background noise. As a rule of thumb, the dwell time for a data point collected using a lock-in amplifier should be at least 3 times the time constant of the lock-in amplifier. Thus, given a certain scanning range, scanning resolution, and dwell time, the image acquisition time can be computed.

54.3.1.4 Optical Elements Laser measurement systems consist of several optical elements that are designed to alter characteristics of the laser beam. Lenses and curved mirrors are used to reshape the laser beam. Dichroic mirrors and gratings are used to separate different frequency components. In this section we discuss two commonly used optical elements: beam splitters and wire-grid polarizers.

A commonly used beam splitter is a thin film. This works based on the Fresnel reflection and transmission coefficients for a thin film. The fraction of light that is reflected, transmitted, and absorbed by the thin film depends on the frequency-dependent refractive index of the material, the film thickness, wavelength, polarization, and the angle of incidence of the incident beam. If the incoming beam is polarized, the physical orientation of the polarizer is determined by the requisite *s*- or *p*-Fresnel coefficient.

Consider a thin Mylar film. Mylar has a complex refractive index of $1.73 + i0.030$ at $513\ \mu\text{m}$. Figure 54.23a and b shows the *s*- and *p*-Fresnel reflectance for this Mylar film ($76.2\ \mu\text{m}$ thick) as a function of incidence angle. For *s*-polarized light, the reflectance is 42% and the transmittance is 53%, while for *p*-polarized light the reflectance is 7.6% and transmittance is 87%. Thus, for the beam splitter to approximate a 50–50 splitter, it is critical that the film be oriented such that the beam polarization is perpendicular to the plane of incidence (*s*-polarized). It is important to note that the reflectance and

(a)



(b)

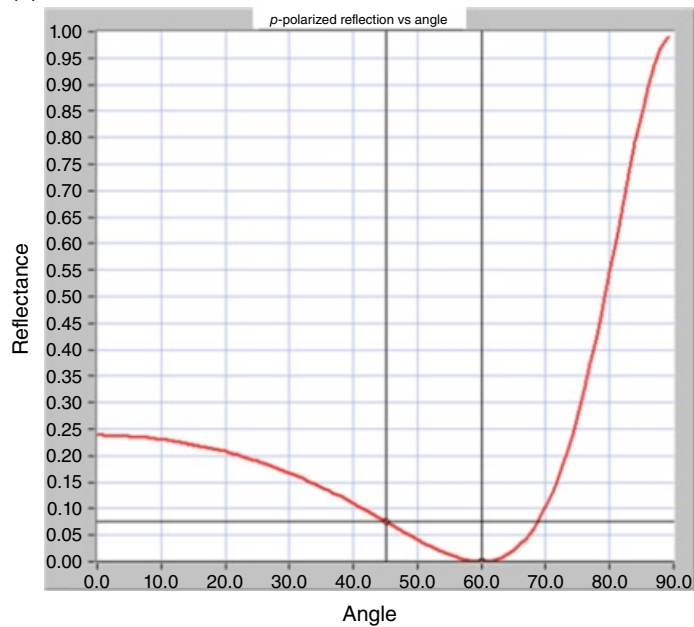


FIGURE 54.23 (a) The s -polarized Fresnel reflectance as function of incident angle and (b) the p -polarized Fresnel reflectance.

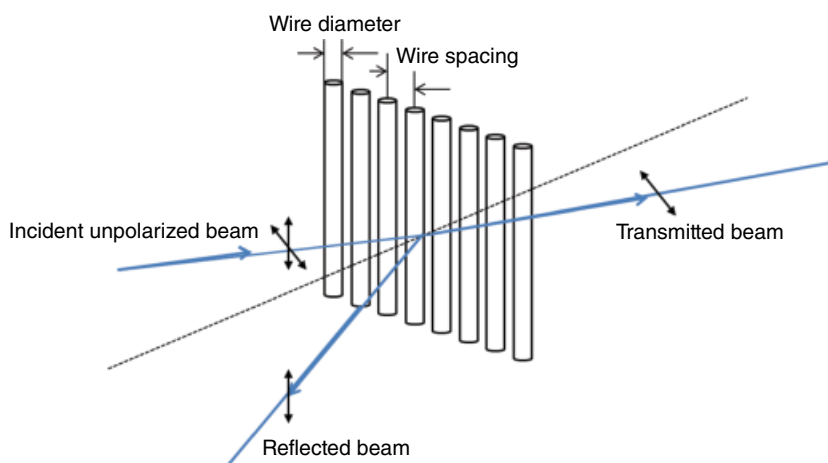


FIGURE 54.24 Reflection and transmission from a wire-grid polarizer.

transmittance do not add up to one, with the remaining signal being absorbed in the film. The signal loss due to absorption is determined by the complex part of the refractive index and film thickness. Beam splitters are typically used when normal-incidence reflectance measurements are desired. Even in the case of a lossless 50–50 beam splitter, a 100% reflective sample, and no signal loss in system optics, at most 25% of the generated laser intensity can reach the detector.

A polarizer is an optical device that selectively transmits or reflects radiation based on the polarization state of the incident light. A metallic wire-grid polarizer, such as the one shown in Figure 54.24, separates different components of linearly polarized light. The wire grid is essentially transparent to radiation that is polarized perpendicular to the orientation of the wires and strongly reflects polarization that is parallel to the wire orientation. This happens because the electric field polarized parallel to the wire grid establishes a current in the wires, which then radiates a reflected beam; however, if the electric field is perpendicular to the wires, no current is set up and the beam is transmitted. The rejection ratio is determined by the wire material, the wire diameter, the wire spacing, and the wavelength of light used. A laser beam is always polarized. The polarization can be linear or circular and quarter-wave plates can be used to convert linearly polarized light to circular polarized and vice versa.

54.3.2 Terahertz Polarized Reflection Imaging of Nonmelanoma Skin Cancers

One of the primary uses for laser imaging is in medical diagnostics. For example, optical frequency systems are used in Optical Coherence Tomography (OCT) and confocal microscopy, among others, for a variety of biomedical imaging applications [17–23]. Terahertz systems are also being developed as possible diagnostic imaging modalities [24–28].

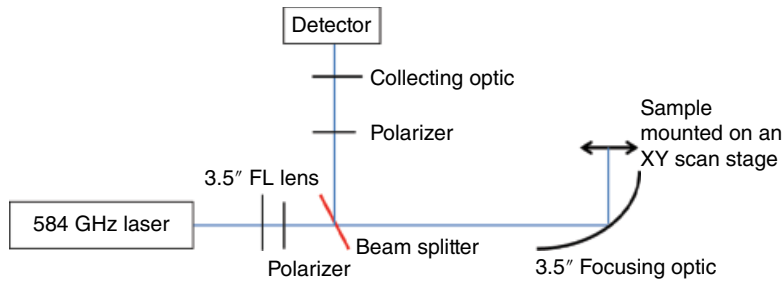


FIGURE 54.25 Schematic of a point scan terahertz reflectance system [29].

In this section we consider Continuous-Wave (CW) terahertz reflectance imaging of skin cancer. Terahertz imaging offers a way to image intrinsic contrast between healthy and diseased tissue and is nonionizing. The presented technique is an example of point scanning and uses the polarized nature of the terahertz laser beam to generate images of intrinsic contrast between normal skin and nonmelanoma skin cancer. This research was performed by Joseph et al. at UMass Lowell [29].

This investigation used a CO_2 optically pumped FIR gas laser as the source. The 584 GHz ($513\ \mu\text{m}$) vertically polarized transition in HCOOH was pumped by the 9R28 transition of the CO_2 laser. A hollow glass waveguide was used to achieve a Gaussian mode, and the measured output power was 10.23 mW. An IR Labs liquid helium-cooled silicon bolometer is used as the imaging detector. The NEP was $1.13\text{E}-13\ \text{W/Hz}^{1/2}$ and the responsivity was $2.75\text{E}+05\ \text{V/W}$. The bolometer had a response time of 5 ms and the gain was 200. A crystalline quartz garnet powdered window on the bolometer cutoff wavelengths below $100\ \mu\text{m}$.

Figure 54.25 depicts the optical configuration for the point scan imaging system. The imaging system can be summarized as follows: the laser source emits a vertically polarized beam at $513\ \mu\text{m}$. This beam is collimated by a TPX lens and then hits a fast-focusing Off-Axis Parabolic (OAP) mirror that focuses the beam down to a spot of waist size $570\ \mu\text{m}$. A sample is placed at the focal spot such that the laser beam is incident normal to the sample surface and the sample is then raster scanned across the focal spot. The acquired signal is correlated to the sample position to generate a 2D image of the sample. The reflected beam retraces the path of the incident beam. A Mylar beam splitter deflects a fraction of the reflected signal into the detection arm. Using a wire-grid polarizer in the detection arm allows one to image different polarization states of the reflected signal and generates images that are co- and cross-polarized relative to the incident beam.

The beam waist at the focal plane determines the imaging resolution. For the 584 GHz optical configuration described in this section, the beam waist was measured to be $0.57\ \text{mm}$ at the focal plane. Standard reflection imaging systems can be set up to either move the sample across the focal spot (as was done in this case) or to scan the beam across a stationary target. During the imaging procedure several factors need to

be taken into account, including scanning resolution and speed. For the experiment described, the laser beam was optically chopped, and the chopping frequency served as the reference frequency for a lock-in amplifier. The data collected by the bolometer was then sent to a lock-in amplifier that had a time constant of 30 ms. The dwell time per point in the image was around 150 ms. For the laser-based reflection imaging experiment described here, the system SNR was measured to be 65 dB.

54.3.2.1 Sample Processing and Mounting The samples imaged were fresh thick excess skin cancer specimens obtained after Mohs micrographic surgeries performed at Massachusetts General Hospital. Prior to imaging, 5 μm thick horizontal sections were cut from the sample for histopathology. These sections were stained with Hematoxylin and Eosin (H&E) and were used to evaluate the results. For imaging, the specimens were covered with a 1 mm thick z -cut quartz window. The z -cut quartz window was selected because it is relatively low loss and its refractive index 2.117 closely matches the refractive index of the human skin (≈ 2.2 at 600 GHz). Index matching allows for efficient transfer of beam power across the quartz-sample interface. To prevent dehydration during the imaging experiment, the samples were placed on a gauze soaked in pH balanced (pH 7.4) saline solution. A total of nine samples, six Basal Cell Carcinomas (BCC) and three Squamous Cell Carcinomas (SCC), were measured during this experiment.

54.3.2.2 Image Processing and Analysis Copolarized and cross-polarized images were acquired by selecting the appropriate orientation with the analyzing polarizer in the reflectance arm of the system. In order to calibrate the percent reflectance of the images, they were calibrated against the full-scale return of a flat front-surface gold mirror. Figure 54.26 shows examples of the co- and cross-polarized images (in logarithmic scale) along with the H&E-stained histopathology of the sample.

In Figure 54.26c, the tumor is outlined with the black dotted line. Easily observed, the cross-polarized image (Fig. 54.26b) correlates low terahertz cross-polarized reflectance with the tumor area, while the copolarized image (Fig. 54.26a) does not correlate well. This same observation was seen in all nine samples. This investigation demonstrates that cross-polarized terahertz reflectance imaging offers intrinsic contrast between normal and cancerous tissue. However the origin of the contrast is unclear. A possible explanation is scattering within the tissue volume. While contrast in the copolarized image is possibly obscured by Fresnel reflections from interfaces, the cross-polarized reflectance requires repolarization, which in turn could be caused by multiple scattering events within the tissue volume. At terahertz wavelengths, the primary mechanism is absorption as water is highly absorbing at these frequencies. Scattering requires structures on the order of the wavelength, which is fairly long for terahertz radiation when compared with most cellular structures; thus scattering is generally neglected when compared to absorption.

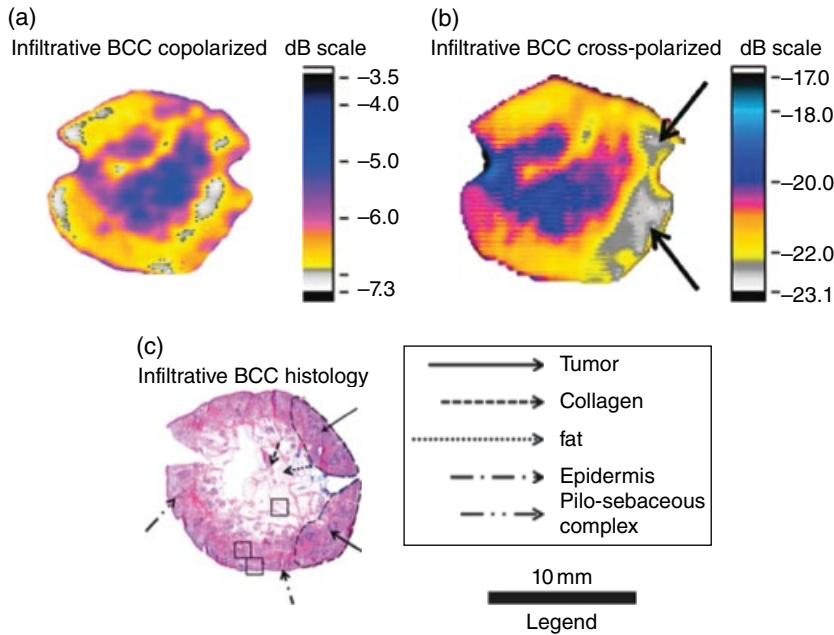


FIGURE 54.26 (a) The copolarized terahertz reflectance, (b) the cross-polarized terahertz reflectance, and (c) the H&E-stained histopathology of the corresponding 5 μm thick section of a sample of infiltrative Basal Cell Carcinoma (BCC).

The image contrast indicates that while scattering of terahertz radiation is low, it may not be negligible. The measured cross-polarized reflectance is very low ($<1\%$) as opposed to copolarized reflectance. However, it indicates that the terahertz radiation possibly undergoes scattering events within the tissue volume. Moreover, in CW-THz imaging, copolarized reflectance suffers from Fresnel artifacts at the air-window interface and at the surface of the tissue, thus obscuring contrast. Using the fact that repolarization of the backscattered beam requires several scattering events in the tissue implies that imaging cross-polarized terahertz imaging penetrates deeper into tissue and the signal collected is representative of the tissue volume.

The postulate that scattering possibly contributes to terahertz contrast is consistent with work done on terahertz dark-field imaging of dehydrated formalin-fixed tumor tissue that also showed intrinsic contrast [30]. Furthermore, work on terahertz polarization-sensitive reflection imaging of colon cancers shows contrast between normal and cancerous colon tissue as well; however in the case of colon cancer the remittance from the cancerous tissue is higher than that of normal colon [31]. The possible cause of this is that cancerous colon contains structures that are of the order of the imaging wavelength ($\approx 500\mu\text{m}$), while normal colon is very homogeneous. Further work is required to study the refractive index variation within the tissue volume at terahertz frequencies in order to determine whether scattering is an additional contrast mechanism as opposed to just water content differences.

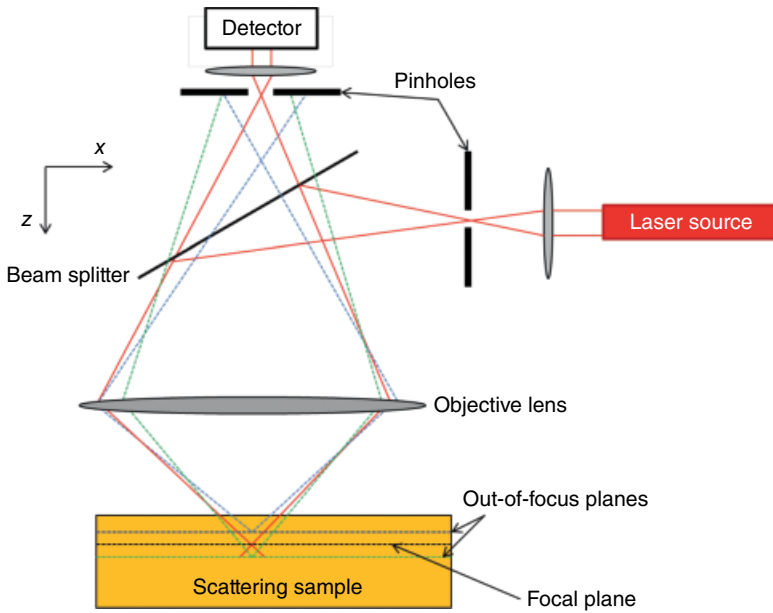


FIGURE 54.27 Schematic diagram of the confocal principle.

54.3.3 Confocal Imaging

Confocal microscopes are based on the confocal principle invented by Marvin Minsky [32]; they are able to use a pinhole to reject light that is scattered from outside the focal plane and produce high-resolution images. Confocal images are generally formed by optomechanically scanning the laser spot across the target area. The confocal principle is used to eliminate scattering from unwanted regions of the sample and thereby improve image contrast. Figure 54.27 presents the schematic of the confocal principle.

As shown in Figure 54.27, when light propagates through scattering media, signal is remitted back to the detector from the entire sample volume. This remission of light from planes that are not in focus leads to blurring and a loss of image quality and contrast. The confocal principle uses pinholes that are placed in conjugate planes to reject the unwanted reflections. As seen in Figure 54.30, the solid line traces the signal from the desired focal plane, while the dotted lines trace the scattered remission from out-of-focus planes. The out-of-focus light is rejected by the detector pinhole, the position of which is in the conjugate plane. Confocal laser scanning microscopes offer very high resolution of the order of the emission wavelength of the laser. The lateral and axial resolutions are determined by the numerical aperture of the objective lens and the emission wavelength of the laser source [21]. The depth discrimination, which is the ability to reject out-of-focus planes, is ultimately a function of the pinhole diameter; smaller pinholes better confine the focal plane. For a confocal setup, it is possible to

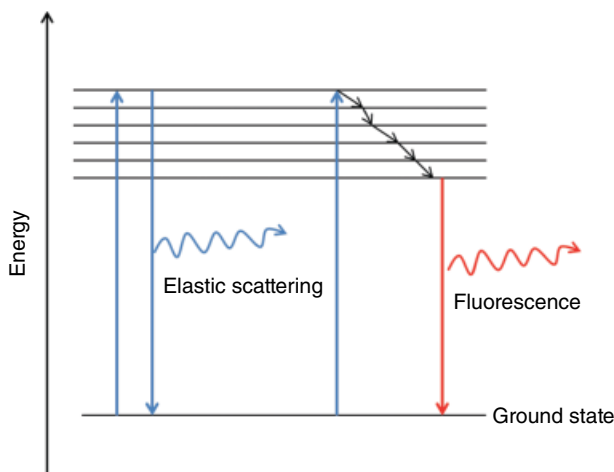


FIGURE 54.28 Absorption and possible scattering mechanisms.

image a scattering sample at varying depths by scanning the sample along the optical axis (z -axis in Fig. 54.27). This generates a sequence of imaging planes across the depth of the sample and can be used to generate a 3D map of the target. This technique is sometimes referred to as “optical sectioning.”

Confocal microscopes can be used for fluorescence and reflectance imaging of biomedical samples [33–35]. Confocal Raman microscopes use the confocal principle to generate images of the Raman scattering cross section across a target volume [36, 37]. For optical wavelengths, the interaction with biological tissue is dominated by the scattering coefficient. Figure 54.28 shows an energy-level diagram showing possible scattering mechanisms.

The absorption of photons in Figure 54.28 takes place at the excitation (or laser emission wavelength). If the photon is elastically scattered, the remitted signal is at the same wavelength as the excitation wavelength and is measured in the reflectance channel. For inelastic scattering one of the mechanisms possible is fluorescence. In this case, as shown in Figure 54.28, after the photon is absorbed, the molecule relaxes via nonradiative process to a lower energy level and then relaxes back to ground state. In this case, as the emitted photon has lower energy than the absorbed photon, it has a longer wavelength than the emission laser. This is measured in the fluorescence channel. Other inelastic scattering processes include phosphorescence and Raman scattering.

Contrast in confocal imaging can be either intrinsic or extrinsic. Intrinsic contrast measures signal remitted from the sample constituents, and for several biomedical applications at optical frequencies, this contrast is limited. Extrinsic contrast is generated by adding a dye or contrast agent that binds to specific aspects of the sample. The fluorescence or reflectance of these dyes is then used to generate high contrast images [33, 34].

54.3.4 Optical Coherence Tomography

54.3.4.1 Introduction Technologies such as OCT have allowed researchers to analyze the 3D structures of various materials. OCT works in the optical to near-infrared range and therefore provides high axial and lateral resolution. Depending upon how the signal is measured, OCT can broadly be divided in two types: time-domain OCT and spectral-domain OCT. OCT has been an established technique for medical imaging. It is commonly used for eye and retinal imaging [38]. Some recent research shows that it can be a useful tool to monitor the progression of glaucoma [39]. It has also been used for the imaging of coronary arteries [40]. As this technique uses optical or near-infrared range of frequencies, it is limited by scattering in biological tissue samples.

Broadband diode lasers are used as a light source in a typical OCT system. For example, a diode laser centered at 800 nm with a bandwidth of 50 nm is used for retinal imaging [38].

54.3.4.2 Time-Domain OCT A typical OCT system uses broadband diode lasers that have low coherence length. This property has been utilized to investigate the axial properties of the sample. The schematic diagram of time-domain coherence tomography using broadband sources is shown in Figure 54.29a.

The optical configuration is simply a Michelson interferometer. The broadband input beam is split between two arms using a beam splitter. One arm acts as a reference with signal reflected from a mirror, while the other acts as a sample arm in which the signal is reflected from the sample under investigation. The combined

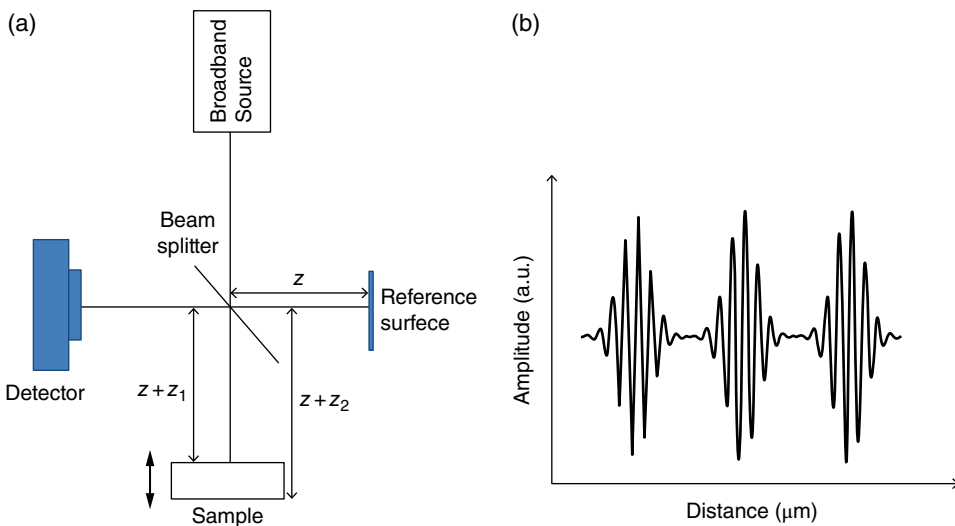


FIGURE 54.29 (a) Time-domain OCT system. (b) Interference pattern of the signal reflected from three different layers of the sample with respect to signal reflected from reference mirror surface.

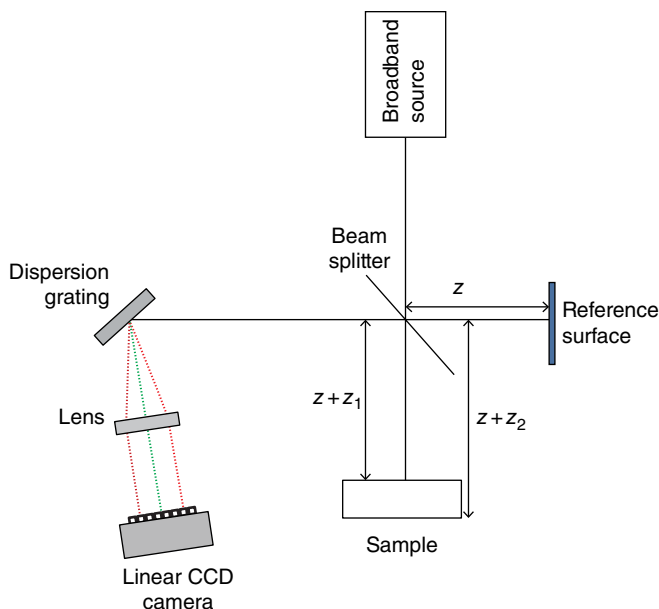


FIGURE 54.30 Schematic of spectral-domain interferometry.

reflected signal from the reference mirror and sample is measured at the detector. The signals reflected from different layers of the sample are delayed in time with respect to each other and also with respect to the reference mirror. In order to obtain the axial profile, the reference mirror is scanned axially. A sinusoidal signal is obtained at the detector when the optical path length of the reference mirror matches with a certain layer of the sample. The envelope of the interference pattern gives the axial profile of the sample. For example, if the sample shown in Figure 54.29a has three layers, it yields the axial profile shown in Figure 54.29b. The 3D profile of the sample is obtained by either raster scanning the probe beam or the sample itself and at the same time scanning the reference mirror to profile the sample axially. The speed of the scanning stage poses a limitation on how fast one can acquire the sample information.

54.3.4.3 Spectral-Domain OCT The spectral-domain low coherence interferometry technique is based on a Michelson interferometer, where a broadband, low-coherence source is used to illuminate a reference surface and a sample as shown in Figure 54.30. Consider a sample composed of two reflecting layers: one at an Optical Path Difference (OPD) of Z_1 and another at an OPD of Z_2 from the reference surface. The reflected light from the two illuminated objects is combined onto a dispersion grating that angularly separates the different wavelength components, which form an interference pattern on a linear CCD camera using a lens.

At the image plane, the phase difference between the signal reflected from the reference surface and the first layer of the sample is given by

$$\phi_1(k) = \phi_0 + \frac{4\pi}{\lambda} z_1$$

where λ is the wavelength, ϕ_0 is the phase change introduced by reflection at the first layer, and z_1 is the OPD between the reference surface and the first layer.

The phase difference can be rewritten in terms of the wavenumber, $k = 2\pi/\lambda$, such that

$$\phi_1(k) = \phi_0 + 2kz_1$$

which is a linear relationship between the phase and the wavenumber. As the frequency along the k -axis is given by the rate of change of the phase with respect to the wavenumber, this leads to

$$f_{k1} = \frac{1}{2\pi} \cdot \frac{\partial \phi_1(k)}{\partial k} = \frac{z_1}{\pi}$$

Similarly for the second reflecting layer in the sample, the frequency of the phase is given by

$$f_{k2} = \frac{1}{2\pi} \cdot \frac{\partial \phi_2(k)}{\partial k} = \frac{z_2}{\pi}$$

One should note that analysis in the k -space is preferred to λ -space since f_k is independent of k , whereas the equivalent frequency in λ -space would vary with λ . Thus, sampling the interferogram intensity data uniformly along the k -axis would cause a broadening in the frequency spectrum, which reduces the axial depth resolution of the system.

The intensity distribution along the k -axis on the linear CCD camera as a result of the interference between the signal reflected from the reference surface and the signal reflected from the two layers of the sample can be expressed as

$$I(k) = I_0(k) + 2\sqrt{I_r(k)I_1(k)}\cos(\phi_1(k)) + 2\sqrt{I_r(k)I_2(k)}\cos(\phi_2(k))$$

where $I_0(k)$ is a DC term; $I_r(k)$, $I_1(k)$, and $I_2(k)$ are the intensity of the signals coming from the reference surface, the first layer, and the second layer, respectively; and $\phi_1(k)$ and $\phi_2(k)$ are the phase differences between the signal from the reference surface and the first layer and the reference surface and the second layer, respectively. Because of the presence of the cosine term and the fact that the phase difference is dependent on k , a modulation in the spectrum intensity along the k -axis is introduced.

The modulation is the result of two signals with frequencies given by $f_{k1} = z_1/\pi$ and $f_{k2} = z_2/\pi$, which are directly proportional to z_1 and z_2 , respectively. One should note that if there were N_i layers in the sample, the signal from each layer would interfere with the reference signal and produce a modulation in the spectrum intensity along the k -axis whose frequencies would be proportional to the OPD between the reference surface and all N_i corresponding reflecting layers. The FT of the intensity pattern at the camera, after proper calibration from pixel to k -space, directly leads to the frequencies of the modulation and thus the position of the reflecting layers within the sample with respect to the reference surface. Such frequency spectrum obtained after the FT of the spectrum intensity is commonly called axial scan (A-Scan), or depth scan. In this technique, the maximum measurable OPD is limited by the depth range of the system, the details of which are discussed further.

In spectral-domain low coherence interferometry, the available spectral bandwidth of the light source is spread over the limited number of CCD pixels. If the total bandwidth of the interfering signal acquired with N_p number of pixels of the CCD camera is Δk , then the interval along the distance axis or the distance per pixel (Δd) after the FT is given by $\Delta d = \frac{1}{2} \frac{2\pi}{\Delta k}$. The 1/2 factor in this expression accounts for the doubling of the OPD after reflection. Thus the maximum OPD that can be measured as a function of N_p is given by

$$d_{\max} = \frac{1}{2} \frac{N_p}{2} \Delta d = \frac{1}{2} \frac{N_p}{2} \frac{2\pi}{\Delta k} = \frac{1}{2} \frac{N_p}{2} \frac{\lambda_0^2}{\Delta \lambda}$$

A factor of 2 appears in the denominator of the above equation, since the signal after an FT is symmetric around zero OPD. Thus, the signals on the opposite sides of zero give the same information. For a Gaussian profiled spectrum, the depth range can be written as

$$d_{\max} = \frac{1}{2} \frac{2 \ln 2}{\pi} \frac{N_p}{2} \frac{\lambda_0^2}{\Delta \lambda}$$

A typical example of the frequency-domain interferometric signal at different depths for a given depth range is shown in Figure 54.31.

54.3.5 Femtosecond Laser Imaging

The branch of ultrashort laser imaging and spectroscopy is used in a wide range of areas, both scientific and industrial. The key point in ultrashort laser pulses is their time- and frequency-domain properties. In the time domain, the output consists of high-intensity pulses on the order of femtoseconds. In the frequency domain, the pulse train produced by a mode-locked laser consists of a broad spectrum of equidistant

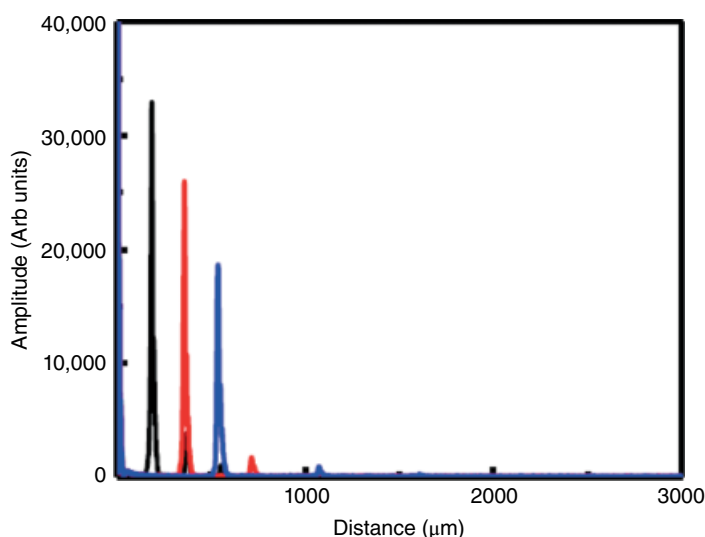


FIGURE 54.31 Change in the FFT signal with change in the optical path difference between sample and reference surface.

modes with a defined phase relationship. Ultrashort pulse lasers make it possible to probe samples on a femtosecond time scale, allowing a number of ultrafast chemical, biological, and physical processes [41–43].

Today, the titanium–sapphire ($\text{Ti}:\text{Al}_2\text{O}_3$) laser is the most widely used commercially available tunable laser. Like a ruby laser, to make Ti:sapphire, Ti_2O_3 is doped into a crystal of Al_2O_3 , and Ti^{3+} ions then occupy some of the Al^{3+} -ion sites in the lattice. The typical concentrations range between 0.1 and 0.5% by weight. The laser is based on a four-level energy scheme. CW Ti:sapphire lasers are pumped by the green output of an argon laser. In pulsed operation, frequency-doubled Nd:YAG or Nd:YLF lasers as well as flashlamps are used. The tuning curve of this laser in a CW regime spans the wavelength range of over 400 nm between 670 and 1050 nm. This laser possesses a favorable combination of properties that are up to now the best among all known broadband laser materials. First, the active medium is solid state, which means long operational time and laser compactness. Second, sapphire has high thermal conductivity, exceptional chemical inertness, and mechanical resistance. Third, it has a very broad generated spectrum. The combination of all these properties made the Ti:sapphire crystal the most popular laser medium in the industry.

After selecting an appropriate lasing medium the next step is to generate ultrashort pulses. The following two techniques are used for generating short pulses:

1. Q-switching
2. Mode-locking

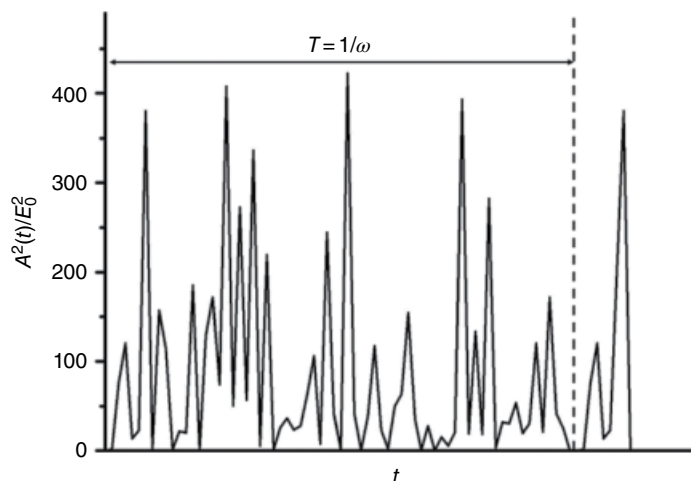


FIGURE 54.32 Squared amplitude of random oscillating modes in a cavity as a function of time.

54.3.5.1 Laser Q-Switching As been discussed, under CW operation, the population inversion reaches to its threshold value when oscillation starts. If a shutter is introduced in the cavity; when the shutter is closed, the population inversion can exceed a value far more compared to value when the shutter is open. When the shutter is closed, there will be gain in the cavity that greatly exceeds losses, and when the shutter is opened suddenly, the stored energy will be released in the form of short and intense light pulses. As this operation involves switching the Q-factor of the cavity from a low to high value, this technique is called Q-switching. This technique allows for the generation of laser pulses of the duration of photon decay times (few tens of nanoseconds) and power of the order of megawatts.

Several techniques have been applied to switch the cavity Q [44]. Broadly the devices can be grouped into active and passive Q-switches. Active Q-switching involves some kind of external operation to actively control the switching mechanism, for example, electro-optical Q-switching, where an external voltage is applied to switch the cavity Q-value. In passive Q-switching the switching takes place automatically using nonlinearity of some medium. An example of a passive Q-switch is the saturable absorber, which has low value of saturation intensity. When the laser power is low, the material does not allow the light to transmit through it; this means the switch is closed. When the laser power exceeds the saturation threshold of the medium, it becomes transparent, which in turn opens the switch.

54.3.5.2 Laser Mode-Locking In a laser cavity, many longitudinal modes can oscillate at a frequency given by $\omega_m = m2\pi \frac{c}{2L}$. These modes exhibit no phase relationship and oscillate independent of each other as shown in Figure 54.32. It is possible to make these random modes oscillate with a definite phase relation within the cavity; the process used is known as mode-locking and such lasers are referred to

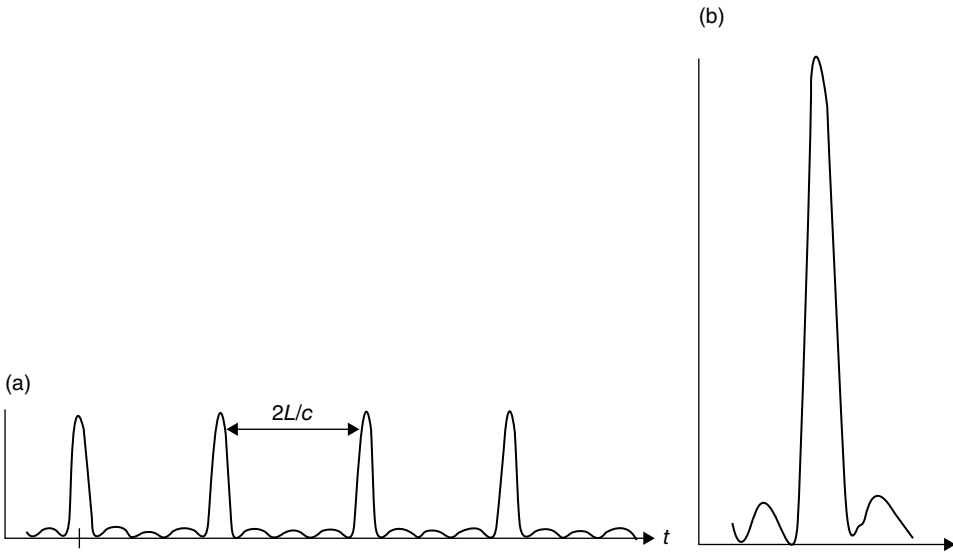


FIGURE 54.33 (a) Phase-locked modes propagating in the cavity. (b) Coherent sum of these pulses results in short but much higher-amplitude pulse.

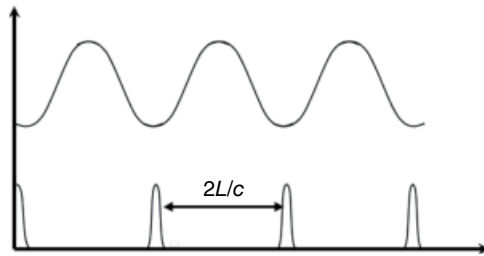


FIGURE 54.34 Cavity round-trip losses for amplitude-modulated mode-locking.

as mode-locked lasers. These locked modes can be considered as a Fourier series expansion of a periodic function in time, given by $T = 1/\omega = 2L/m2\pi c$, in which case they constitute a periodic pulse train, as shown in Figure 54.33a. The coherent sum of all the locked phases would interfere constructively, and a pulse with high peak power is obtained as shown in Figure 54.33b.

Like Q-switching the mode-locking mechanism can also be classified as active and passive. In this section, only active mode-locking (amplitude modulation) is discussed, and other active and passive mode-locking techniques are discussed in detail in reference [44]. In the case of amplitude modulation mode-locking, the modulator is placed at one end of cavity.

Figure 54.34 shows cavity round-trip losses, which are modulated by a time interval $T = 2L/c$. If the frequency of the modulation (ω_m) is equal to the frequency difference between two consecutive modes ($\Delta\omega$), the light pulses will pass through the modulator at the time of minimum loss. This is the steady-state condition: as if a light pulse passes

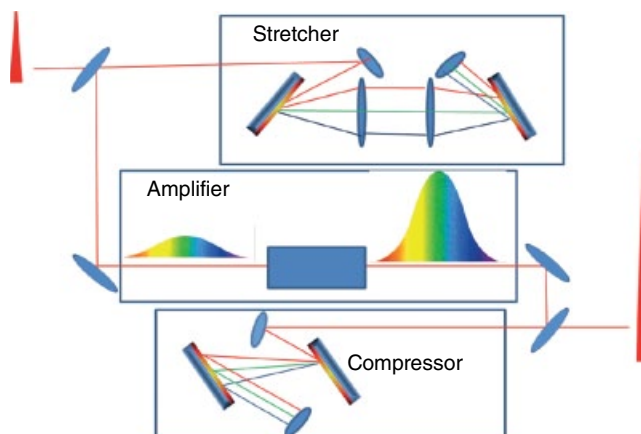


FIGURE 54.35 Chirped pulse amplification of a femtosecond laser pulse.

through the modulator at a time of minimum loss, which is the round-trip time of the cavity as well ($2L/c$), it will return to the modulator at its minimum loss as well. The time-varying losses of the modulator will damp the pulses reaching before and after the minimum loss time, thus mode-locking the laser by forcing a definite phase relationship between the propagating modes. In this case after each pass the pulse duration gets shorter, as the leading and trailing edge of the pulse attenuates in each pass. The shortening of the pulse is eventually limited by bandwidth of the gain medium. As the pulse width becomes shorter, the spectrum becomes large and at one point would fill the bandwidth of the laser medium. In this situation, the wings of the spectrum would no longer be amplified due to losses of the modulator, and this would limit the pulse duration of the laser.

54.3.5.3 Chirped Pulse Amplification Using mode-locking it is possible to generate laser pulses less than 10 fs, producing a peak power on the order of megawatts. Further amplification of the laser pulse is limited by nonlinear processes due to the high power and intensity of the laser pulse in the gain medium. To amplify the pulse further a technique adopted from radar technology is used, where a few tens of femtosecond short pulse is first stretched in time to several picoseconds, then amplified, and then finally compressed back to the ultrashort femtosecond pulse. The stretching reduces the intensity of the laser beam by 3–4 orders of magnitude and thus makes further amplification of the pulse possible. A typical Chirped Pulse Amplification (CPA)-based Ti:sapphire laser amplifier system is shown in Figure 54.35, where a grating pair is used as a stretcher and another pair as compressor.

54.3.5.4 Two-Photon Fluorescence Microscopy One application of high-power lasers for imaging is the two-photon fluorescence microscopy. In typical single-photon fluorescence microscopy, a molecule in its ground state will absorb a photon and is

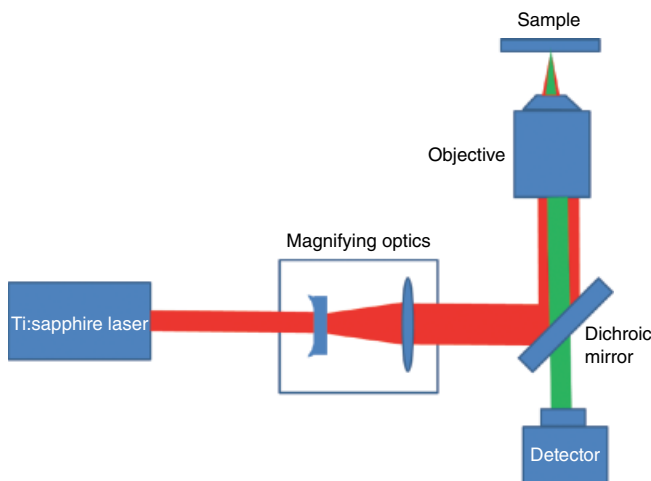


FIGURE 54.36 Schematic of two-photon fluorescence microscopy imaging setup.

excited to a higher energy state if the energy of the photon is equal to or higher than the energy difference between the two states. After relaxing to a lower vibrational state, the molecule will return to its electronic ground state, emitting a lower-energy photon (fluorescence) compared to the absorbed one (Fig. 54.28). The same process can take place if the sum of the energies of two photons is enough to reach the first excited state of the molecule. In this case, the probability of simultaneous absorption of two photons depends on the square of the intensity of the exciting beam.

The experimental design for a two-photon microscopy system is shown in Figure 54.36. In the setup, the femtosecond laser pulse from Ti:sapphire is magnified using negative and positive lens. The expanded beam is reflected using a dichroic mirror, which reflects infrared and transmits visible light. The reflected beam is focused using an objective lens to the diffraction limit at the sample position. At the focus, due to the high intensity of the beam, two-photon absorption will take place. Visible light, for example, green, will be generated due to fluorescence. The dichotic mirror will transmit the visible light to the detector.

Two-photon microscopy offers several advantages over other conventional techniques like confocal microscopy. The two-photon wavelength is twice as high as the wavelength of single photon; this wide difference between the excitation and emission wavelength makes sure that the excitation light and other scattering can be filtered out from the fluorescence signal. Secondly, two-photon microscopy is well suited for optically thick samples, as infrared radiation used in the two-photon excitation has greater penetration and less absorption/scattering in biological samples compared to green, blue, or ultraviolet light. The use of pinhole in confocal microscopy limits the number of photons reaching the detector, as the scattering of fluorescence deviates the path of the reflected photons. In two-photon microscopy the signal reaching at the detector is much higher as the deviated light will still reach the detector. Several

biomedical applications of two-photon fluorescence microscopy are currently under investigation [45].

54.3.6 Laser Raman Spectroscopy

Given the myriad of spectroscopic measurement systems developed, Raman spectroscopy exemplifies one of the methodologies engineered for performing chemical analysis. Pioneering the technique, C.V. Raman recognized that an energy shift occurred for photons scattered off molecules. He realized that the inelastic process of scattering light caused vibrational and/or rotational transitions of the molecules that shifted the scattered photons to lower or higher energy levels.

If the molecule transitions to a higher energy level when illuminated, then the photon scatters with a lower energy level exhibiting a longer wavelength referred to as a Stokes shift. Likewise, if the molecular transition is to a lower level, the scattered photon carries away the excess energy and thus has a higher energy and shorter wavelength than the incident light referred to as an anti-Stokes shift. The Raman shift is expressed in units of wavenumber as

$$\Delta w(\text{cm}^{-1}) = \left(\frac{1}{\lambda_i} - \frac{1}{\lambda_s} \right)$$

where λ_i incident wavelength and λ_s is the scattered wavelength. Raman received the 1930 Nobel Prize in Physics for his discovery.

Since the vibrational and/or rotational transitions of a molecular structure are uniquely driven by the chemical bonds, Raman spectroscopy can be used to identify chemical species of molecules. However the Raman scattering is weak. A variety of techniques have been developed to improve sensitivity and resolution. Taking advantage of the phenomena, the ultraviolet wavelength pulses of an excimer laser can be passed through a gas cavity to generate wavelengths as short as 35.5 nm, the seventh harmonic of KrF, enabling spectrometry in the extreme ultraviolet. Raman interactions in molecular gases such as hydrogen produce smaller frequency shifts when a tightly focused excimer laser beam is passed through the hydrogen gas cell. Though the Raman shift can produce longer or shorter wavelengths, generally longer wavelength light is generated more efficiently.

Given the commercial availability of Raman frequency shifters, Raman spectroscopy can be readily used to observe vibrational, rotational, and other low-frequency modes in molecules. With both pulsed and CW lasers generating sufficient power densities for driving the harmonic generation technique, the inelastic scattering of monochromatic laser light in the ultraviolet, visible, and near-infrared regime yields information about the vibrational modes of molecules and therefore the molecular structure of materials. Raman spectroscopy is particularly useful for mineral identification because chemical species exhibit unique spectral characteristics and can

identify molecular structure by chemical bonds [46]. Techniques such as Resonance Raman Scattering (RRS), Surface-Enhanced Raman Scattering (SERS), and coherent anti-Stokes Raman spectroscopy have been developed to observe low Raman Scattering cross sections [23].

54.3.7 Suggested Reading

Several reference books cover different imaging and spectroscopic systems introduced in this section. *Terahertz Techniques* by Bründermann, Hübers and Kimmitt, published by Springer, provides an excellent overview of terahertz frequency measurement systems and covers various aspects of imaging and detector technology. *Handbook of Photonics for Biomedical Science*, edited by V. Tuchin and published by CRC Press, covers a wide range of imaging systems used in biomedical applications including OCT, terahertz imaging, confocal microscopy, and Raman spectroscopy. *Handbook of Biological Confocal Microscopy* by J. B. Pawley, published by Springer, also provides an excellent textbook for confocal microscopy.

REFERENCES

1. Siegman, A. E., “*Lasers*,” University Science Books, Mill Valley, CA, 1986.
2. Kogelnik, H. and Li, T., “Laser beams and resonators,” *Applied Optics*, 5(10), 1550–1567, 1966.
3. Guenther, R. D., “*Modern Optics*,” John Wiley & Sons, Inc., New York, 1990.
4. Weber, M. J., “*CRC Handbook of Laser Science and Technology, Volume II Gas Lasers*,” CRC Press, Inc., Boca Raton, FL, (1982).
5. Pozar, D. M., “*Microwave Engineering*,” John Wiley & Sons, Inc., Hoboken, NJ, 2005.
6. Danylov, A., “Frequency stabilization, tuning, and spatial mode control of terahertz quantum cascade lasers for coherent transceiver applications,” Ph.D. dissertation, Department of Physics and Applied Physics, University of Massachusetts Lowell, USA, 2010.
7. Faist, J., Capasso, F., Sivco, D. L., Sirtori, C., Hutchinson, A. L., and Cho A. Y., “Quantum cascade laser,” *Science*, 264(5158), 553–556, 1994.
8. Danylov, A. A., Waldman, J., Goyette, T. M., Gatesman, A. J., Giles, R. H., Linden, K. J., Neal, W. R., Nixon, W. E., Wanke, M. C., and Reno, J. L., “Transformation of the multi-mode terahertz quantum cascade laser beam into a Gaussian, using a hollow dielectric waveguide,” *Applied Optics*, 46(22), 5051–5055, 2007.
9. Waldman, J., Fetterman, H. R., Goodhue, W. D., Bryant, T. G., and Temme, D. H., “Submillimeter modeling of millimeter radar systems,” *Proceedings of SPIE*, 259 Millimeter Optics, 152–157, 1980.
10. Chen, T. J., Chu, T. H., and Chen, C., “A new calibration algorithm of wideband polarimetric measurement system,” *IEEE Transactions of Antennas and Propagation*, 39(8), 1188–1192, August 1991.

11. DeMartinis, G. B., Coulombe, M. J., Horgan, T. M., Giles, R. H., and Nixon, W. E., "A 240 GHz Polarimetric Compact Range for Scale Model RCS Measurements," Antenna Measurements Techniques Association (AMTA), Atlanta, GA, pp. 3–8, October 2010.
12. Goyette, T. M., Dickinson, J. C., Waldman, J., and Nixon, W. E., "1.56-THz compact radar range for W-band imagery of scale-model tactical targets," *Proceedings of SPIE*, 4053, 615–622, Algorithms for Synthetic Aperture Radar Imagery VII; Edmund G. Zelnio; Ed. August 2000.
13. Goyette, T. M., Dickinson, J. C., Waldman, J., and Nixon, W. E., "Three dimensional fully polarimetric W-band ISAR imagery of scale-model tactical targets using a 1.56THz compact range," *Proceedings of SPIE*, 5095, 66–74, Algorithms for Synthetic Aperture Radar Imagery X; Edmund G. Zelnio, Frederick D. Garber; Ed. September 2003.
14. Brundermann, E., Hubers H.-W., and Kimmit, M. F., "*Terahertz Techniques*," Springer Series in Optical Sciences, Springer, Berlin/London, 2012.
15. Suzaki, Y. and Tachibana, A., "Measurement of the μm sized radius of Gaussian laser beam using the scanning knife-edge," *Applied Optics*, 14(12), 2809–2810, 1975.
16. Khosroffian, J. M. and Garetz, B. A., "Measurement of a Gaussian laser beam diameter through the direct inversion of knife-edge data," *Applied Optics*, 22(21), 3406–3410, 1983.
17. Huang, D., Swanson, E. A., Lin, C. P., Schuman, J. S., Stinson, W. G., Chang, W., Hee, M. R., Flotte, T., Gregory, K., Puliafito, C. A., and Fujimoto, J. G., "Optical coherence tomography," *Science*, 254, 1178–1181, 1991.
18. Schmitt, J. M., "Optical coherence tomography (OCT): a review," *IEEE Journal of Selected Topics in Quantum Electronics*, 5(4), 1205–1215, July/August 1999.
19. Wojtkowski, M., Leitgeb, R., Kowalczyk, A., Bajraszewski, T., and Fercher, A. F., "In vivo human retinal imaging by Fourier domain optical coherence tomography," *Journal of Biomedical Optics*, 7(3), 457–463, 2002.
20. Paddock, S. W., "Principles and practices of laser scanning confocal microscopy," *Molecular Biotechnology*, 16(2), 127–149, 2000.
21. Pawley, J. B., "*Handbook of Biological Confocal Microscopy*," 3rd ed., Springer Science + Business Media, New York, 2006.
22. Rajadhyaksha, M., Grossman, M., Esterowitz, D., Webb, R. H., and Anderson, R. R., "In vivo confocal scanning laser microscopy of human skin: melanin provides strong contrast," *Journal of Investigative Dermatology*, 104(6), 946–952, 1995.
23. Tuchin, V. V. (Editor), "*Handbook of Photonics for Biomedical Science*," CRC Press, Taylor & Francis Group, Boca Raton, FL, 2010.
24. Pickwell, E. and Wallace, V. P., "Biomedical applications of terahertz technology," *Journal of Physics D: Applied Physics*, 39(17), R301, 2006.
25. Kim, S. M., Baughman, W., Wilbert, D. S., Butler, L., Bolus, M., Balci, S., and Kung, P., "High sensitivity and high selectivity terahertz biomedical imaging," *Chinese Optics Letters*, 9(11), 110009, 2011.
26. Oh, S. J., Choi, J., Maeng, I., Park, J. Y., Lee, K., Huh, Y. M., Suh, J., Haam, S., and Son, J. H., "Molecular imaging with terahertz waves," *Optics Express*, 19(5), 4009–4016, 2011.

27. Arbab, M. H., Dickey, T. C., Winebrenner, D. P., Chen, A., Klein, M. B., and Mourad, P. D., "Terahertz reflectometry of burn wounds in a rat model," *Biomedical Optics Express*, 2(8), 2339–2347, 2011.
28. Son, J. H. (Editor), "*Terahertz Biomedical Science and Technology*," CRC Press, Taylor & Francis Group, Boca Raton, FL, 2014.
29. Joseph, C. S., Patel, R., Neel, V. A., Giles, R. H., and Yaroslavsky, A. N., "Imaging of ex vivo nonmelanoma skin cancers in the optical and terahertz spectral regions. Optical and terahertz skin cancers imaging," *Journal of Biophotonics*, 7, 295–303, 2014, published online 2012.
30. Löffler, T., Bauer, T., Siebert, K., Roskos, H., Fitzgerald, A., and Czasch, S., "Terahertz dark-field imaging of biomedical tissue," *Optics Express*, 9, 616–621, 2001.
31. Doradla, P., Alavi, K., Joseph, C., and Giles, R., "Detection of colon cancer by continuous-wave terahertz polarization imaging technique," *Journal of Biomedical Optics*, 18(9), 090504–090504, 2013.
32. Minsky, M., "Memoir on inventing the confocal scanning microscope," *Scanning*, 10(4), 128–138, 1988.
33. Wirth, D., Snuderl, M., Sheth, S., Kwon, C. S., Frosch, M. P., Curry, W., and Yaroslavsky, A. N., "Identifying brain neoplasms using dye-enhanced multimodal confocal imaging," *Journal of Biomedical Optics*, 17(2), 0260121–0260127, 2012.
34. Snuderl, M., Wirth, D., Sheth, S. A., Bourne, S. K., Kwon, C. S., Ancukiewicz, M., Curry, W. T., Frosch, M. P., and Yaroslavsky, A. N., "Dye-enhanced multimodal confocal imaging as a novel approach to intraoperative diagnosis of brain tumors," *Brain Pathology*, 23(1), 73–81, 2013.
35. Al-Arashi, M. Y., Salomatina, E., and Yaroslavsky, A. N., "Multimodal confocal microscopy for diagnosing nonmelanoma skin cancers," *Lasers in Surgery and Medicine*, 39(9), 696–705, 2007.
36. Caspers, P. J., Lucassen, G. W., Carter, E. A., Bruining, H. A., and Puppels, G. J., "In vivo confocal Raman microspectroscopy of the skin: noninvasive determination of molecular concentration profiles," *Journal of Investigative Dermatology*, 116(3), 434–442, 2001.
37. Caspers, P. J., Lucassen, G. W., and Puppels, G. J., "Combined in vivo confocal Raman spectroscopy and confocal microscopy of human skin," *Biophysical Journal*, 85(1), 572–580, 2003.
38. Swanson, E., Izatt, J., Lin, C., Fujimoto, J., Schuman, J., Hee, M., Huang, D., and Puliafito, C., "In vivo retinal imaging by optical coherence tomography," *Optics Letters*, 18, 1864–1866, 1993.
39. Schuman, J. S., Hee, M. R., Puliafito, C. A., Wong, C., Pedut-Kloizman, T., Lin, C. P., Hertzmark, E., Izatt, J. A., Swanson, E. A., and Fujimoto, J. G., "Quantification of nerve fiber layer thickness in normal and glaucomatous eyes using optical coherence tomography: a pilot study," *Archives of Ophthalmology*, 113, 586–596, 1995.
40. Jang, I.-K., Tearney, G. J., MacNeill, B., Takano, M., Moselewski, F., Iftima, N., Shishkov, M., Houser, S., Aretz, H. T., Halpern, E. F., and Bouma, B. E., "In vivo characterization of coronary atherosclerotic plaque by use of optical coherence tomography," *Vascular Medicine*, 111, 1551–1555, 2005.

41. Stolow, A., Bragg, A. E., and Neumark, D. M., "Femtosecond time-resolved photoelectron spectroscopy," *Chemical Reviews*, 104, 1719–1758, 2004.
42. Kukura, P., McCamant, D. W., and Mathies, R. A., "Femtosecond stimulated Raman spectroscopy," *Annual Review of Physical Chemistry*, 58, 461–488, 2007.
43. Holzwart, A. R., "Applications of ultrafast laser spectroscopy for the study of biological systems," *Quarterly Reviews of Biophysics*, 22, 239–326, 1989.
44. Svalto, O., "*Principles of Lasers*," 5th ed., Springer Science + Business Media, LLC, New York, 2010.
45. So, P. T., Dong, C. Y., Masters, B. R., and Berland, K. M., "Two-photon excitation fluorescence microscopy," *Annual Review of Biomedical Engineering*, 2(1), 399–429, 2000.
46. Chen, H. and Stimets, R. W., "Fluorescence of trivalent neodymium in various materials excited by a 785 nm laser," *American Mineralogist*, 99(2–3), 332–342, 2014.

MAGNETIC FORCE IMAGES USING CAPACITIVE COUPLING EFFECT

BYUNG I. KIM

Department of Physics, Boise State University, Boise, ID, USA

55.1 INTRODUCTION

There are several conventional approaches for mapping magnetic field distributions. Optical techniques based on the Kerr effect have moderate spatial resolution of about $0.5\mu\text{m}$. Bitter pattern technique causes degradation of the sample surface. Although electron beam imaging techniques like Lorentz microscopy, scanning electron microscope with polarization analysis (SEMPA), and differential phase contrast (DPC) STEM are known to have higher spatial resolution, sample preparation and operation are difficult. But magnetic force microscopy (MFM) needs no special sample preparation and provides different and somewhat complementary information to e-beam imaging techniques as well as high spatial resolution (10–100 nm).

Owing to these merits, MFM has been applied to the study of recording media such as longitudinal media and magnetic multilayer film [1]. As the recording density is increased, it is important to understand the submicron magnetic structure behavior. In this respect, MFM could be a promising tool. Furthermore, MFM has been successfully applied to the study of fundamental science: For example, the macroscopic quantum tunneling (MQT) [2], the biological magnetism [3], the domain-structure change with the film thickness [4], the magnetic field [5], and so forth. There also have been several

recent attempts to study the vortex structure of the high-temperature superconductors with MFM [6, 7].

MFM has matured and developed into a routine method for imaging magnetic surface structures [8]. Although MFM is one of the most important imaging tools [9] of nanoscale magnetic structures such as interspin interactions [10], vortex ratchets and cores [11, 12], carrier-controlled ferromagnetism [13], superconducting vortices [14], the separation of magnetic and topographic signals in MFM has been a long-standing issue since its development 20 years ago [8]. This issue still remains largely unsolved and thus has limited the current capability of the MFM as a quantitative magnetic imaging tool [9, 15].

Since magnetic forces between the tip and the sample are very weak ($\sim 10^{-12}$ N), an ac-mode operation with a small vibration amplitude between the tip and the sample is used to detect these weak magnetic interactions [16]. When the oscillation amplitude is used as a feedback signal, the image obtained represents a constant force gradient contour of the magnetic sample surface. Most of the systems currently implemented in MFM are based on mechanically driven vibrations utilizing a piezo device called a bimorph (see Fig. 55.1a).

The bimorph-driven system is limited in its ability to separate the MFM signal from the topographic signal on surfaces where topographic features match or exceed the tip height where imaging occurs [15, 17, 18]. Mixing of the topographic and magnetic

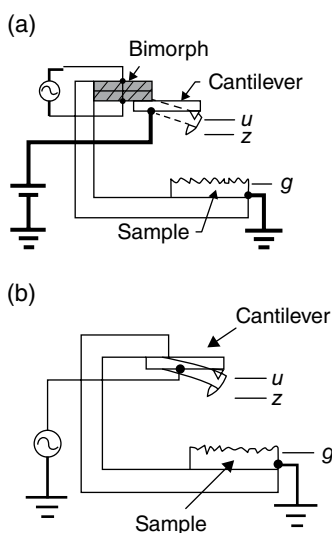


FIGURE 55.1 (a) Schematic of the bimorph-driven system. (b) Schematic of the electrostatic force modulation system. Source: Reprinted with permission from Ref. 25. Copyright (2012), American Institute of Physics.

signals can make it difficult to extract the intrinsic magnetic structure from MFM images, thereby limiting our understanding of magnetic surfaces.

Existing separation methods have focused on monitoring magnetic signals as the tip follows the sample topography. Schönenberger et al. developed an MFM technique that monitors the dc magnetic force (magnetic image) under the constant amplitude (topography image) between the sample and the tip [19, 20]. The technique gives *reasonably good* separation of topographic and magnetic information. Later, Giles et al. made an important contribution to the separation of the two signals through the development of a two-pass technique, the so-called “tapping/lift mode” MFM [21]. In this technique, the first scan is for the topographic signal, then the second scan is repeated to record magnetic information (either as variations in amplitude, frequency, or phase of the cantilever oscillation) in the same line scan at a constant elevated height above the surface. The technique is good at eliminating topographic features in the magnetic image, but the tip contact to the sample during the tapping scan causes some problems.

First, the tip stray field can frequently and significantly disturb the magnetization distribution in a sample, especially in a soft magnetic material, during tapping [22]. Second, the sharp magnetic tip may be easily worn in the presence of topographic variations on the sample surface [15]. Third, the imaging mode places higher demands on instrument stability, or the continual alternating line scan causes serious reduction of the correlation between the two scans due to drift [15]. To minimize these problems, the topographic features need to be separated from MFM images *without touching the sample surface*, thus minimizing the influence of the tip stray field on a sample magnetization.

As we shall see in this chapter, however, magnetic images taken with the conventional noncontact amplitude modulation (AM) MFM frequently picked up topographic features even at larger distances (100–200 nm) [15, 23]. Here, the mixing mechanism of the two signals in the conventional AM MFM is identified using nonlinear dynamics between the tip and the surface. An MFM method has recently been designed and developed using electrostatic force modulation to separate the magnetic domain structure from the topographic structure on magnetic samples with rough surfaces [24, 25]. In this method, a capacitive coupling is introduced between the tip and the sample using electrostatic force modulation (see Fig. 55.1b).

In this chapter, an electrostatic force modulation technique is introduced and the detailed mechanism of the stability improvement is described through a direct comparison of the traditional bimorph method and the electrostatic force modulation method. This will be done by comparing the amplitude–frequency curves for the two systems. The effects on imaging stability of tip–sample separation and perturbations (e.g., collisions between the cantilever tip and a tall hillock structure) are also investigated. Through this investigation, it will be revealed that the superior performance of the electrostatic force modulation system results from the long-range electrostatic capacitance effect and the direct force modulation effect.

55.2 EXPERIMENT

55.2.1 Principle

The technique can be understood by considering the equation of motion of the cantilever probe with a single-point mass m undergoing one-dimensional (1D) forced harmonic oscillation along the vertical z axis:

$$m \frac{\partial^2 z}{\partial t^2} + \gamma \frac{\partial z}{\partial t} + k(z - u) = F(z) \quad (55.1)$$

The probe–surface interaction force can be incorporated as

$$F(z) = F_0 + F'(z_0)\zeta \quad (55.2)$$

and the positions of the bimorph and the lever, u and z , respectively, can be given by

$$u = u_0 + a \exp(i\omega t) \quad (55.3)$$

$$z = z_0 + \zeta \quad (55.4)$$

Then,

$$m \frac{\partial^2 \zeta}{\partial t^2} + \gamma \frac{\partial \zeta}{\partial t} + k[\zeta - a \exp(i\omega t)] = F'\zeta \quad (55.5)$$

The amplitude of vibration of lever is given by

$$A_b(\omega, F') = \frac{ak \exp(i\theta)}{k' - \omega^2 m + i\omega\gamma} \quad (55.6)$$

where $k' = k - F'$. When the driving frequency is near the resonance frequency, $\omega \sim \omega_0$, the local force gradient F' will shift the resonance frequency by an amount $\Delta\omega_0 \sim \omega_0(F'/2k)$. This resonance frequency shift causes the change of the amplitude in the cantilever vibration. The constant force gradient contour of a selected area can be obtained by taking the servo electronics input as data.

55.2.2 Instrumentation

A heavily doped Si cantilever coated with a thin magnetic Co layer is used (nanosensors) for the detection of the magnetic force. The cantilever has a force constant of 2.1 N/m and a resonant frequency of about 106 kHz. The lock-in phase was set to make the driving signal be “in phase” with the response signal of the cantilever at

the tip-sample distance greater than 1000 nm from the sample surface. The in-phase ω component amplitude, $X_\omega (=R_\omega \cdot \cos \phi_\omega)$, was measured as a function of tip-sample distance to understand its behavior. The in-phase amplitude is selected as a feedback signal because of its higher sensitivity with the tip-sample distance over the amplitude R_ω [25]. The tip is magnetized by placing the cantilever in a magnetic field of 0.2 T, aligned perpendicularly to the lever, for 3 min [18]. The electrical contact for the tip and sample was made with a silver paste. The scan rate is 1 Hz, and the time constant is set at 100 μ s in a lock-in amplifier (EG&G Princeton Applied Research, Model 5302). A CoCr film deposited on the glass substrate by the dc magnetron sputtering method up to a thickness of 300 nm was used as a sample for this study. All the data shown here are obtained with a commercial AutoProbe LS in air [26].

MFM cantilevers attached on the bimorph were driven by the voltage-controlled oscillator (VCO). For MFM imaging using the bimorph-driven method in Figure 55.1a, a dc voltage +10 V was applied between the tip and the sample to provide the servo force F_c for feedback to keep the tip from crashing into the surface during scanning [27]. For the electrostatic force modulation method in Figure 55.1b, a sinusoidal signal $V_{ac} \sin(\omega t)$ was applied using a function generator (Hewlett Packard, HP 33120A) to modulate the gap between the tip and the sample. In this experiment, we applied $V_{ac} = 14$ V between the tip and the sample surface at an operating frequency of 53 kHz, half the resonance frequency of the cantilever. For both magnetic imaging methods, magnetic and topographic images were obtained under a constant amplitude feedback condition by using the difference between the output of the amplitude and set amplitude as the error signal. The typical scan rate was 1 Hz. The cantilevers used here are the microfabricated triangular Si_3N_4 cantilevers with pyramidal tips.

The AutoProbe LS system employed an optical beam deflection-detection method, which is one of the most commonly used methods to detect interaction between the magnetic tip and the magnetic sample. The optical-detection system is composed of a laser diode, a bicell photodiode as a position sensitive detector, and a mirror for aligning laser beam on the backside of cantilever. Under the cantilever, there is the sample stage mounted on the tube piezo scanner. The tip can be positioned on the selected area of the sample surface by an XY micrometer stage combined with CCD camera.

55.2.3 Approach

The tip is approached to the sample after loading the sample and focusing the laser beam on the backside of the cantilever. At the moment the cantilever amplitude drastically decreases, the tip is retracted from the contact point about 100 nm by the stepping motor. This trapping of the tip can also be observed easily in the CCD image. It implies that there is a strong adhesive force directly on the magnetic sample. From the amplitude-frequency curve obtained at that point, we choose an optimum frequency at which the amplitude is 81.6% of the resonance peak [28]. In this experiment, the

operating frequency is 53 kHz. The absolute vibration amplitude of the cantilever can be determined using the Michelson–Morley interferometry at this operating frequency before the experiment. After the scan size, the offset, the time constant, the gain, and the slope setting are determined, the images are obtained.

The amplitude changes are measured with the lock-in amplifier as the error signal of the feedback. If the output of the bicell photodiode is the error signal of feedback directly, MFM can be operated as an atomic force microscope (AFM). This is useful when a topographic image of a magnetic sample is needed. The feedback circuits are composed of the control circuit and the high-voltage amplifier. The feedback-control parameters, like the time constant and the gain, can be manually adjusted for the optimum feedback condition.

It takes 5–10 min to get an MFM image for 256×256 data acquisition. In the data file, the scan frequency, the scanning area, the offset, the bias voltage, the data gain, and the force gradient set value are also stored for each image. Image process and analysis can be done after data acquisition. When data are acquired in the constant force gradient mode with the feedback on, the image represents a mapping of the force gradient experienced by the magnetic tip as a function of position. Maximum scanning area is $50 \times 50 \mu\text{m}$ with the system used. The finest structure observed in our MFM images is in the region of 50–100 nm. All data presented here are raw data.

55.3 RESULTS AND DISCUSSION

We must approach to improve the resolution of MFM and study any relation between topography and magnetic structure, if we want magnetic tip to the surface more closely ($\sim 200 \text{ \AA}$). In this case, it is important to separate the topographic and magnetic features of the sample. One way to solve this problem is to apply a sinusoidal voltage to the tip to induce the modulated electrostatic force. The information of modulated electrostatic signal extracted by the lock-in will be used to control the height of the scan [24].

If we assume that the cantilever is parallel to the sample and that the tip is a point dipole with fixed moment, $\vec{m} = (m_x, m_y, m_z)$, in the stray magnetic field of the sample surface, $\vec{B} = (B_x, B_y, B_z)$, then we can write down force derivative as

$$F' = m_x \frac{\partial^2 B_x}{\partial z^2} + m_y \frac{\partial^2 B_y}{\partial z^2} + m_z \frac{\partial^2 B_z}{\partial z^2}. \quad (55.7)$$

Since the force derivative change is related with the second derivative of the sample stray field, the contrast will be obvious at the abrupt change of the vertical component and the horizontal component of the magnetization like the domain boundary. Therefore, the lines of the image seem to be the natural magnetic domain walls with the thickness of about 50–100 nm. For more complete understanding, this value must be compared with the result based on the magnetic anisotropy measurement [29].

55.3.1 Separation of Topographic Features from Magnetic Force Images Using Capacitive Coupling Effect

55.3.1.1 Topographic Features in a Magnetic Force Image A magnetic tip is mechanically vibrated at the resonance frequency of the cantilever with the free oscillation amplitude of 96 nm by the acoustic excitation method such as the bimorph-driven system (inset of Fig. 55.2a). A dc bias voltage $V_{dc} = +10\text{V}$ is applied between the tip and the surface to create a long-range attractive electrostatic interaction, which is essential to make the feedback polarity stay constant, regardless of the attractive or repulsive magnetic forces for stable feedback [20, 23]. In the in-phase amplitude–distance curve, the amplitude decreases monotonically as the tip moves toward the surface for the distance between 64 and 1000 nm (“noncontact region”) in Figure 55.2a. For magnetic imaging, the average tip–sample distance of the oscillating tip should be positioned in the noncontact region where the magnetic signal is dominant over the topographic signal. Since the sign of slope is positive (i.e., $\partial X_{\omega}(d)/\partial d > 0$) in the noncontact region, it is necessary to set the feedback polarity to positive for stable magnetic imaging. The linear tapping region also has the same sign of slope [30, 31], indicating that two stable states exist for a given set-amplitude ($X_{\omega,SP}$): one in the tapping region and the other in the noncontact region.

The key approach of this chapter is to solve a 20-year-old problem by using the fundamental understanding of nonlinear stochastic physics in the tip–sample interactions of AM AFM, recently discovered and published by Garcia and San Paulo [32].

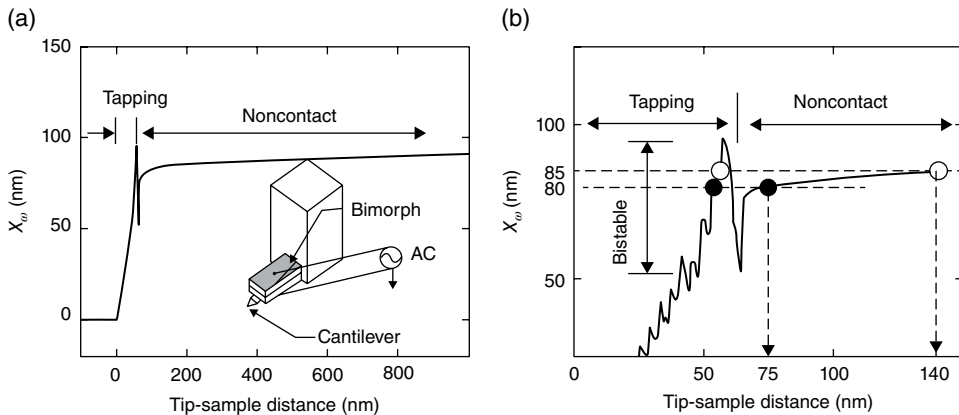


FIGURE 55.2 (a) A typical in-phase amplitude–distance curve at the operating frequency of resonance 106 kHz and the free oscillation amplitude of 96 nm on a CoCr film. (inset) A schematic sketch of the acoustic excitation method. (b) An enlarged amplitude–distance curve in the distance range between 0 and 150 nm. Horizontal dashed lines represent the feedback set amplitudes of 80 and 85 nm that correspond to the average noncontact distances of 75 and 140 nm from the surface, respectively, as marked with dashed arrows. The two stable states are marked with two circles for each set amplitude. Source: Reprinted with permission from Ref. 24. Copyright (2009), American Institute Physics.

Garcia and San Paulo associated abrupt changes in height of topographic features with the continual switching of the oscillating tip between the two stable states during the tip scanning over the surface in the AM AFM. Similarly, in this chapter, the topographic features in the magnetic image are attributed to the switching between the bistable states. This effect is analogous to the sudden transition from the noncontact to contact states in contact mode AFM [22]. In principle, both effects result from the intrinsic nonlinear mechanical bistability of the sensor-sample assembly. In the contact AFM, the so-called “snap-to-contact” issue has been resolved by removing the bistability using a voltage-activated force feedback [33, 34]. Similarly, in this chapter, this issue has been addressed by removing the stable state in the tapping region. The bistability originates from the dc bias voltage to induce *monotonic decrease* of the long-range amplitude as the distance becomes smaller in the noncontact region between 64 and 1000 nm (Fig. 55.2a).

In order to verify this concept, the magnetic imaging was performed repeatedly at two different set-amplitudes ($X_{o,SP}$) of 80 and 85 nm at the same tip location. The average operating spacing can be determined graphically by projecting to the x -axis from the intersecting points where the open-loop curve and a feedback set point line meet, as shown in Figure 55.2b. The origin is chosen as the point where the linear-extrapolated line in the tapping region crosses with the x -axis, and thus the x -coordinate represents the average tip-sample distance [31]. The average noncontact distances are determined to be 75 and 140 nm for the set amplitudes of 80 and 85 nm, respectively. Figure 55.3a and b shows stripe-like structures that correspond to the magnetic domain features because of a higher contrast variation at a larger tip-sample distance (see line scans in Fig. 55.3a and b), a well-known signature for magnetic features [35]. Topographic hillocks and grains appear as spots with diameters of 1000–2000 nm and spots with diameters of 100–300 nm, respectively, in Figure 55.3a (see more details in Fig. 55.5b). Most of the topographic features drastically disappear except a few hillocks marked with arrows in Figure 55.3b. The remarkable change of pickup ratio results from the noncontact lift-height change by 65 nm ($= 140 - 75$ nm) in Figure 55.2b. In the constant amplitude mode, the average tip-sample distance continues to vary in order to maintain the set-amplitude constant. Because the set-amplitude 80 nm is comparable to the average tip-sample distance 75 nm, feedback perturbations allow for switching between the bistable states for picking up topographic features. The switching between two stable states would be almost equally probable during the data acquisition, explaining the pickup ratio of nearly one as shown in Figure 55.3a. At the set amplitude of 85 nm, the average noncontact distance of 140 nm is somewhat bigger than, but still smaller than, the heights of bigger hillocks of 200–300 nm. The oscillating tip spends most of its time collecting magnetic features *in the noncontact region* except in the regions of the bigger hillocks, explaining the small pickup ratio in Figure 55.3b. The bistable states exist for almost all of the average tip-sample spacing from 64 nm up to >1000 nm in Figure 55.2a. This indicates that topographic features always have a chance to appear in a magnetic image, except for smooth and homogeneous sample surfaces.

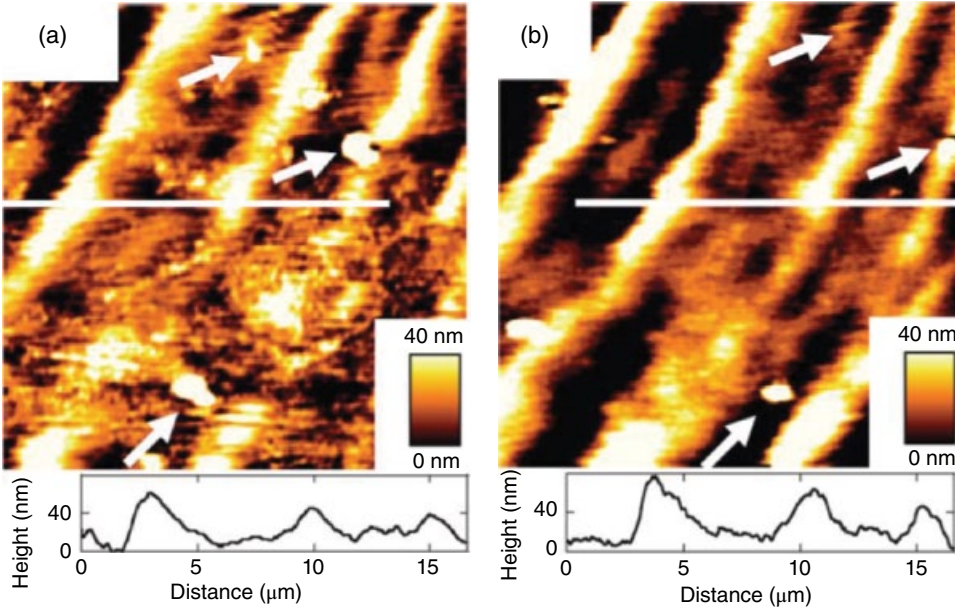


FIGURE 55.3 (a) A stripe-like magnetic domain image with hillocks and magnetic grains on the CoCr film (scan area: $20 \times 20 \mu\text{m}$). (b) The same stripe-like magnetic domain image with sporadic hillocks with $4 \mu\text{m}$ shift to the left from the position of (a) (scan area: $20 \times 20 \mu\text{m}$). (insets) Line scans along the white lines in each image for comparison of the contrast variation between two images. Common topographic hillocks in both images are as marked with arrows for comparison. Source: Reprinted with permission from Ref. 24. Copyright (2009), American Institute Physics.

55.3.1.2 Separation of Topography and Magnetic Structures As a method to make the *amplitude increase* in the noncontact region as the tip approaches the surface, an electrostatic force modulation method is introduced to use the capacitive coupling effect for magnetic imaging (inset of Fig. 55.4a). The same rough surface is used for a direct comparison between both methods in separating topographic features from an MFM image. A sinusoidal signal $V_{ac} \cdot \sin(\omega t)$ with $V_{ac} = 10 \text{ V}$ is applied between the same tip and the same surface using a function generator (Hewlett Packard, HP 33120A). The operating frequency (f_{op}) was set to 53 kHz where in-phase 2ω component amplitude, $X_{2\omega} (= R_{2\omega} \cdot \cos \phi_{2\omega})$, has the resonance peak for the cantilever with the resonance frequency of 106 kHz [36]. In Figure 55.4a, the in-phase amplitude increases in the noncontact region as the tip approaches the surface. The enlarged curve (Fig. 55.4b) shows that the horizontal set-amplitude line ($X_{2\omega} = X_{2\omega,SP}$) meets twice with the in-phase amplitude–distance curve, but the signs of the slope $\partial X_{2\omega} / \partial z$ (i.e., the feedback polarity) at the two intersecting points are different from each other. The result shows that only one stable state (in either tapping or noncontact region) is available for a given feedback polarity.

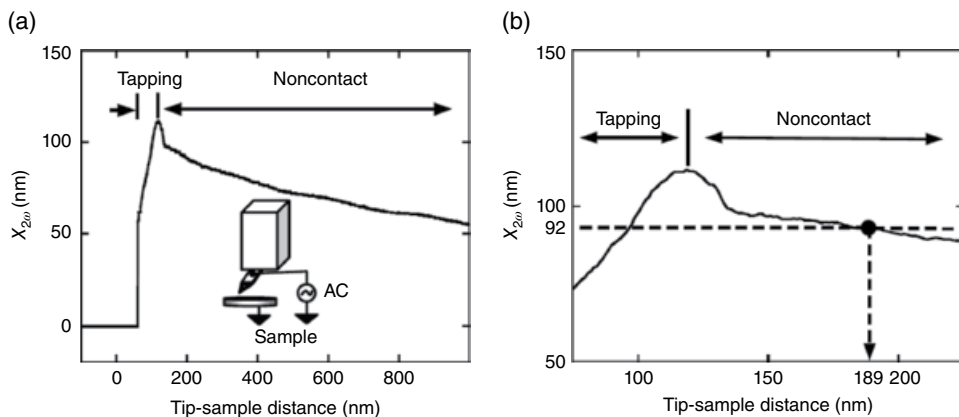


FIGURE 55.4 (a) A typical in-phase amplitude–distance curve at the operating frequency of resonance 53 kHz. (inset) A schematic sketch of the electrostatic force modulation method. (b) An enlarged amplitude–distance curve in the distance range between 75 and 225 nm. A dashed line representing the feedback set amplitude of 92 nm corresponds to the average non-contact distance of 189 nm from the surface. A stable state is marked with a solid circle in the noncontact region. Source: Reprinted with permission from Ref. 24. Copyright (2009), American Institute of Physics.

Magnetic imaging of the same CoCr magnetic film was performed under the feedback condition of a constant set amplitude of 92 nm in the noncontact region to investigate the effect of the capacitive coupling on the separation. The imaging condition is similar to the conditions of Figure 55.3b because the noncontact distance of 189 nm is much bigger than the set amplitude of 92 nm, but still smaller than the heights of the hillocks. Figure 55.5a shows a magnetic image of well-connected labyrinthine structures with the periodicity of 4.5–5 μm . The image does not show any evidence of the topographic features. More magnetic images are collected repeatedly in different scan areas from 100 \times 100 nm to 40 \times 40 μm for the different set values between 82 and 96 nm (not shown). However, neither image shows the topographic features, indicating that there is no mechanical contact between the tip and the surface during MFM data acquisition. The reproducibility of MFM data without the tip crashing toward the surface suggests the enhanced stability of MFM method using the capacitive coupling over the conventional AM MFM. In order to observe the topographic features, the oscillating tip is brought into the tapping region by reversing feedback polarity with the same set amplitude of 92 nm (see Fig. 55.4b). Figure 55.5b shows several hillocks with the diameters of 1000–3000 nm and heights of 200–300 nm and aggregations of small grains with the sizes of 100–300 nm and the heights of 10–40 nm, consistent with those observed in Figure 55.3a and b. The consistency indicates that the charging effect due to the electrostatic modulation is not usually important for general conductive magnetic samples [19].

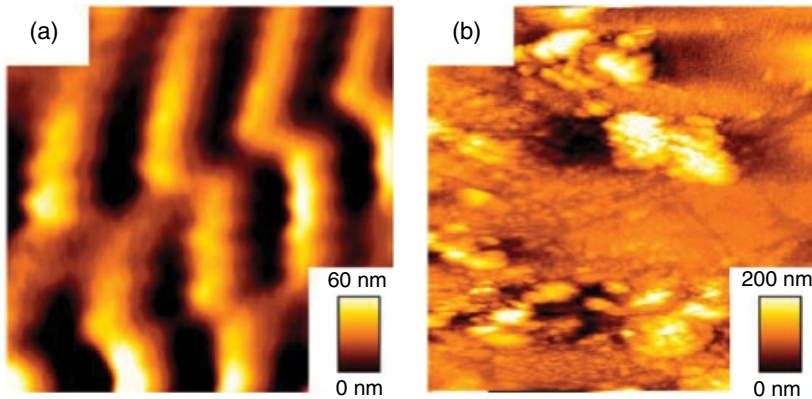


FIGURE 55.5 (a) The stripe-like magnetic domain image in the noncontact regime with the set amplitude of 92 nm (scan area: $20 \times 20 \mu\text{m}$). (b) Topographic image of CoCr film taken with the electrostatic tapping mode with the set amplitude of 92 nm (scan area: $20 \times 20 \mu\text{m}$). (insets) Line scans along the white lines in each image for comparison of the contrast variation between two images. Common topographic hillocks in both images are as marked with arrows for comparison. Source: Reprinted with permission from Ref. 24. Copyright (2009), American Institute Physics.

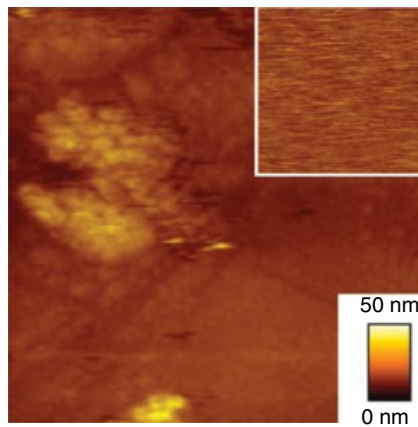


FIGURE 55.6 A topographic image taken by a nonmagnetic tip in the tapping regime with the set amplitude of 92 nm (scan area: $20 \times 20 \mu\text{m}$). (inset) An image taken by the same nonmagnetic tip in the noncontact regime with the set amplitude of 92 nm (scan area: $20 \times 20 \mu\text{m}$). Source: Reprinted with permission from Ref. 24. Copyright (2009), American Institute Physics.

An image taken in a tapping regime by a nonmagnetic commercial conductive Si cantilever [36] is shown in Figure 55.6 where the topographic features are similar to those in Figure 55.5b. When an image was taken in a noncontact regime (inset of Fig. 55.6), it was completely featureless without showing any evidence of magnetic features within the current noise level of the system. The result indicates that topographic interactions (e.g., electrostatic force) are too weak to provide observable

topographic features. It also suggests that the magnetic interactions should be dominant over the topographic interactions at a noncontact distance around approximately 200 nm where the magnetic image of Figure 55.5a was taken.

55.3.2 Effects of Long-Range Tip–Sample Interaction on Magnetic Force Imaging: A Comparative Study Between Bimorph-Driven System and Electrostatic Force Modulation

In this section, the detailed mechanism of the stability improvement is described through a direct comparison of the traditional bimorph method and the electrostatic force modulation method. Figure 55.7a and b compares the amplitude–frequency curves for both the electrostatic force modulation MFM system and the bimorph-driven MFM system. In the bimorph-driven system (Fig. 55.7a), unwanted peaks around the cantilever resonance result in unstable feedback conditions during MFM imaging. This poor frequency response of the cantilever is due to the additional frequency components integrated near the primary resonance peak, as shown in Figure 55.7a. The existence of several resonance peaks imposes drastic restrictions on the operating frequency range at which useful measurements can be made. In the electrostatic force modulation system (Fig. 55.7b), the cantilever frequency response has two well-defined resonance peaks (ω and 2ω components only). This indicates that the cantilever is the only component being driven by electrostatic force modulation.

Figure 55.8a and b shows maze-like magnetic domain structures with periodicity 3–5 μm . These results are consistent with our previous study [24]. However, topographic features frequently overlap in magnetic images obtained by MFM using bimorph-driven modulation, as shown in Figure 55.8a. The overlapped topographic features appear to be associated with previously observed large hillocks with diameter 5–10 μm and height 100–300 nm [24]. Figure 55.8b shows an MFM image of the striped magnetic domains on the CoCr magnetic film using electrostatic force modulation. We reproduced the same image repeatedly at the same location, indicating the improved stability of the electrostatic force modulation technique.

Figure 55.9a and b illustrates the differences in imaging stability obtained by bimorph-driven and electrostatic MFM systems. Figure 55.9a shows an MFM image of stripe domain structures with periodicity 4.5–5 μm on CoCr magnetic film, consistent with previously reported MFM images [16]. As the bimorph-driven tip scans over the magnetic surface, it appears to crash into a large hillock that is taller than the tip height. The crash of the tip and a hillock structure (as marked by an arrow in the middle of y-scan) creates a feedback disturbance, overcoming the barrier for the imaging mode transition from magnetic mode to topographic mode [30]. After the crash, the tip remains in tapping mode as it continues to scan in the y-direction [30]. The topographic image mode stays until the tip completes scanning, different from previous observations of reversible switches between noncontact and topographic modes [24, 32]. The persistence of topographic imaging mode in the bimorph-driven system indicates that

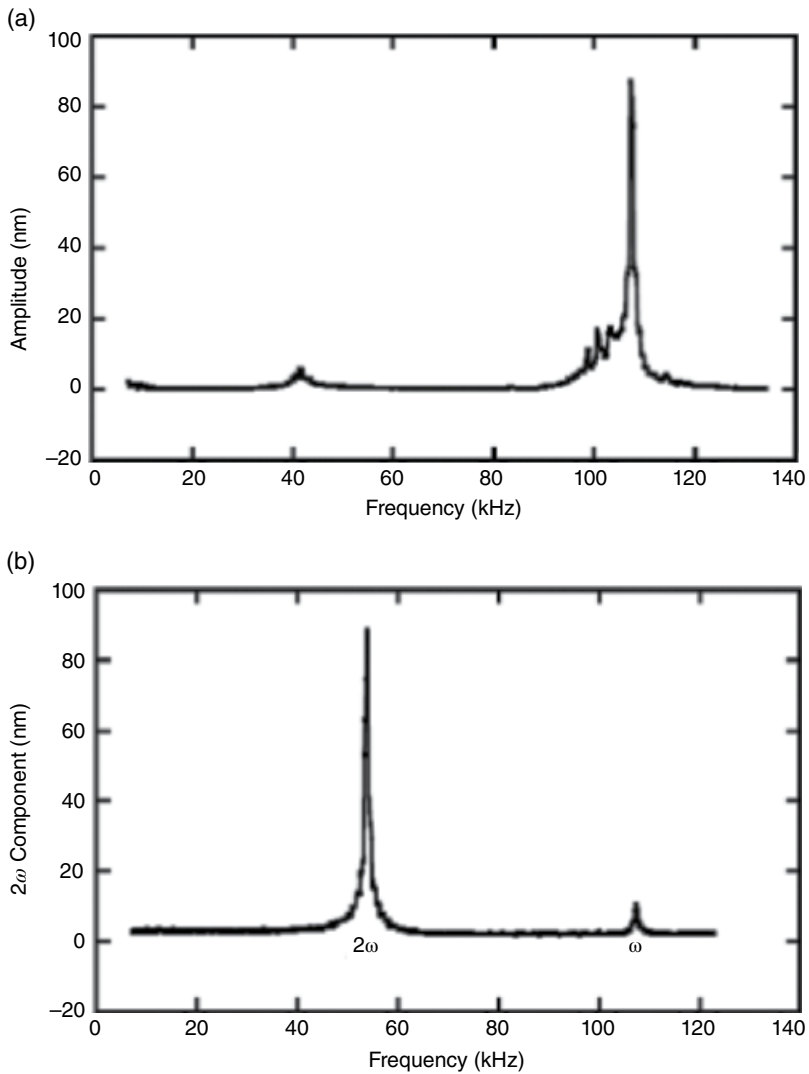


FIGURE 55.7 Amplitude–frequency response for the cantilever. (a) The bimorph-driven system. (b) The electrostatic force modulation system. Source: Reprinted with permission from Ref. 25. Copyright (2012), American Institute Physics.

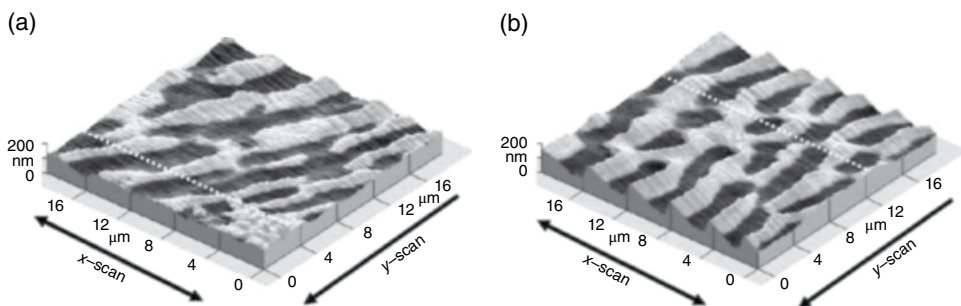


FIGURE 55.8 (a) MFM images showing maze-like magnetic domain structures with periodicity 3–5 μm (scan area: 20×20 μm) taken with bimorph-driven system and (b) electrostatic force modulation system. Source: Reprinted with permission from Ref. 25. Copyright (2012), American Institute Physics.

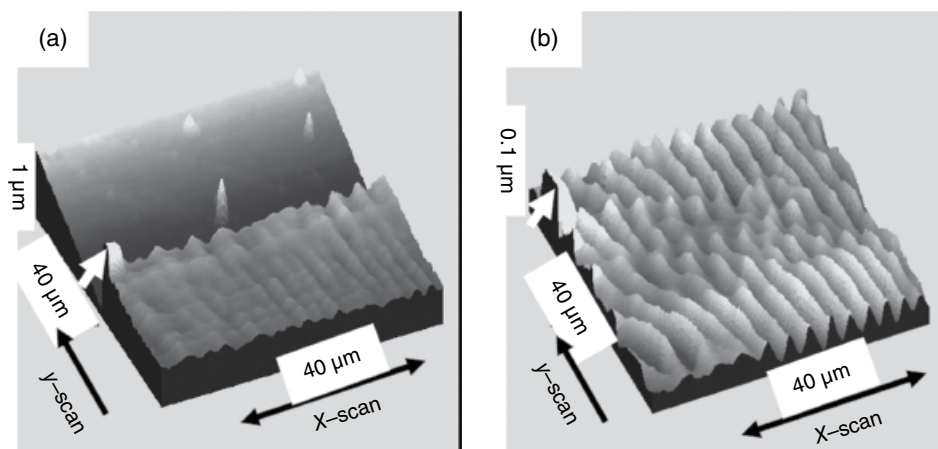


FIGURE 55.9 (a) Transition from magnetic domain imaging mode to topographic imaging mode during tip scanning of CoCr film with set-amplitude 85 nm using bimorph-driven modulation (scan area: $40 \times 40 \mu\text{m}$). The bimorph-driven system is more susceptible to large topographic features such as the feature indicated by an arrow, which causes the imaging mode to change from magnetic to topographic. (b) Magnetic domain image of CoCr film in the noncontact regime with set-amplitude 95 nm operating at the resonance frequency 53 kHz using the electrostatic force modulation system (scan area: $40 \times 40 \mu\text{m}$). The system is robust against frequent tip collisions with topographic hillocks such as the feature indicated by an arrow. Source: Reprinted with permission from Ref. 25. Copyright (2012), American Institute Physics.

there is a hysteresis in switching between tapping mode and the noncontact magnetic mode. This irreversible process indicates that there might be a change in the energy barrier between the noncontact and tapping regions in the amplitude–distance curve. The disappearance of noncontact mode in Figure 55.9a can possibly be explained by a change in the tip structure: a tip radius decrease during the crash causes the noncontact region of the amplitude–distance curve to move upward due to a capacitance decrease. This upward move eliminates the intersection point between the horizontal set amplitude and the amplitude–distance curve in the noncontact region of Figure 55.10b [24].

Figure 55.9b shows a magnetic image from the same sample surface, this time using electrostatic force modulation. As the electrostatically driven tip scans over the magnetic surface, it again encounters the large hillock that is taller than the tip height. After the encounter, the tip stays in noncontact mode as it scans in the y-direction. The persistence of noncontact mode in the electrostatically driven system can possibly be understood by the opposite polarity of slope of the amplitude–distance curve in the tapping and noncontact regions, respectively, as shown in Figure 55.10a [24]. In addition, the completeness of the magnetic image beyond the encounter indicates that the enhanced barrier height in amplitude (i.e., the difference between peak height and the horizontal set amplitude) of the electrostatic force modulation system (over the

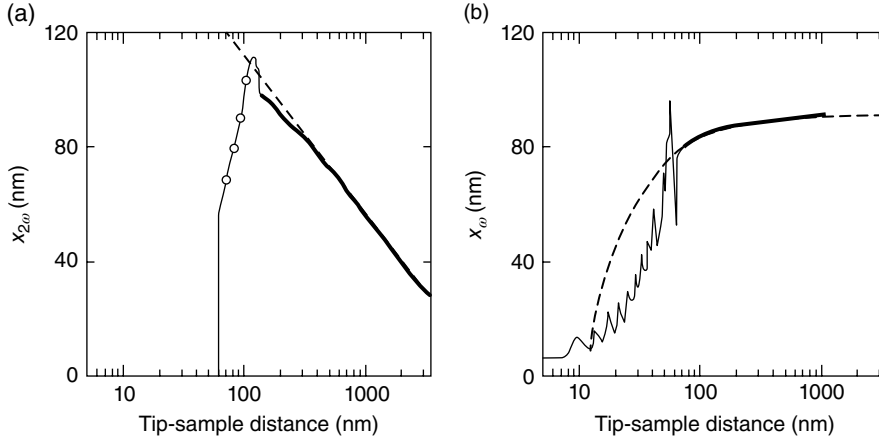


FIGURE 55.10 (a) Electrostatic force modulation system: Solid curve fits the amplitude–distance data. Note the barrier between the noncontact and tapping regions and the opposite feedback polarity (sign of slope) in each region. Dashed line fits the data to the logarithmic function $b \cdot \log(D/z)$ over the entire noncontact region from 200 nm to 4 μm with the fitting parameters $b = 1.9 \text{ nN/V}^2$, $D = 1 \mu\text{m}$. (b) Amplitude–distance data for the bimorph system. Note the identical feedback polarity in the noncontact and tapping regions. Source: Reprinted with permission from Ref. 25. Copyright (2012), American Institute Physics.

bimorph system) prevents the cantilever from snapping upon contact with the surface. Since there are no appreciable height changes in the magnetic image, the tip appears to be more secure under the electrostatically modulated system.

In the bimorph-driven system, feedback bistability between the noncontact and tapping regions in the amplitude–distance curve frequently causes overlap of the topographic signal onto the magnetic signal during scanning. Figure 55.10a shows the amplitude change as the tip approaches the sample surface in the electrostatic force modulation system. The amplitude increases in the noncontact region but decreases in the tapping region. This behavior is very different from that obtained with the bimorph-driven system, as shown in Figure 55.10b. In the bimorph-driven system, the feedback polarity is the same in both the noncontact and tapping regimes.

In the usual constant force gradient mode with bimorph-driven modulation, a dc voltage is applied between the tip and the sample to prevent the tip from crashing into the surface during scanning. Instead, we employed the electrostatic force modulation technique for self-actuation of the cantilever to satisfy the requirements for MFM imaging. For the direct force modulation in a novel MFM, we utilized capacitive force coupling between the cantilever and the sample to induce the modulation (Fig. 55.1b). To avoid mixing of the topographic signal with the MFM image, we must increase the operating distance of the noncontact region to avoid the tip crashing near the tapping region. The opposite feedback polarity between noncontact mode and tapping mode is necessary to exclude possible mixing of topographic signal due to the tapping mode

operation during noncontact MFM imaging. Figure 55.1b shows the electrostatic force modulation as a servo force to keep the polarity of feedback constant for both repulsive and attractive magnetic forces. The electrostatic force keeps the total force gradient always positive regardless of changes in the polarity of the magnetic force. The decrease in amplitude results from the sudden involvement of the repulsive tapping force. The peak in Figure 55.10a represents a transition from an electrostatic noncontact interaction to tapping interaction between the tip and the sample. The linear part in the tapping regime is used to find the amplitude of the tip vibration in the noncontact regime because the z -piezo is precisely calibrated [32, 37, 38].

However, in the bimorph-driven system, the linear tapping region also has the same sign of slope as the noncontact region [24]. This indicates that there exist two *stable* states for a given set-amplitude ($X_{\omega,SP}$): one in the tapping region and the other in the noncontact region. The coexistence of two stable states is analogous to that in the AFM [32, 39]. Garcia and San Paulo associated continual switching of the oscillating tip between the two stable states with abrupt changes in height of topographic features. Similarly, we attribute the appearance of topographic features in a magnetic image to the switching between the bistable states.

If this switching mechanism is the major channel of picking up topographic features in a magnetic image, the pickup ratio should depend upon the average distance of the noncontact state from the sample surface. In order to check this idea, we performed the magnetic imaging as a function of feedback set point amplitude $X_{\omega,SP}$ at the same tip location repeatedly. Using electrostatic force modulation, tip crashing toward the surface upon encountering the peak barrier in Figure 55.10a rarely happens during repeated magnetic imaging. Hong et al. reported the same type of stable imaging condition in tapping mode AFM using the electrostatic force modulation [40].

The equation of motion of the cantilever probe is employed to reveal the stability of the mechanism: the probe with a single-point mass m undergoes 1D forced harmonic oscillation along the vertical z axis in Figure 55.1a and b for bimorph-driven modulation and electrostatic force modulation, respectively. This oscillation is described by the following equation:

$$\frac{\partial^2 z}{\partial t^2} + \frac{\omega_0}{Q} \cdot \frac{\partial z}{\partial t} + \omega_0^2 \cdot (z - u) = \frac{1}{m} (F_m + F_c + F_{vdw}) \quad (55.8)$$

where z is the coordinate perpendicular to the sample surface, t is time, u is the undeflected cantilever position, k is the force constant, ω_0 is the resonance angular frequency, Q is the quality factor, and ω_0/Q corresponds to the damping constant per unit mass in the presence of magnetic force F_m , capacitance force F_c , and van der Waals force F_{vdw} . Equation 55.8 can be solved analytically when the average tip-sample distance u is much bigger than the set-amplitude $X_{\omega,SP}$, for both bimorph-driven and electrostatic force modulation systems. The position u in the equation of motion (55.8) is constant while ac voltage is applied.

For electrostatic force modulation, the square law dependence of the capacitive force F_C on the driving signal $V_{ac} \cos(\omega t)$ induces a mechanical vibration of the cantilever:

$$F_C = \frac{1}{4} \frac{\partial C}{\partial z} V_{ac}^2 (1 + \cos 2\omega t) \quad (55.9)$$

where C is the capacitance between the tip and the sample. The in-phase amplitude of the 2ω component, $X_{2\omega}$, is predicted as follows:

$$X_{2\omega} = \frac{1}{2m} \cdot \frac{-F_C(z, V_{ac}) \cdot 2(\omega_0'^2 - 4\omega^2)}{\sqrt{(\omega_0'^2 - 4\omega^2)^2 + (2\omega\omega_0/Q)^2}} \sin(\phi_{2\omega} + \phi_0) \quad (55.10)$$

where ϕ_0 is 108° that maximizes the $X_{2\omega}$ near $u = 100$ nm.

$$\tan \phi_{2\omega} = \frac{2\omega\omega_0}{Q \cdot (\omega_0'^2 - 4\omega^2)} \quad (55.11)$$

where $\omega_0'^2 = \omega_0^2 \left(1 - \frac{1}{k} \left(F_m' + \frac{1}{2} \cdot F_C'(z, V_{ac}) \right) \right)$.

$$F_C(z, V_{ac}) = \frac{1}{2} \frac{\partial C}{\partial z} V_{ac}^2 \quad (55.12)$$

$$F_C'(z, V_{ac}) = \frac{1}{2} \frac{\partial^2 C}{\partial z^2} V_{ac}^2 \quad (55.13)$$

The Lorentzian Equation 55.10 accounts for the behavior of $X_{2\omega}$ in Figure 55.7b. The earlier predicted solutions, Equations 55.10 and 55.13, are compared with the experimental data to find the dependence of the interaction with distance in Figure 55.10a and b. The amplitude increases in the noncontact region using electrostatic force modulation since the change of the electrostatic capacitance force $-F_C(z, V_{ac})$ in Equation 55.10 increases as the distance z becomes smaller. Experimental data in Figure 55.10a shows that the electrostatic interaction extends to $3 \mu\text{m}$ (equivalent to the limit of our z -piezo). The solid line of the prediction of Equation 55.10 matches well with the experimental data. In the prediction, the electrostatic capacitance force $-F_C(z, V_{ac})$ is described by assuming a conical tip. In this case, the capacitance gradient in Equation 55.12 is known to vary with a single logarithmic function $\partial C / \partial z = b \cdot \log(D/z)$, where the constants b and D depend on the angle and length of the cone [41, 42]. They are determined to be 1.9 nN/V^2 and 1000 nm , respectively, from the predictive amplitude curve (solid line). The amplitude data curve (Fig. 55.9a) in the distance range $200\text{--}4000 \text{ nm}$ fits with the predictive curve that assumes the logarithmic function (i.e., conical tip shape). The fitting result suggests that the amplitude depends logarithmically on the

tip-sample distance z . Due to the logarithmic long-range electrostatic interaction, good feedback stability was achieved during the image acquisition of magnetic domains in the electrostatic force-driven MFM.

In case of the bimorph-driven system, the equation of motion (55.8) has $F_c = \frac{1}{2} \frac{\partial C}{\partial z} V_{dc}^2$. The bimorph-driven modulation can be described as follows:

$$u = u_0 + a \cdot e^{i\omega t} \quad (55.14)$$

where u_0 is the midpoint of the undeflected cantilever position, a is the driving amplitude, and ω is the angular frequency. For bimorph-driven modulation, the in-phase component is predicted as follows:

$$X_\omega = \frac{a \cdot \omega_0^2}{Q} \cdot \frac{a \cdot \omega_0^2 \cdot (\omega_0'^2 - \omega^2)}{(\omega_0'^2 - \omega^2)^2 + (\omega \omega_0 / Q)^2} \quad (55.15)$$

where $\omega_0'^2 = \omega_0^2 \left(1 - \frac{1}{k} (F'_m + F'_c(z, V_{dc})) \right)$ and

$$F'_c(z, V_{dc}) = \frac{1}{2} \frac{\partial^2 C}{\partial z^2} V_{dc}^2 \quad (55.16)$$

Due to the long-range nature of the electrostatic capacitance force $-F_c(z, V_{ac})$, the amplitude varies significantly for a much longer range (up to 3000 nm) in the noncontact region when electrostatic force modulation is used.

The concavity of the approaching curve is down since the amplitude in the noncontact region is determined by the change of capacitance with the tip-sample distance, $\partial C / \partial z$, in Equation 55.10; we fitted the curve in Figure 55.10a, assuming the conical tip model as a good approximation in the noncontact regime [35, 41]. To support this argument, the $X_{2\omega}$ versus distance curve was fitted with a single logarithmic function representing capacitance coupling between a conical tip and a flat sample surface [40]. Figure 55.10a shows an excellent agreement of the logarithmic function with the curve up to the distance 4000 nm that the current z -piezo tube can maximally extend from the surface, indicating that the conical tip model is a good approximation in the noncontact region. The contrast mechanism during the image acquisition of magnetic domain in the electrostatic force modulation results from the long-range electrostatic servo interaction that depends on the tip-sample distance z logarithmically. Due to the long-range nature of capacitance coupling in the electrostatic interaction, the noncontact region is much greater for electrostatic force modulation than for bimorph-driven modulation in Figure 55.10b. Therefore, the electrostatic force $F_c(z, V_{ac})$ with logarithmic z dependence is the origin of the long-range servo interaction rather than the electrostatic force gradient

$$F'_c(z, V_{dc}) \propto \frac{1}{z},$$

commonly used as the servo interaction in bimorph-driven modulation [43]. Figure 55.10b also shows that the $1/z$ function represents nicely over the noncontact region in the in-phase $X_{2\omega}$ -distance curve of the bimorph-driven system. Due to the long-range electrostatic interaction, good feedback stability during the imaging of magnetic domain can be achieved in the electrostatic force-driven MFM up to the operation distance of a few hundred nanometers.

As the tip approaches the surface, $\partial C/\partial z$ in the numerator $F_C(z, V_{ac})$ contributes more to the $X_{2\omega}$ amplitude than the denominator due to the change of $F'_C(z, V_{ac})$ with z . In the bimorph-driven system, on the other hand, the in-phase vibration amplitude $X_{2\omega}$ is controlled by the change of $F'_C(z, V_{ac})$ due to the absence of the factor $\partial C/\partial z$ in the numerator. When the tip experiences the change of magnetic force $\Delta F'_m$ during a lateral move across a domain wall, both the electrostatic force $F_C(z, V_{ac})$ and the electrostatic force gradient $F'_C(z, V_{ac})$ compensate for a given change in magnetic force gradient by moving the tip vertically relative to the sample surface to maintain the $X_{2\omega}$ constant under feedback. In a bimorph-driven MFM, only an electric force gradient $F'_C(z, V_{ac})$ plays this role. The bigger amplitude in $X_{2\omega}$ at a closer distance in the non-contact region indicates that the in-phase amplitude of the 2ω component is controlled more by $F_C(z, V_{ac})$ than by $F'_C(z, V_{ac})$. The change of magnetic force gradient ($\Delta F'_m$) is compensated by the factor $F_C(z, V_{ac})$ to maintain the in-phase amplitude constant under feedback through the movement of the piezo tube. In the MFM using bimorph-driven modulation, the change of magnetic force gradient ($\Delta F'_m$) is simply compensated by the change of the servo force gradient ($\Delta F'_C$) [16, 35, 44]. The shorter range indicates that due to the collision, the excited tip can easily overcome the cross-over barrier.

Based on the understanding of the interactions of the tip and the sample surface of Figure 55.10a and b, the separation mechanism is described through the comparison between the bimorph-driven system and electrostatic force modulation system in Figure 55.11a and b. In Figure 55.11a, topographic features (shaded profile) and the magnetic features (solid profile line extracted from the dotted line in Fig. 55.8a) appear together across the vertical dashed line due to the involvement of the short-range interaction during data acquisition. This is because the tip is likely to crash with tall topographic features during the oscillatory motion of the cantilever, as illustrated on the right side of Figure 55.11a. Such a coexistence of topographic and magnetic structures on an MFM image is known to depend on the distance between the tip and the surface and oscillatory amplitude [24]. In the electrostatic force modulation system, the long-range nature of the interaction keeps the tip from being crashed into the topographic features, as shown in Figure 55.11b. Due to such long-range of interactions, the MFM image has only magnetic structures without having any topographic features as found in Figure 55.8b. This is because the oscillatory amplitude is smaller than the tip-sample distance as depicted in Figure 55.11b.

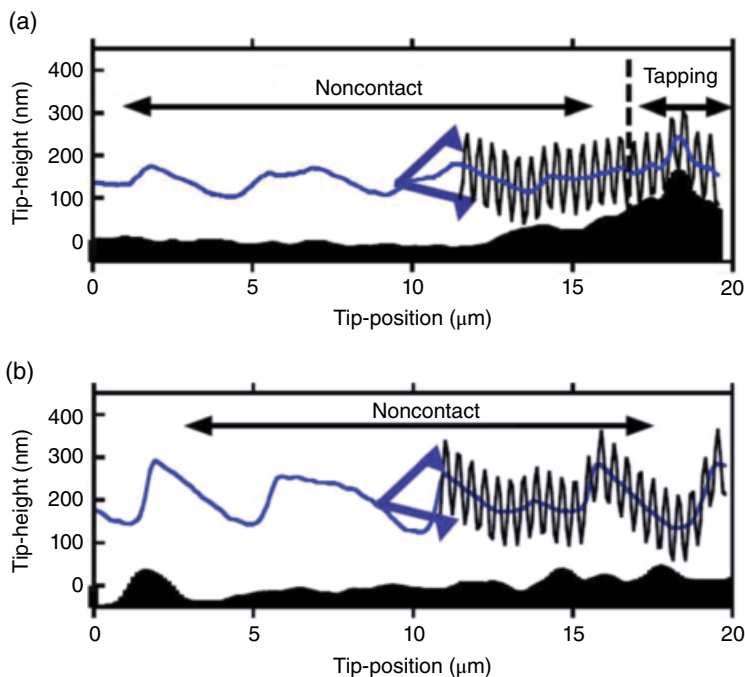


FIGURE 55.11 A comparison between the bimorph-driven system (a) and electrostatic force modulation system (b) during the magnetic force imaging acquisition. The shaded profiles represent the topographic features, whereas the solid lines represent the sectional profiles in the magnetic force images. The oscillatory lines depict the motion of the cantilever with magnetic probes. Source: Reprinted with permission from Ref. 25. Copyright (2012), American Institute Physics.

55.4 CONCLUSION

For improvement of the MFM resolution and study of the relation between topography and magnetic structure, it is important to separate the topographic and magnetic features of the sample. One way to do this is to use modulated electrostatic bias field between the tip and the sample and then the information of modulated electrostatic signal as feedback input.

Both nonmagnetic images again support the separation of the topographic features from the magnetic images through the removal of one stable state using the capacitive coupling. The novel approach presented in this chapter should have a dramatic impact on not only on the MFM field, but also on other scanning probe microscopy fields such as electrostatic force microscopy and Kelvin probe microscopy. Furthermore, the technique is expected to allow for less invasive observation of superconducting vortex structures and soft magnetic structures by avoiding the sudden transition from the noncontact state to the tapping state in the conventional bimorph-driven system, thus

leading to a better understanding of the relationship between magnetic structures and topographic pinning sites such as grain boundaries.

We have developed MFM using an electrostatic force modulation that shows excellent separation of the magnetic signal from the topographic signal. Compared to the bimorph-driven system, the observed magnetic images do not show any topographic features, clearly indicating the separation of topographic and magnetic signals in the noncontact region. We attribute this separation to the opposite feedback polarity in the noncontact region to the one in tapping mode for topographic imaging, thus preventing the magnetic signal from mixing with the topographic signal under feedback condition for the constant amplitude. The origin of the feedback polarity difference is discussed with the electrostatic capacitive coupling and is compared with the bimorph-driven system. This enhanced stability may come from differences between the two systems in terms of the servo forces and modulation method. We attribute the higher stability of MFM using electrostatic force modulation (instead of bimorph-driven modulation) to the long-range electrostatic interaction between the tip and the sample surface.

We also report the successful application of this system for imaging the magnetic structure and topography of CoCr thin film surfaces through separation of the magnetic and topographic structures, with enhanced stability. The system has been tested on the magnetic domain structures of CoCr magnetic film, with a periodicity of 4.5–5 μm . The system has good stability due to the long-range electrostatic capacitance force and also the well-defined single resonance peak in the frequency response. This system will be a promising tool for studying magnetic and topographic structures on the magnetic sample surfaces. In addition to the magnetic imaging, the system can be easily expanded to image the electronic potential structures using ω component of the output signal due to the tip–sample interaction. The simultaneous measurement of magnetic structure and the charge–potential distribution would be especially useful for studying the relationship between magnetic structures and electronic structures in semiconducting magnetic materials for future spintronics applications.

REFERENCES

1. M. S. Valera, A. N. Farley, S. R. Hoon, L. Zhou, S. McVitie, and J. N. Chapman, *Appl. Phys. Lett.* 67, 2566 (1995).
2. G. Bochi, H. J. Hug, D. I. Paul, B. Stiefel, A. Moser, I. Parashikov, H.-J. Güntherodt, and R. C. O’Handley, *Phys. Rev. Lett.* 75, 1839 (1995).
3. M. Lederman, S. Schulz, and M. Ozaki, *Phys. Rev. Lett.* 73, 1986 (1994).
4. R. Proksch, T. E. Schaffer, B. M. Moskowitz, E. D. Dahlberg, D. A. Bazylinski, and R. B. Frankel, *Appl. Phys. Lett.* 66, 2582 (1995).
5. M. Löhndorf, A. Wadas, R. Wiesendanger, and H. W. van Kesteren, *J. Vac. Sci. Technol. B* 14, 1214 (1996).

6. A. Moser, H. J. Hug, I. Parashikov, B. Stiefel, O. Fritz, H. Thomas, A. Baratoff, and H.-J. Güntherodt, *Phys. Rev. Lett.* 74, 1850 (1995).
7. C. W. Yuan, Z. Zheng, A. L. de Lozanne, M. Tortonese, D. A. Rudman, and J. N. Eckstein, *J. Vac. Sci. Technol. B* 14, 1210 (1996).
8. Y. Martin and H. K. Wickramasinghe, *Appl. Phys. Lett.* 50, 1455 (1987).
9. M. R. Freeman and B. C. Choi, *Science* 294, 1484 (2001).
10. U. Kaiser, A. Schwarz, and R. Wiesendanger, *Nature* 446, 522 (2007).
11. C. C. de Souza Silva, A. V. Silhanek, J. Van de Vondel, W. Gillijns, V. Metlushko, B. Ilic, and V. V. Moshchalkov, *Phys. Rev. Lett.* 98, 117005 (2007).
12. T. Shinjo, T. Okuno, R. Hassdorf, K. Shigeto, and T. Ono, *Science* 289, 930 (2000).
13. J. Philip, A. Punnoose, B. I. Kim, K. M. Reddy, S. Layne, J. O. Holmes, B. Satpati, P. R. Leclair, T. S. Santos, and J. S. Moodera, *Nat. Mater.* 5, 298–304 (2006).
14. E. W. J. Straver, J. E. Hoffman, O. M. Auslaender, D. Rugar, and K. A. Moler, *Appl. Phys. Lett.* 93, 172514 (2008).
15. S. Porthun, L. Abelmann, and C. Lodder, *J. Magn. Magn. Mater.* 182, 238 (1998).
16. H.-J. Güntherodt and R. Wiesendanger, eds., *Scanning Tunneling Microscopy I-II* (Springer-Verlag, Berlin, 1992).
17. (a) X. Zhu, P. Grütter, V. Metlushko, and B. Ilic, *Phys. Rev. B* 66, 024423 (2002); (b) X. B. Zhu and P. Grutter, *IEEE Trans. Magn.* 39, 3420 (2003).
18. B. I. Kim, J. W. Hong, J. I. Kye, and Z. G. Khim, *J. Korean. Phys. Soc.* 31, s79 (1997).
19. C. Schönenberger and S. F. Alvarado, *Z. Phys. B* 80, 373 (1990).
20. C. Schönenberger, S. F. Alvarado, S. E. Lambert, and I. L. Sanders, *J. Appl. Phys.* 67, 7278 (1990).
21. R. Giles, J. P. Cleveland, S. Manne, P. K. Hansma, B. Drake, P. Malvald, C. Boles, J. Gurley, and V. Elings, *Appl. Phys. Lett.* 63, 617 (1993).
22. E. Meyer, H. Heinzelmann, P. Grütter, Th. Jung, Th. Weisskopf, H.-R. Hidber, R. Lapka, H. Rudin, and H.-J. Güntherodt, *J. Microsc.* 269, 152 (1988).
23. P. Grütter, H. J. Mamin, and D. Rugar, *Scanning Tunnelling Microscopy II* (Springer, Berlin, 1992), pp. 151–207.
24. B. I. Kim, *Rev. Sci. Instrum.* 80, 023702 (2009).
25. B. I. Kim, *J. Appl. Phys.* 111, 104313 (2012).
26. Bruker AFM Probes, Camarillo, CA, <http://www.brukerafmprobes.com> (accessed December 19, 2015).
27. H. J. Mamin, D. Rugar, J. E. Stern, B. D. Terris, and S. E. Lambert, *Appl. Phys. Lett.* 53, 1563 (1988).
28. D. Sarid, *Scanning Force Microscopy* (Oxford University, New York, 1991), p. 31.
29. R. M. White and T. H. Geballe, *Long Range Order in Solids* (Academic, New York, 1979), Chapter VIII.
30. Q. Zhong, D. Inniss, K. Kjoller, and V. B. Ellings, *Surf. Sci. Lett.* 290, L688 (1993).
31. F. Perez-Murano, G. Abadal, N. Barniol, X. Aymerich, J. Servat, P. Gorostiza, and F. Santz, *J. Appl. Phys.* 78, 6797 (1995).

32. R. Garcia and A. San Paulo, *Phys. Rev. B* 61, R13381 (2000).
33. S. A. Joyce and J. E. Houston, *Rev. Sci. Instrum.* 62, 710 (1991).
34. J. R. Bonander and B. I. Kim, *Appl. Phys. Lett.* 92, 103124 (2008).
35. Y. Martin, D. Rugar, and H. K. Wickramasinghe, *Appl. Phys. Lett.* 52, 244 (1988).
36. B. I. Kim, U. H. Pi, Z. G. Khim, and S. Yoon, *Appl. Phys. A* 66, s95 (1998).
37. B. Anczykowski, D. Kruger, and H. Fuchs, *Phys. Rev. B* 53, 15485 (1996).
38. A. Kuhle, A. H. Sorensen, J. B. Zandbergen, and J. Bohr, *Appl. Phys. A* 66, S329 (1998).
39. B. N. J. Persson, *Sliding Friction: Physical Principles and Applications*, 2nd ed. (Springer, Heidelberg, 2000), pp. 54–77.
40. J. W. Hong, Z. G. Khim, A. S. Hou, and S. I. Park, *Appl. Phys. Lett.* 69, 2831 (1996).
41. H. Yokoyama, T. Inoue, and J. Itoh, *Appl. Phys. Lett.* 65, 3143 (1994).
42. S. Belaidi, P. Girard, and G. Leveque, *J. Appl. Phys.* 81, 1023 (1997).
43. A. S. Hou, Ultrafast electric force microscope for probing integrated circuits, PhD dissertation, Stanford University, Palo Alto, CA (1995).
44. A. Wadas, P. Grütter, and H.-J. Güntherodt, *J. Appl. Phys.* 67, 3462 (1990).

SCANNING TUNNELING MICROSCOPY

KWOK-WAI NG

Department of Physics and Astronomy, University of Kentucky, Lexington, KY, USA

56.1 INTRODUCTION

Scanning tunneling microscope (STM) was first invented by Binnig and Rohrer in the early 1980s to study surface reconstruction on silicon surface [1]. This was the first time that atoms were imaged in real space. Besides its unsurpassed resolution, STM also has the convenience that imaging can be performed at ambient atmosphere or even chemical solution. Comparing to scanning electron microscope, the cost of STM can be significantly lower since there is no expensive electron optics in STM, and high vacuum equipment is not required in many STM applications. We should note that STM is a surface probe and its operation highly depends on the cleanliness and purity of the surface. A heavily oxidized surface will make imaging impossible or produce data that is hard to interpret. For this reason, ultrahigh vacuum environment is quite common for STM—especially as a high-resolution surface study tool. This will unavoidably increase the cost of the setup.

Quantum mechanically it is possible for an electron to coexist in two conductors separated by a small gap. This phenomenon is called quantum tunneling. When a voltage bias is applied between the two conductors, a tunneling current can be established along the direction of the bias. Obviously if we look at the electron as a wave, the separation between the conductors has to be comparable to the lattice parameter for tunneling to occur. From this we can expect the magnitude of the tunneling current depends sensitively on the separation between the conductors. In STM, this tunneling current is measured and used to deduce the distance between a conducting

sharp tip and the conducting sample surface. If we taste the tip on the surface and measure and record the distance at the grid points, a topographical image can then be formed. One can see that the hardware of an STM is extremely simple. It basically comprises of a battery and some sensitive electronics to measure small currents. Most STM nowadays can measure a current as small as a few pA. We can find many cases in the Internet that even a hobbyist can build an STM at home with a very low budget. This performance per price ratio makes STM a very powerful research tool.

STM has one major shortcoming in requiring the sample under study to be conducting or at least semiconducting. To overcome this shortcoming, since the inception of STM in the 1980s, many sensing techniques other than tunneling current have been introduced. The atomic force microscope (AFM) is a good example [2]. In an AFM the deflection of a cantilever is measured either with a laser beam [3] or shift in resonance frequency [4] as the cantilever scans along the surface. Since no current is involved, even the surface of an insulator (I) can be imaged. AFM is now becoming very common and can be found in many laboratories. AFM is most conveniently used to image larger features, and it is more difficult to achieve ultrahigh resolution like the STM. Many other probing techniques like the sensing of magnetic force (magnetic force microscope (MFM)), magnetic field (scanning Hall probe), and electromagnetic radiation (scanning near-field optic microscope (SNOM or NSOM)) have been used to image different surface properties accordingly. Besides the different sensing methods, the control and instrumentation of these microscopes are quite similar. This type of microscopy technique is in general known as scanning probe microscopy. In this review we will focus on the STM.

56.2 THEORY OF OPERATION

In quantum mechanics a particle is described by a wave function $\psi(x)$ satisfying the time-independent Schrödinger equation:

$$-\frac{\hbar^2}{2m} \frac{d^2}{dx^2} \psi(x) + V(x) \psi(x) = E \psi(x)$$

where $V(x)$ is the potential and E is the energy of the particle. The probability of finding the particle in the range $[a, b]$ is given by

$$P(a \leq x \leq b) = \int_a^b \psi^* \psi dx$$

where ψ^* is the complex conjugate of ψ . This probability allows the particle to be found even in the classical forbidden region and leads to the tunneling phenomenon. Consider

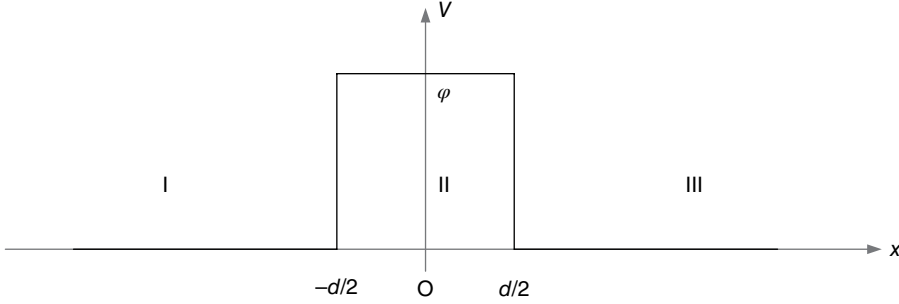


FIGURE 56.1 A square potential area of width d and height ϕ . In the discussion here the energy of the electron is less than ϕ , so the barrier is classically forbidden.

a simple square potential in Figure 56.1; if a particle entering from the left with $E < V_0$, classically, it will be bounced back by the potential barrier.

Quantum mechanically, solving Schrödinger equation for this potential will give the wave function of the particle as

$$\begin{aligned}\psi_I &= e^{ikx} + Re^{-ikx} \\ \psi_{II} &= Ae^{\kappa x} + Be^{-\kappa x} \\ \psi_{III} &= Te^{ikx}\end{aligned}$$

where $k = \frac{\sqrt{2mE}}{\hbar}$ and $\kappa = \frac{\sqrt{2m(\phi - E)}}{\hbar}$

The transmission coefficient is important here, and it can be gotten by the continuity conditions in ψ and $d\psi/dx$ between the wave functions at $x = -d/2$ and $x = d/2$:

$$|T|^2 = \frac{(2k\kappa)^2}{(k^2 - \kappa^2)^2 \sinh^2 \kappa d + (2k\kappa)^2 \cosh^2 \kappa d}$$

The fact that ψ_{III} is not zero means there is a certain probability for the particle to tunnel through the barrier, giving rise to a probability current

$$j = \frac{i\hbar}{2m} \left(\psi \frac{\partial \psi^*}{\partial x} - \psi^* \frac{\partial \psi}{\partial x} \right) = \frac{\hbar k}{m} |T|^2$$

In low tunneling rate limit $\kappa d \gg 1$, $\sinh \kappa d \sim \cosh \kappa d \sim e^{\kappa d}$ and $|T|^2 \sim (2k\kappa / (k^2 + \kappa^2))^2 e^{-2\kappa d}$. While k and κ are mostly constant, $I \propto e^{-2\kappa d}$, the current is decaying exponentially with the barrier thickness d . A slight increase in d will produce a sharp drop in the current. For a reasonable tunneling rate, we require $2\kappa d \sim 1$. Roughly speaking, this means the

barrier thickness d should be the order of the wavelength of the particle given by $\lambda = 2\pi/k$. The wavelength of a 1 eV electron is about 1.23 nm. If the barrier height ϕ is not a constant across the barrier, we can conveniently replace it with the average height because of the exponential dependence:

$$\kappa = \frac{\sqrt{2m(\bar{\phi} - E)}}{\hbar}$$

The preceding discussion can be applied to the real situation when two metals are separated by a vacuum gap (the barrier) of distance z . The barrier height will now become the average work potential of the metals at two sides. All conducting electrons of different energies at both sides will involve in the tunneling process.

For simplicity we will ignore thermal excitation and assume the Fermi–Dirac distribution function is a step function. Only the electrons in the energy range $0 \leq E \leq |eV|$ at the left side in Figure 56.2 can tunnel because only they can find unoccupied states at the opposite side for tunneling to occur. If $\rho_L(E)$ and $\rho_R(E)$ are the electron density of states of the materials at the left- and right-hand sides in the figure, the tunnel current will be given as

$$I \sim \int_0^{eV} e \left(\frac{2k\kappa}{k^2 + \kappa^2} \right)^2 e^{-2\kappa z} \cdot \rho_L(E) \rho_R(E) dE$$

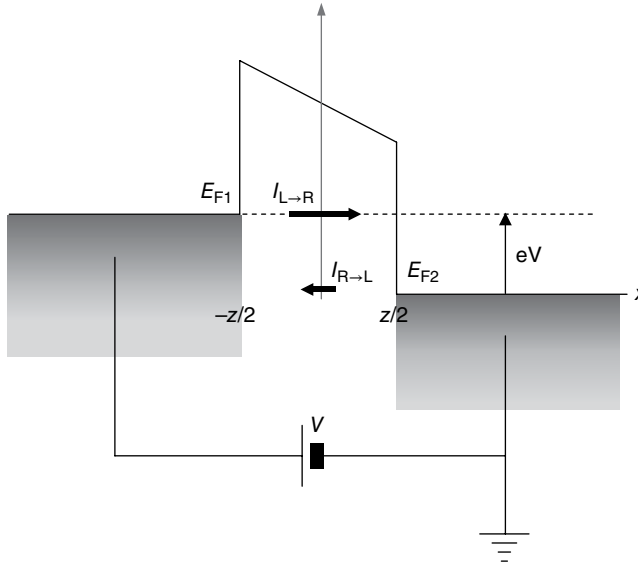


FIGURE 56.2 The Fermi level at one side of the barrier is raised by the applied potential V . This will tip the balance and create a net tunnel current through the barrier.

Note that the current is a convolution of the occupied and unoccupied density of states at both sides. Useful result can be derived if we assume the density of state of one material is more constant than the other, say, $\rho_R(E) = \rho_R(0) = \text{constant}$. The earlier equation can now be reduced to

$$I \sim e\rho_R(0) \int_0^{eV} \left(\frac{2k\kappa}{k^2 + \kappa^2} \right)^2 e^{-2\kappa z} \cdot \rho_L(E) dE$$

Differentiating I with respect to d ,

$$\frac{d}{dz} \ln I = \frac{1}{I} \frac{dI}{dz} = -2\kappa = -\frac{\sqrt{8m(\bar{\phi} - E)}}{\hbar}$$

From this equation, we can see how an STM can be used to determine the average work function of the two metals by a dI/dz measurement, where z is the sample–tip separation.

Besides the work function, STM can also be used to measure the density of state of a material. We should point out that the transmission coefficient is not as well understood and well predicted as described previously. This is especially true in the case of STM, when the wave function is not really a plane wave [5] and it is highly sensitive to the surface condition. For this reason, we will replace the transmission coefficient with an arbitrary function $M(E, z)$:

$$I \sim e\rho_R(0) \int_0^{eV} |M(E, z)|^2 \cdot \rho_L(E) dE$$

and

$$\frac{dI}{dV} \sim e^2 \rho_R(0) |M(eV, z)|^2 \cdot \rho_L(eV)$$

If the variation of $M(E, z)$ is not as strong as $\rho_L(E)$ and it can be approximated as constant in the energy range of measures, then the tunneling conductance dI/dV is proportional to the density of state $\rho_L(eV)$. Tunneling spectroscopy is a powerful technique to measure the electron density of states in a conductor, and it is widely used to measure the energy gap [6] and phonon spectrum [7] in a superconductor. STM can also be used to measure the local density of state, and this type of measurement is known as scanning tunneling microscopy and spectroscopy (STMS). In the STM community, it is a common practice to rid the unknown proportional constant by dividing dI/dV with I/V and quote it as “normalized density of states,” which actually equal to

$$\frac{1}{I/V} \frac{dI}{dV} = V \cdot \frac{e^2 \rho_R(0) |M(eV, z)|^2 \cdot \rho_L(eV)}{e\rho_R(0) \int_0^{eV} |M(E, z)|^2 \cdot \rho_L(E) dE} = eV \cdot \frac{\rho_L(eV)}{\int_0^{eV} \rho_L(E) dE}$$

56.3 MEASUREMENT OF THE TUNNEL CURRENT

The previous probabilistic current can be measured as an electrical current because electron carries charge. This tunnel current depends on the bias voltage and the sample–tip distance. For topographical imaging, the bias voltage should be set as high as possible to increase the tunnel current. However, this bias voltage should be less than the work functions of the sample and tip as field emission will replace quantum tunneling at higher voltages. The work function of most metals is in the order of a few electron volts. For this reason, the bias voltage used for most imaging work is between 0.1 and 1 V or slightly higher. With such a bias voltage, the tunnel current should be in the range of nA (10^{-9} A) to pA (10^{-12} A). A greater current indicates direct touch between the tip and sample. Measuring a smaller current is pushing the limit of the electronics, and noise will eventually become intolerable. It is theoretically correct to use as small a tunnel current as possible. Since the tunnel current depends exponentially on the sample–tip distance, pulling back the tip by merely an angstrom will require a lot of reduction in tunnel current, so we should not compromise too much on the noise level in achieving a small tunnel current.

To measure such a small current, most STM put the first-stage amplifier as close to the junction as possible, in many cases just next to the tip or sample. This first-stage amplifier is actually a current to voltage converter, converting the tunneling current to a few volts at the low-impedance output. This output can then be connected to the main electronics over long cables. Since the circuit is close to the junctions and very likely mounted on the STM body, it has to be simple with not too many components. Many STMs just use an operational amplifier to accomplish this, schematically shown in Figure 56.3. Since the current is less than 1 nA with a bias voltage of 1 V, so the junction resistance is at least 1 G Ω . For this large load resistance, the bias voltage can be easily provided by a constant voltage source. In contrast, the junction resistance of a planar tunnel junction is often less than 10 Ω . Small-resistance junction should be current biased with a constant current source, so STM electronics is not very suitable for spectroscopic measurement of planar junctions. It is easier to ground and guard the tip than the sample, so the tip in Figure 56.3 is virtually grounded, and the bias is applied directly to the sample. If the tunnel current is I , the output of this current to voltage converter will be $-IR$ where R is the circuit feedback resistance of the circuit. In many cases, a logarithmic amplifier is used to remove the exponential dependence of the tunnel current. In constant height mode (Section 56.5 of this chapter), this data will be recorded as the sample to tip distance, and its magnitude will be represented by a gray scale in the topographic image. In constant current mode, the output from the current to voltage converter will be compared to a preset value, and the difference or error will then be proportional, integration, and differentiation (PID) processed, magnified to high voltage and fed back negatively to the z -electrode of the scanner to maintain a constant tunnel current. The voltage to the

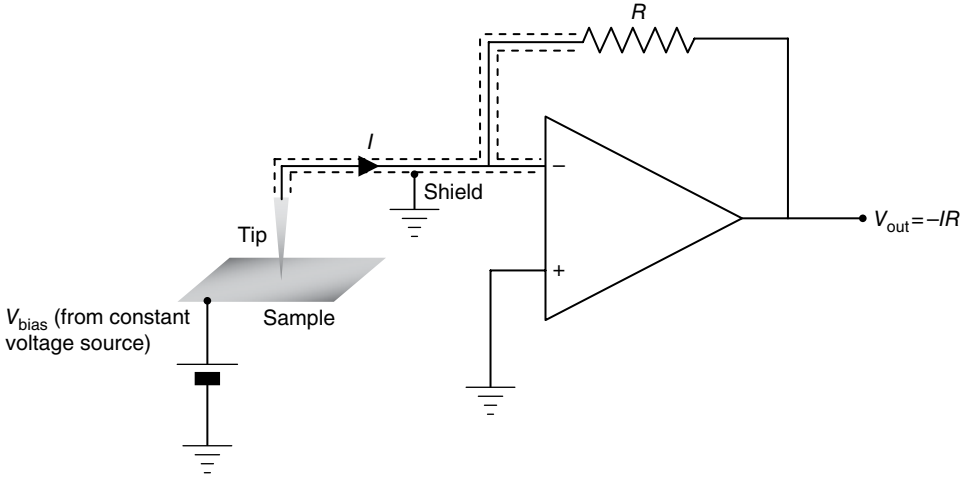


FIGURE 56.3 A current to voltage converter can be used to measure the tunnel current. The sensitivity is determined by the feedback resistance R . The operational amplifier has to be extremely low in input bias current.

z -electrode will be attenuated and recorded as the tip position that represents the topographic height.

In many experiments it is necessary to measure dI/dV for information in density of states. While one can do numerical differentiation on I/V data, it is preferable to measure dI/dV directly with a lock-in amplifier for a better signal-to-noise ratio. To do this a small sinusoidal modulation $\delta V \sin \omega t$ is added to the DC bias applied to the sample; the output of the converter will then equal to $(\delta I \sin \omega t) R$. Commercial STM electronics often provide external modulation input for this purpose. The modulation signal and the current to voltage converter output should be connected to the reference and voltage input of the lock-in amplifier, respectively. The output of the lock-in amplifier $V_{\text{lock-in}}$ is DC and equals to the product of the amplitude of the $\sin \omega t$ component in the input voltage and cosine of the phase difference between the input and reference voltage,

$$V_{\text{lock-in}} = -\delta I R \cos \phi = -\left(\frac{dI}{dV}\right) \delta V \cos \phi$$

so if δV is kept constant, $V_{\text{lock-in}}$ is proportional to dI/dV . The phase of the lock-in amplifier can be adjusted for maximum signal when $\cos \phi = 1$. We now can measure I from the converter output and dI/dV from the lock-in amplifier output simultaneously. Note that for planar tunnel junction, when the junction resistance is small, current source has to be used, and δI will be the control variable, so only dV/dI can be measured in this case (Fig. 56.4).

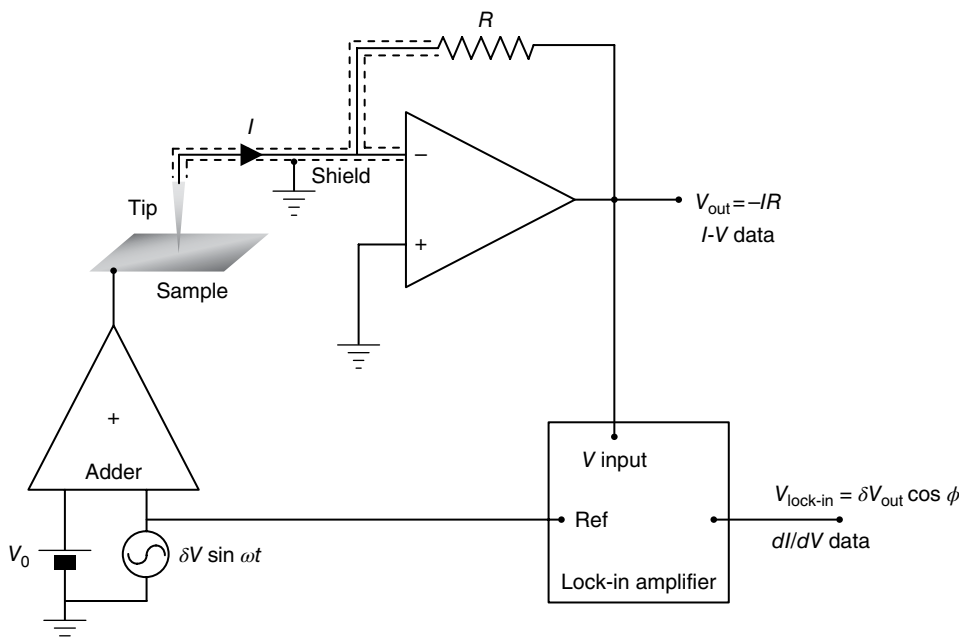


FIGURE 56.4 Lock-in amplifier can be used to measure dI/dV . A constant voltage modulation δV is added to the bias, and the lock-in amplifier is used to measure the resultant current modulation δI , which is proportional to dI/dV .

56.4 THE SCANNER

For atomic imaging, it is necessary to control the tip position to angstrom (10^{-10} m) resolution. The only mechanical method to achieve this is the use of piezoelectric materials to produce the motion. Piezoelectricity is the production of accumulated charges and electric field when the material is under mechanical stress. We are using the reverse of this effect to produce motion by applying an electric field to the piezoelectric material. The electric field is conveniently generated by applying a voltage difference across two electrodes deposited on the surface of the component. The piezoelectric material is lead zirconate titanate (PZT), and there are many types of them like PZT 4, PZT 5A, PZT 5H, PZT 8, etc. [8]. All these have slightly different physical properties, and each STM may choose a type that is most suitable for the experimental conditions it is designed for. The piezoelectric component is poled by applying a high voltage across the electrodes when it is manufactured. This will align the polarization of the domain and induce a permanent dipole in the ceramic. The material can be depolarized when it is heated above the Curie temperature, or too high a voltage is applied in reverse to the poling voltage. Piezoelectricity will be lost after the material is depolarized, and the component has to be replaced or repolarized when this happens.

Applying a voltage with the same polarity as the poling voltage will further align the polarization, and the ceramic will elongate between the electrodes (Fig. 56.5b). In reverse, it will contract if the applied voltage is in reverse to the poling voltage (Fig. 56.5c).

The polarity of a component can be determined by reducing the dipole. The dipole will become smaller when the ceramic is compressed between the electrodes or heat to a slightly higher temperature. A voltage in the same polarity as the poling voltage will be induced at the two electrodes, and the polarity of this voltage can be measured with either a voltmeter or oscilloscope (Fig. 56.6).

The most common shape of piezoelectric material used in STM scanner is probably a cylindrical tube, so we will use this as an example in the discussion here. The two electrodes are on the inner and outer walls of the tube. Let us assume the dipole is pointing radially outward, so when a higher voltage is applied to the outer electrode (opposite polarity to the poling voltage), the element will contract radially and elongate along the length of the tube. Reversing the polarity of the applied voltage will make the tube contract. It is this motion that can be used to scan and move the tip. The capacitance between the two electrodes can be measured as $C = 2K_{33}^T \epsilon_0 \pi L / \ln \left(\frac{OD}{ID} \right)$,

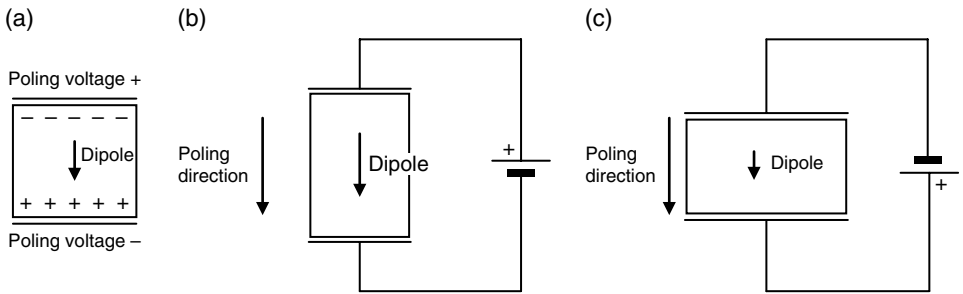


FIGURE 56.5 (a) A piezoelectric material has polarity indicated by the poling direction. In most cases it will (a) elongate when an electric field is applied in the same direction as the poling direction and (b) contract when the electric field is opposite to the poling direction.

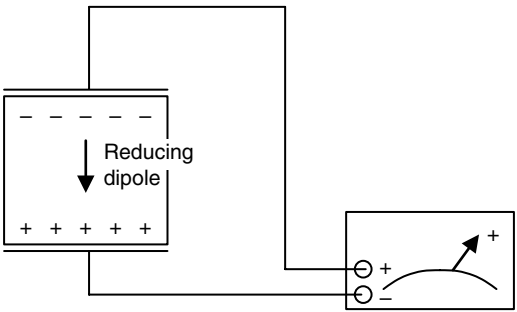


FIGURE 56.6 Determination of the polarity of a piezoelectric component.

TABLE 56.1 d_{31} , d_{33} , and K_{33}^T Values for Different Piezoelectric Materials

	PZT 4	PZT 4D	PZT 5A	PZT 5B	PZT 5J	PZT 5H	PZT 5R	PZT 7A	PZT 7D	PZT 8
d_{31}	-122	-135	-171	-185	-220	-274	-195	-60	-100	-97
d_{33}	285	315	374	405	500	593	450	153	225	225
K_{33}^T	1300	1450	1700	2000	2600	3400	1950	425	1200	1000

d_{31} and d_{33} are in units of 10^{-12} m/V, and K_{33}^T is in 10^{-15} .

where K_{33}^T is the free dielectric constant (given in Table 56.1 for different PZT materials), L is the length of the tube, and OD and ID are the outer and inner diameter, respectively. The capacitance has to be measured at a frequency much lower than the resonance frequency of the tube. In some cases, with some detective work, the type of the ceramic used in an STM can be determined by measuring the capacitance and the K_{33}^T value. Though the capacitance will not tell the condition of the piezoelectric tube, it can be used to confirm the proper connection of the wiring to the electrodes. The extension of the tube depends on the potential difference across the two electrodes as

$$\Delta L = \frac{2d_{31}L}{OD - ID} \Delta V, \text{ where } d_{31} \text{ is the piezoelectric constant with the value of different}$$

materials given in Table 56.1. With the scanner tube we can easily produce motion with angstrom resolution. For example, for a quarter-inch-long PZT 8 tube of 1/8-inch outer diameter and 0.020-inch wall thickness, 1 V can extend or contract the tube by 25 Å in length.

We can now construct the scanner of the STM with piezoelectric tube described in the preceding text. There are two common types of scanner, the tripod and the single-tube scanner. The tripod uses three tubes, one for each axis of motion, connected orthogonally at the tip holder as shown in Figure 56.7a. The (x , y , z) coordinates of the tip will be represented by the voltages applied to the three piezoelectric tubes. Another more compact design is to use a single tube to produce all the motion along the three axes. The outer electrode is sliced into four equal pieces along the length of the tube (Fig. 56.7b). By adjusting the voltage applied to a quadrant electrode, we can just expand or contract that quarter of the tube. This will bend the tube and move the tip in the x - y plane. To move the tip along the tube (i.e., z -) direction, we just need to set the voltage to the inner electrode, which is common to all four quadrants. The inner electrode of some scanners has to be permanently grounded to better shield the tip and the wiring running through it. In this case we can electronically add the z -voltage to the four outside voltages as common. Since the applied voltage cannot exceed the depolarization voltage, this will reduce the dynamic range of the scanner and cause inconvenient complication during operation. Alternatively we can also use a longer tube, with one-half for the z -motion and the other half for the x - and y -motion. This will unavoidably lower the resonant frequency of the scanner.

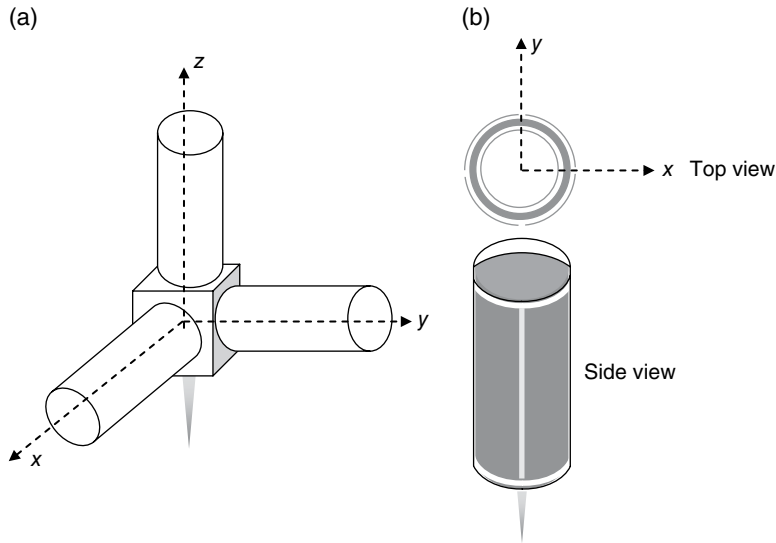


FIGURE 56.7 (a) Tripod design of the tip scanner. (b) A single-tube scanner.

56.5 OPERATING MODE

There are two operating modes of STM. In constant height mode, the tip is scanned with a constant z -voltage, and the tip is maintained at a constant height (Fig. 56.8a), disregarding the topographical variation as it scans along the surface. The tunnel current will be used to represent the topographical height and plotted as a pixel in the 2D image according to a color or gray scale. This requires an atomically flat surface, so the constant height mode is mostly for atomic resolution imaging. For larger scan area, there will be atomic steps and other higher features that the tip may crash into. The STM has to be operated in constant current mode in this case. In constant current mode, the current is compared to a preset value, and the difference is considered as an error between the actual position of the tip and the preset position. Either the positive or negative of this error is added to the z -voltage and negatively fed back to the piezoelectric tube to keep the tip position as close to the preset value as possible. The PID parameters of the feedback loop have to be adjusted to compromise between accuracy, response time, and stability for the best result. The total z -voltage (the offset together with the feedback) will now be used as topographic height to construct the image. The tunneling current should now be roughly constant because of the feedback, but it is advisable to monitor the current map also to ensure the feedback is operating properly.

Since tunnel current is directly measured without any feedback delay in constant height mode, it is easier to obtain crispy high-resolution pictures in constant height mode. However, constant current mode should always be used to scan an unknown area first. The same area can then be imaged with constant current mode after the surface flatness is confirmed.

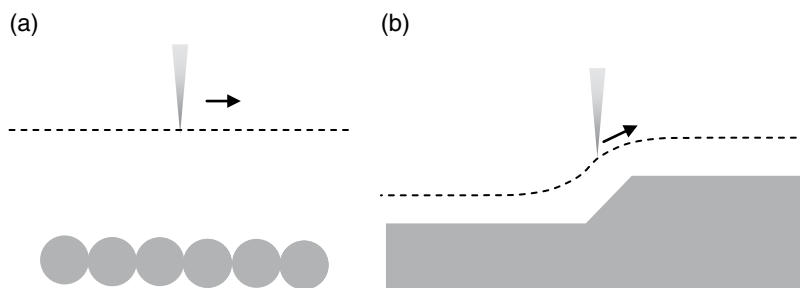


FIGURE 56.8 (a) Constant height mode. (b) Constant current mode.

56.6 COARSE APPROACH MECHANISM

A piezoelectric tube can produce very high-resolution motion, but the range it can scan is very limited. If 1 V can extend the tube by 25 \AA , then it can extend its length at most by 5000 \AA with 200 V. Higher voltage can break down the piezoelectricity and cause permanent damage to the scanner. Until now, STM with long-range capability is still not readily available. The reason is resolution and scan range are two contradictory properties of an STM. To extend the scanning range of an STM from a single-tube scanner, additional mechanism has to be introduced. These extra constructions will unavoidably make the STM more bulky and massive and weaken the rigidity of the structure. This will lower the resonance frequency of the STM and make it more vulnerable to external vibration and eventually reduce the resolution. The short motion of a scanner also poses a problem in positing the sample within its scanning range without crashing the tip to the sample surface. The coarse approach mechanism is one of the most critical components in the construction of an STM, as it will ultimately affect the performance of an STM.

A coarse approach mechanism is some kind of a mechanical device that can move the scanner and the tip over a long distance of several millimeters but with a poor resolution. Though the mechanism has a poor resolution in its motion, it should still be able to move in step significantly smaller than the maximum range of the scanner. With the coarse approach mechanism, the sample can be placed millimeters away from the tip. The scanner will slowly push the tip toward the sample, and tunnel current is monitored at the same time. If there is no tunnel current after the scanner is fully extended, the scanner will retreat to its natural length. Since the sample is still out of the scanning range, it is safe for the coarse approach mechanism to move one step forward. This cycle will be repeated until a tunnel current is established.

It is not difficult to produce fine motion that is good enough for the coarse approach purpose. Earlier STMs often used simple machines like lever, differential springs and screws, gears, or a combination to reduce the motion (Fig. 56.9). The mechanism is most likely driven by a computer-controlled step motor. These simple machines can produce reliable and robust motions, but they are in general massive, bulky, and not

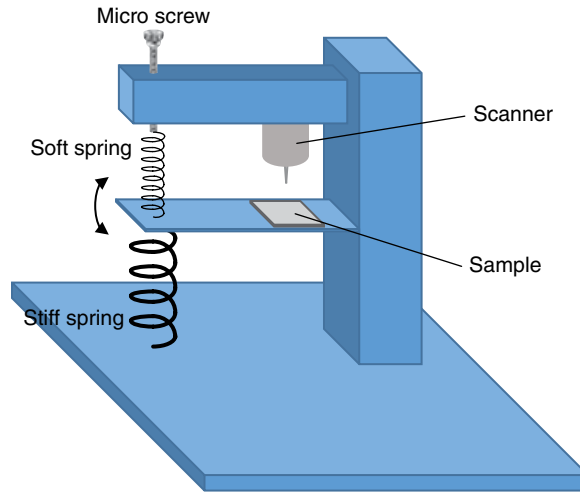


FIGURE 56.9 An example of coarse approach by mechanical components, using the fine motion of a micro screw, and further reduction of motion by differential spring and lever.

very rigid. STMs nowadays mostly employ some kinds of piezoelectric motor for coarse approach. While piezoelectric motors are readily available, most STMs have special designs to integrate the coarse approach motor with the STM structure for the most needed compactness and stability. In here we will discuss some major development in designs over the years as examples to demonstrate the idea.

The first STM developed by Binnig and Rohrer used a device called “louse,” which was a kind of piezoelectric motor already (Fig. 56.10a). The louse is a triangular piezoelectric plate that can expand or contract by applying a voltage across the two sides of the plate. Each vertex is attached to a smooth circular metal foot (MF) coated with a thin I. This structure is placed on three smooth ground plates (GP) of large size. As shown in Figure 56.10, the MFs are grounded to zero volt. When a voltage V_F is applied to a GP, it will be held and locked to the fixed GP by the electric force. Let us label the three vertices as A, B, and C. If we lock down B and C and expand the piezoelectric plate, it can only expand in the vertex A direction. We can then lock down A and unlock B and C before relaxing the piezoelectric plate to its natural size. Repeating this sequence will move the louse with A as spearhead. By the similar method we can choose to move the louse in the direction with any one of the vertices as spearhead. The louse can move to any point on the ground plane, providing a two-dimensional movement. Binnig and Rohrer utilize the louse to transport the sample and use one of the dimensions for coarse approach, as shown in Figure 56.10b. The remaining dimension can be used to move the tip over a long distance over the sample. All piezoelectric coarse approach motors use similar inching mechanism like the louse. The louse uses electric force to cramp one end of the actuator while it is extending or contracting. Another approach is to make use of the static friction, and this type of device is often called inertial motor.

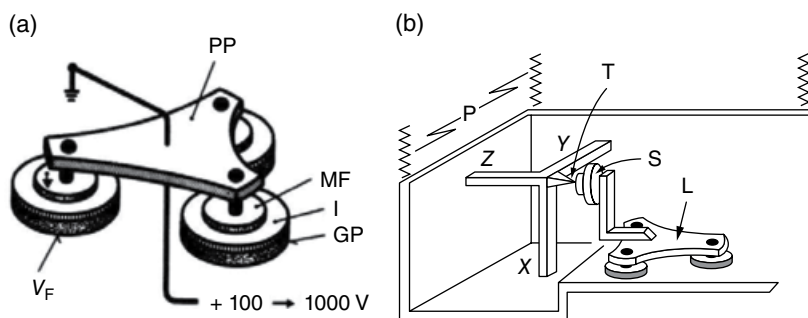


FIGURE 56.10 (a) The louse used by Binnig and Rohrer. GP, ground plate; I, insulator; MF, metal foot; PP, piezoelectric plate; V_F , voltage applied to the ground plate [1]. (b) Louse can be used to coarse approach the sample toward the tip [9]. X, Y, and Z are piezoelectric scanner in tripod configuration, L, louse; S, sample; T, tip. P is springs for vibrational isolation.

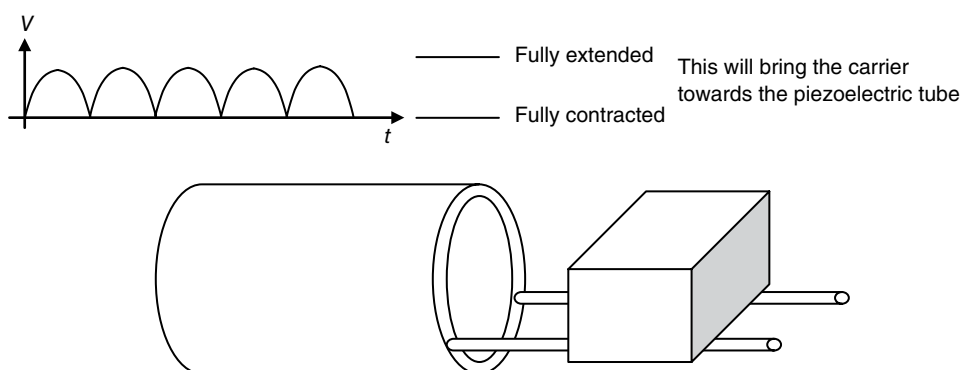


FIGURE 56.11 An example of inertial approach mechanism in which the piezoelectric tube expands and contracts at different rates causing the carrier to slip toward or away from the tube. Top: voltage pattern applied to the piezoelectric tube.

The inertial approach was first introduced by Pohl in 1987 [10]. We will use Figure 56.11 to demonstrate the idea. In this particular example, two rails or the supporter are attached to the end of a piezoelectric tube. The piezoelectric tube has only inner and outer electrodes for extension and contraction. A carrier (mostly for the sample) is then placed on the rails. We need to pay attention when the tube is fully extended and begin to contract, and vice versa, because acceleration is maximum at these extreme points. Suppose we want to move the carrier toward the piezoelectric tube. The acceleration has to be maximized when the tube is fully contracted. Slipping will occur if the acceleration exceeds the limit of the static friction, and the carrier will stay roughly at the same place without following the rail. Now when the rail is fully extended, it has to return and contract with a small acceleration not exceeding the friction limit, and the rail will pull the carrier toward the piezoelectric tube by repeating

this cycle. The upper insert of Figure 56.11 shows the voltage applied to the piezoelectric tube. Note that the acceleration is proportional to dV/dt . Besides the value of the acceleration at the turnaround points, the actual waveform is not that important in most cases, unless in some special situations like vertical or low-temperature applications when the performance becomes very critical. The scanner tube is often installed concentrically inside the piezoelectric tube holding the rails. If both tubes have the same length and are made of the same material, they will compensate each other in thermal expansion and make the STM less vulnerable to thermal fluctuation and variation. It has been demonstrated that springs could be added to hold the carrier in place, and the inertial motor could be installed in the vertical direction [11].

K. Besocke [12] later extended the inertial motor idea to construct a two-dimensional transporter like the louse but without the cramping force. This device is called a beetle for its appearance, and it is still used by many STMs today because of its versatility. A beetle has three legs built of scanner piezoelectric tube (each four outer electrodes and one inner electrode) supporting a platform at the top. Each tube is attached to a ball bearing at the other end for point contact with the floor (Fig. 56.12a). The x - and y -electrodes of these tubes have to be aligned in parallel. By applying x - and y -voltages to a piezoelectric tube and if the acceleration exceeds the friction limit, the ball will slip, and the tube will swing to the designated point on the floor. If all three legs swing with the same displacement, the whole station will transport in the same amount.

The scanner is often installed at the center of the platform, vertically with the tip pointing downward to the sample at the floor. By applying different combinations of voltages to the three piezoelectric tubes, it can produce more complicated motion. For example, if the three legs swing tangentially along the circle joining them, the beetle will rotate around that circle. This provides a method for coarse approach by forcing the beetle to rotate up and down a circular ramp, as shown in Figure 56.12b.

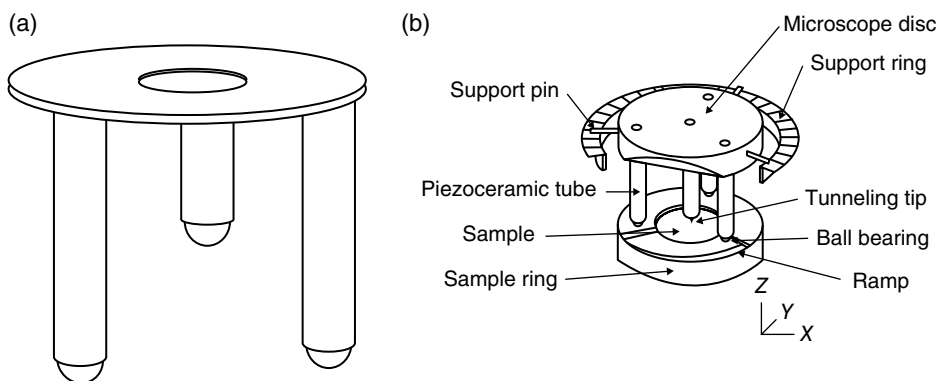


FIGURE 56.12 The beetle. (a) The three legs are piezoelectric tubes like the one used in single-tube scanner. Each leg can swing and slip in any predetermined direction to cause the beetle to rotate or translate to any position. (b) Rotating the beetle on a circular ramp will force it to move up and down, which is useful for coarse approach [13].

While the beetle structure has the advantage of thermal compensation and the possibility of three-dimensional motion, its stability is not outstanding, and it also has to move together with many high-voltage wires attached to the piezoelectric tube. There is a newer design [14] commonly used by homebuilt STM in research laboratories, and it is becoming more popular lately. In this design the motion is produced by shear mode piezoelectric plates. The polarization direction of a shear mode electric plate is within the plane of the plate. A voltage applied across the two sides of the plate will cause them to “slide” on each other and produce the shear motion. Several plates are stacked together to enhance the magnitude of the motion. The STM body has a V-shaped groove along which the carrier will move. Two stacks of piezoelectric plates are glued to each side of the groove. The carrier is a sapphire prism resting on the top of the piezoelectric stacks. The assembly is completed by clamping the prism with another two piezoelectric stacks attached to the top plate, as shown in Figure 56.13a. Figure 56.14 shows the voltage sequence applied to these piezoelectric stacks. The basic idea is to apply a voltage step to a piezoelectric one by one with a time delay, and this will cause the stack to slip on the prism surface. The voltages will then return to zero slowly, and this will cause the piezoelectric plates to relax to the neutral position all together with the prism. Though this motor can only produce one dimension of motion, it can naturally move in a vertical direction, and the rigidity makes it one of the most stable designs. Unlike the beetle, the piezoelectric plates are not part of the

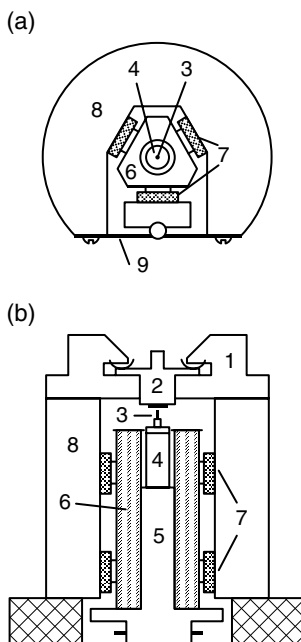


FIGURE 56.13 (a) Top view. (b) Side view. 1, Sample receptacle; 2, sample holder; 3, tip; 4, single-tube scanner; 5, scanner holder; 6, sapphire prism; 7, shear piezo stacks; 8, macor body; 9, spring plate [14].

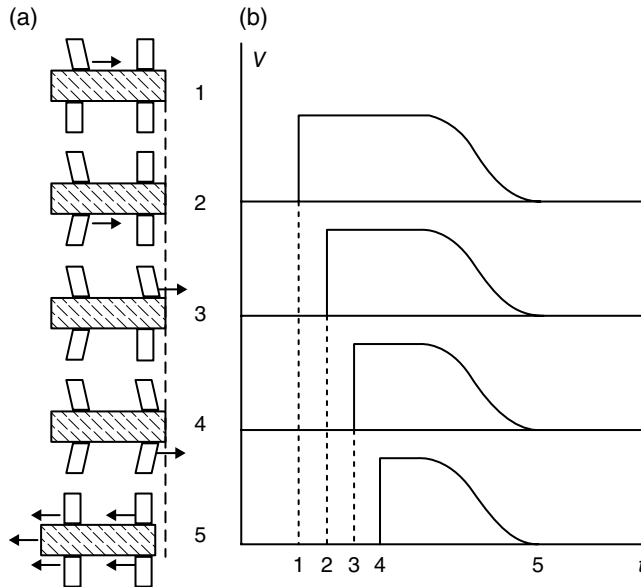


FIGURE 56.14 (a) Positions of the piezoelectric stacks at different times (1–5) when the voltage pattern in (b) is applied to the appropriate stack. Note that for simplicity only four stacks are shown here to demonstrate the idea [14].

carrier, and hence the high-voltage wirings are stationarily attached to the STM body. This reduces the carrier load significantly, and the motion produced by this design should be more reliable.

56.7 SUMMARY

In this chapter we have reviewed the principle of operation of an STM. An STM can be operated in an ambient atmosphere or ultrahigh vacuum, at room temperature or ultralow temperatures. It can provide high-resolution topographical image of conducting surfaces. STM is not just a high-power microscope. Measurement of dI/dV will give information on the electron density of states, and the surface work function can be measured by dI/dz . STM has also been used to manipulate individual atom on metallic surfaces. Other types of scanning probe microscope like the AFM and the SNOM or NSOM are developed based on the idea of STM.

One major advantage of STM is in the simplicity of instrumentation. The highest voltage used in an STM is at most 100–200 V. It does not involve expensive electron optics. The component that gives rise to the high resolution is a small tip that is commercially available at a low cost or can be prepared by simple methods in a lab (not reviewed here). The most critical part in the electronics is probably in the measurement of the extremely small tunnel current in nA to pA range. A preamplifier is often put

next to the tip in converting the tunnel current to a voltage signal, and the requirements on measurement will be minimal after this point. A good STM should be small in size but rigidly built. For this reason the coarse approach mechanism will determine the performance of an STM, and it has to be carefully designed.

REFERENCES

1. G. Binnig and H. Rohrer, "Scanning tunneling microscopy", *Helvetica Physica Acta* 55, 726 (1982).
2. G. Binnig, C. F. Quate, and Ch. Gerber, "Atomic force microscope", *Physical Review Letters* 56, 930 (1986).
3. G. Meyer and N. M. Amer, "Novel optical approach to atomic force microscopy", *Applied Physics Letters* 53, 1045 (1988).
4. F. J. Giessibl, "Advances in atomic force microscopy", *Reviews of Modern Physics* 75, 949 (2003).
5. J. Tersoff and N. D. Lang, "Theory of scanning tunneling microscopy", *Methods of Experimental Physics* (Ed. J. A. Strosio and W. J. Kaiser), Academic Press, San Diego, pp. 1–29 (1993).
6. M. Tinkham, *Introduction to Superconductivity*, 2nd edition, McGraw-Hill, New York, pp. 71–78 (1996).
7. E. L. Wolf and G. B. Arnold, "Proximity electron tunneling spectroscopy", *Physics Reports* 91, 33 (1982).
8. H. Jaffe and D. A. Berlincourt, "Piezoelectric transducer materials", *Proceedings of the IEEE* 53, 1372 (1965).
9. G. Binnig and H. Rohrer, "Scanning tunneling microscopy", *IBM Journal of Research and Development* 30, 355 (1986).
10. D. W. Pohl, "Sawtooth nanometer slider: a versatile low voltage piezoelectric translation device", *Surface Science* 181, 174–175 (1987).
11. Ch. Renner, Ph. Niedermann, A. D. Kent, and O. Fischer, "A vertical piezoelectric inertial slider", *Review of Scientific Instruments* 61, 965 (1990).
12. K. Besocke, "An easily operable scanning tunneling microscope", *Surface Science* 181, 145–153 (1987).
13. J. Frohn, J. F. Wolf, K. Besocke, and M. Teske, "Coarse tip distance adjustment and positioner for a scanning tunneling microscope", *Review of Scientific Instruments* 60, 1200 (1989).
14. S. H. Pan, E. W. Hudson, and J. C. Davis, "³He refrigerator based very low temperature scanning tunneling microscope", *Review of Scientific Instruments* 70, 1459 (1999).

MEASUREMENT OF LIGHT AND COLOR

JOHN D. BULLOUGH

Lighting Research Center, Rensselaer Polytechnic Institute, Troy, NY, USA

57.1 INTRODUCTION

This chapter provides the reader with some of the basic terminology and concepts used in the measurement of light (photometry) and color (colorimetry) pertaining to lighting systems. Also described are some of the types of instrumentation used to make photometric and colorimetric measurements.

57.2 LIGHTING TERMINOLOGY

57.2.1 Fundamental Light and Color Terms

57.2.1.1 Light Light is defined as radiant energy in the electromagnetic spectrum that is capable of producing a visual sensation in humans through stimulation of the retina [1]. Electromagnetic radiation includes gamma rays, X-rays, ultraviolet (UV) energy, infrared (IR) energy, and radio frequencies as illustrated in Figure 57.1. Light comprises of only a small portion of the entire electromagnetic spectrum.

The wavelength band containing light is from approximately 380 to 780 nm ($1 \text{ nm} = 10^{-9} \text{ m}$). Light itself (i.e., rays of light) cannot be perceived directly but must be directed from a luminous surface or reflected from an object toward the eye. When the light reaches the retina (the photosensitive layer in the back of the eye), the photoreceptors in the retina transmit signals to the brain, which are interpreted by the visual centers of the brain as visual information about the luminous environment. Importantly,

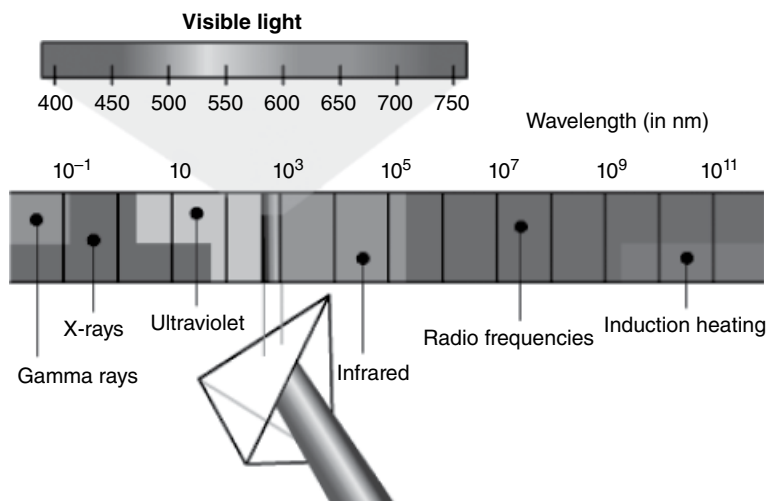


FIGURE 57.1 Electromagnetic spectrum showing the location of visible light. Source: Reproduced with permission of the Lighting Research Center.

this definition of light is made with reference to human visual responses and not the responses of any other organism to radiant energy in the wavelength band defined as light. Some animal species, for example, may be able to respond to UV energy and in some sense could be considered to be light for that species, but there are no formal definitions of light that are not related to human visual responses. Light is the only of the fundamental physical quantities (the others are length, mass, time, electric current, temperature, amount of substance) that depends upon human experience [2].

Collectively, light, in addition to electromagnetic radiation in the bands adjacent to light, UV, and IR radiation, is often referred to as optical radiation because these forms of radiant power can enter and interact with the optical tissues of the human eye (primarily the cornea and lens).

57.2.1.2 Spectral Power Distribution The radiant output of a light source, such as the sun or an electric lamp, can be expressed graphically as a spectral power distribution (SPD) as illustrated in the following. The SPD curve shows the relative distribution of radiant power in the different visible spectral bands (see Fig. 57.2).

Very approximately, the wavelength bands of visible light correspond to different perceived colors as follows:

- 380–430nm: violet
- 430–490nm: blue
- 490–560nm: green
- 560–600nm: yellow
- 600–620nm: orange
- 630–780nm: red

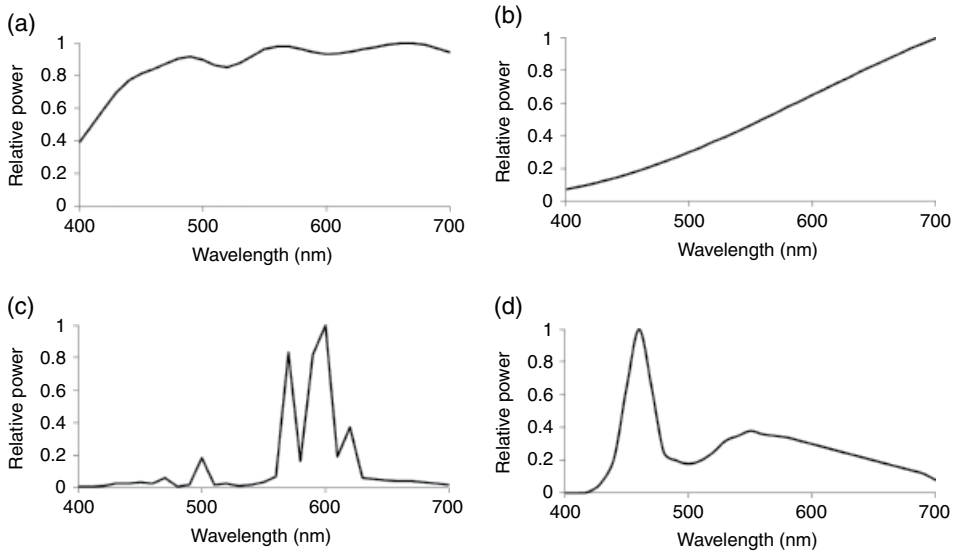


FIGURE 57.2 SPD curves for various light sources. (a) Sunlight, (b) incandescent, (c) high-pressure sodium, and (d) light-emitting diode (LED).

SPD curves like those in Figure 57.2 provide some clues about how different colored objects might appear under a given light source. For example, sunlight (Fig. 57.2a) produces a broad SPD with radiant power across the entire visible light bandwidth, and incandescent light bulbs (Fig. 57.2b) produce more radiant power in the longer visible wavelengths, corresponding to light that is perceived as yellow, orange, and red. This accounts for the warm or yellowish appearance of incandescent bulbs compared to other sources. A high-pressure sodium (HPS) lamp (Fig. 57.2c), commonly used for street and outdoor lighting, produces most of its energy in the portion of the spectrum perceived as yellow and orange (570–620 nm) resulting in the very yellowish appearance of this light source.

57.2.1.3 Correlated Color Temperature As the tungsten filament of an incandescent bulb is heated, it produces light, first a reddish light when the filament is beginning to warm up, then yellowish light until finally it produces its familiar warm white light. Tungsten behaves almost exactly like an ideal blackbody, a material that radiates energy (including light) as a function of its temperature. The higher the temperature, the greater the proportion of short visible wavelengths (bluish light) relative to long visible wavelengths (reddish light) is. The temperature of an ideal blackbody is used as a common metric of the relative coolness or warmth of the appearance of the light. Since no real-world light source is an ideal blackbody, the term correlated color temperature (CCT) is used to define the temperature of an ideal blackbody (in kelvins: K) that most closely matches the color of a light source.

Somewhat confusingly [3], low CCTs that might be considered relatively cool (in terms of temperature) actually produce more yellowish light commonly called warm

light, and high CCTs corresponding to a higher blackbody temperature produce more bluish light commonly called cool light:

- CCTs below 3200 K: warm white light
- CCTs between 3200 and 4000 K: neutral white light
- CCTs above 4000 K: cool white light

57.2.1.4 Color Rendering Index As might be deduced from the SPD curves in Figure 57.2, different light sources will perform differently regarding the way they make objects of different colors look. Under sunlight (Fig. 57.2a), which produces energy across the entire visible spectrum, objects of any colors might be expected to look natural, but a blue car seen under HPS illumination (Fig. 57.2c), which has very little spectral power in the wavelength band corresponding to blue light (~430–490 nm), might not even be seen as blue but rather a distorted color like black or brown, whereas a yellow car under HPS might be expected to look yellow because of the large amount of spectral output in the wavelength range between approximately 560 and 600 nm.

The color rendering index (CRI) is a measure of the color shift of a range of colors that occurs for a given light source compared to the color shift under a comparison, reference source. By convention, the reference source is an ideal blackbody (almost identical to an incandescent bulb’s tungsten filament) when the lamp’s CCT is 5000 K or lower, and the reference source is daylight when the lamp’s CCT is higher than 5000 K. Since incandescent lamps have traditionally been the most common warm white (low CCT) light source experienced in many locations (especially in residences), and daylight was the most common cool white (high CCT) source experienced in the early to middle part of the twentieth century, the CRI is often used as a measure of the naturalness of the appearance of colored objects under a given source. It is expressed as a numerical quantity with a maximum value of 100, representing color appearance identical to that under the reference source. Lower CRI values represent larger color shifts and often less natural appearance (at least, compared to incandescent or daylight illumination). Table 57.1 lists CRI values for several common light sources.

Several alternatives and supplements to CRI are presently under discussion in the lighting industry [4], but at present, CRI is the primary metric used to assess the quality of color rendering of a light source.

TABLE 57.1 CRI Values for Several Common Light Sources

Light Source	CRI
Incandescent lamp	99
4000 K fluorescent lamp	89
High-pressure sodium	22
5000 K light-emitting diode	78

57.2.2 Terms Describing the Amount and Distribution of Light

57.2.2.1 Luminous Intensity Luminous intensity is a measure of the amount of light emitted by a light source in a particular direction, within a particular angular cone from the source. It is measured in units of candelas (cd) and is sometimes referred to as candlepower. As stated in its definition, luminous intensity is direction specific; the luminous intensity of a light source in one direction can, and often will, differ from that in another direction. Luminous intensity is an inherent characteristic of a light source and is not dependent upon the distance from the source. As long as the direction from the light source remains the same, the luminous intensity from the source in that direction will also remain the same. Historically, the candela was defined in terms of the luminous intensity of a standard candle flame, giving the term its name.

57.2.2.2 Luminous Flux Luminous flux is a measure of the total amount of light produced by a light source in all directions around the source. It is measured in lumens (lm). It is the quantity seen on lamp packages to express the total light output of the lamp. Geometrically, the lumen is the amount of luminous flux produced by a point source with a uniform luminous intensity of 1 cd, within an angular cone having a solid angle of 1 steradian. A steradian (Fig. 57.3) is the solid angle subtended by a cone that, when projected onto a sphere, has an area equal to the square of the sphere's radius. Therefore, 1 cd represents 1 lm/steradian (in a given direction).

To determine luminous flux (lumen) ratings of light sources, the conditions under which the light source is measured must be carefully controlled, including the ambient temperature, the orientation of the lamp, the input voltage and current, and vibration, because most light sources are sensitive to these conditions [1].

57.2.2.3 Luminous Intensity Distribution Luminous flux is a useful quantity to help understand how much light a given source produces when the specific direction of the light is not important, such as for general room lighting, but it is less useful for directional lights such as flashlights, spotlights, vehicle headlights, or display lighting, where the lighting system must produce a narrower distribution of light. Two

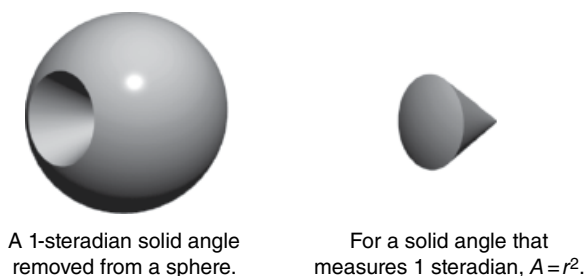


FIGURE 57.3 Graphical representation of a steradian. Source: Reproduced with permission of the Lighting Research Center.

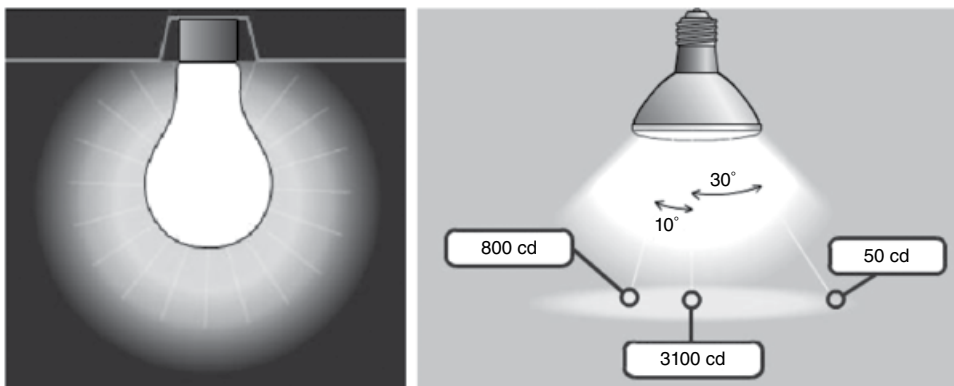


FIGURE 57.4 The general service incandescent lamp shown in the left panel may produce a uniform luminous intensity of 200 cd in most directions. The spot lamp shown in the right panel can produce an equivalent luminous flux but would have much higher intensity in one particular direction and much lower intensity in other directions. Source: Reproduced with permission of the Lighting Research Center.

incandescent lamps, a general service light bulb and a spot lamp (Fig. 57.4), with the same wattage, will have similar lumen outputs, but the spot lamp will be more useful as a headlight or flashlight because it has a high intensity in a particular direction and relatively low intensity elsewhere, whereas the general service bulb will have a modest intensity in nearly every direction from the bulb.

The luminous intensity distribution from a light source can be represented on a polar coordinate graph (Fig. 57.5) that shows the luminous intensity from the light source as a function of the angular direction from the front of the source, represented by a polar angle of 0°.

The luminous intensity distribution can be represented graphically as illustrated in Figure 57.5 or in tabular form for different angles from the light source.

When a light source or system would be expected to have different distributions in different directions, two or more luminous intensity distribution curves can be shown. For many fluorescent lighting systems where the light source is a linear tube, the distribution along the length of the lamp would differ substantially from that across the lamp. In this case, two luminous intensity distributions (Fig. 57.6) would be used to illustrate the distribution of light from such a lighting system.

Sometimes the luminous intensity distribution is graphed using rectangular rather than polar coordinates; this is common for many LED sources.

57.2.3 Terms Describing Lighting Technologies and Performance

57.2.3.1 Luminaire A luminaire, commonly called a light fixture, is a complete lighting unit consisting of a lamp or lamps, a ballast or driver (if needed to operate the lamp), and parts designed to position and protect the lamps, to connect the lamps or

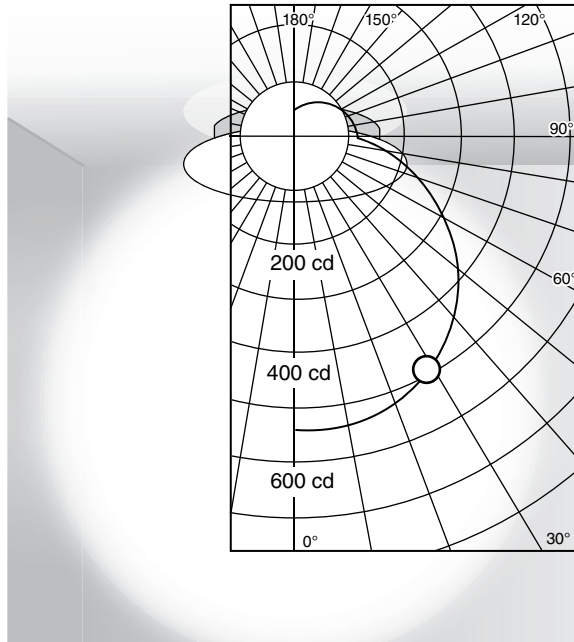


FIGURE 57.5 A luminous intensity distribution of a ceiling luminaire plotted on a polar coordinate graph; the luminous intensity is approximately 550 cd at 0° (directly below the source) and is 500 cd at 30°. Source: Reproduced with permission of the Lighting Research Center.

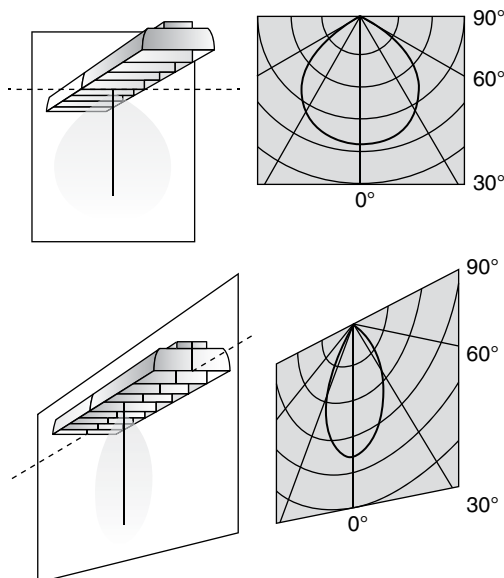


FIGURE 57.6 Top: luminous intensity distribution for a fluorescent luminaire, in the plane across the lamp. Bottom: luminous intensity distribution for the same luminaire, in the plane along the length of the lamp. Source: Reproduced with permission of the Lighting Research Center.

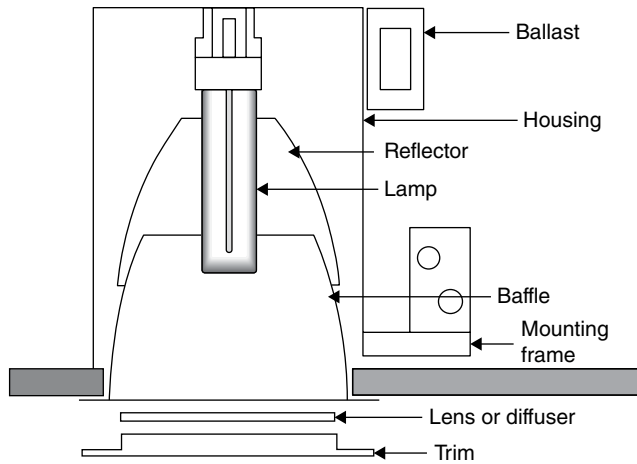


FIGURE 57.7 Cutaway diagram of a compact fluorescent downlight luminaire, showing the individual parts. Source: Reproduced with permission of the Lighting Research Center.

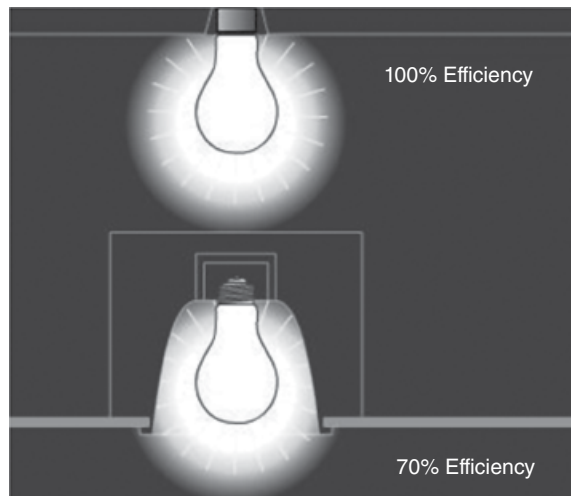


FIGURE 57.8 Top: a bare lamp without a luminaire emits all of its lumens. Bottom: only a percentage of the lumens emitted by the lamp will exit a luminaire; this percentage is the luminaire's efficiency. Source: Reproduced with permission of the Lighting Research Center.

ballasts to the power supply, and to direct the light. Light-directing components may be reflectors, diffusers, baffles, or lenses. An example of a luminaire is the downlight illustrated in Figure 57.7.

57.2.3.2 Luminaire Efficiency The efficiency of a luminaire is defined as the ratio (in percent) of the luminous flux emitted by a luminaire to that emitted by the lamp or lamps within the luminaire. It is a percentage of the lamp lumens that are ultimately emitted by the entire luminaire (see Fig. 57.8).

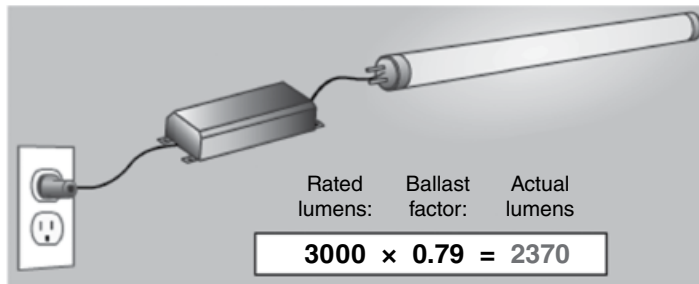


FIGURE 57.9 A fluorescent lamp operated on a reference circuit might produce 3000lm but when operated on one specific ballast might produce 2370lm, a ballast factor (BF) of 0.79. Source: Reproduced with permission of the Lighting Research Center.

57.2.3.3 Ballast Factor For a luminaire that contains a ballast to operate the lamp (such as those using fluorescent or high-intensity discharge (HID) lamps), the ballast uses some of the power necessary to operate the luminaire. Different ballasts will use different amounts of power, depending upon the type of functionality they provide (e.g., dimming, cold temperature operation, etc.). When determining the luminous flux produced by a lamp, a specific reference ballast circuit for each type of lamp is specified [1] to ensure repeatable and consistent results, but the ballast in a particular luminaire is likely to differ from this specific type of reference circuit. The ratio between the luminous flux produced by a lamp using a particular ballast and that produced when using the reference ballast circuit is defined as the ballast factor (commonly abbreviated BF; see Fig. 57.9).

The BF value allows a lighting specifier to predict the lamp lumens in a given luminaire, relative to the rated lumen value for the lamp(s) provided by the lamp manufacturer for the reference ballast circuit. Most of the time, but not always, BF values are less than 1.0.

57.2.3.4 Rated Life For most conventional light sources, the rated life is defined as the amount of time a large group of lamps would be operated (usually, in hours), before half of the lamps in the group would be expected to have failed or burned out (Fig. 57.10). Lamp life is assessed under specific conditions (e.g., temperature, lamp orientation, voltage) and using specific operating cycles [1]. For example, the life of an incandescent lamp is determined by operating the lamp continuously until failure. For fluorescent lamps, where failure of the electrodes in the lamp is a common failure mechanism, lamps are operated on a constant cycle of 3 h on, followed by 20 min off, continuously until failure. For HID lamps, the lamps are operated for 11 h followed by 1 h off until failure.

Because burnout is not a common mechanism for the failure of light-emitting diode (LED) sources, the lighting industry has worked to develop more practical and meaningful measures of useful life for these sources. A common approach [5] is to specify

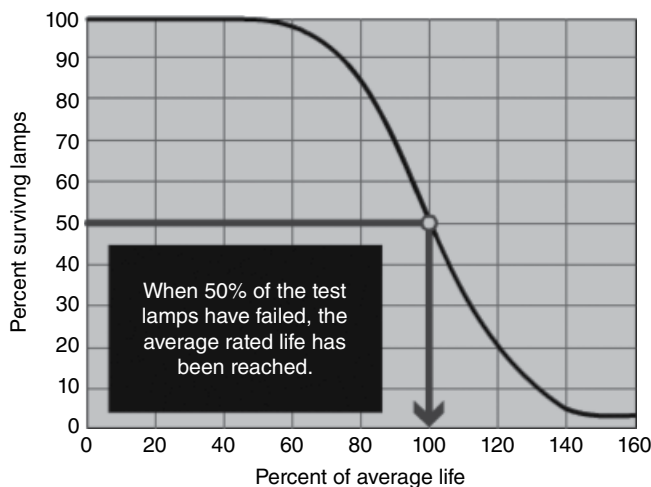


FIGURE 57.10 Rated lamp life is usually defined by the operating hours at which 50% of the lamps in a sample have failed. Source: Reproduced with permission of the Lighting Research Center.

the operating time (in hours) to reach a reduction in light output (such as to 50 or 70%) from the initial light output.

57.2.3.5 Luminous Efficacy (Electrical) Luminous efficacy is defined as the quotient of the total luminous flux (in lm) produced by a lamp or luminaire, by the total electrical power input (W) of the lamp or luminaire. For luminaires, the luminous flux should be modified by the BF, and the power should include power used by both lamp(s) and ballast (if any). Luminous efficacy is expressed in units of lumens per watt (lm/W).

Sometimes the stated luminous efficacy of LED sources is the optical luminous efficacy, given as the lumens per watt of radiant power produced by the source, not the watts used to provide power to the source. Despite having identical units, optical luminous efficacy and electrical luminous efficacy values are different and cannot be compared in a meaningful way.

57.2.4 Common Quantities Used in Lighting Specification

57.2.4.1 Illuminance

The definition of illuminance is the density of luminous flux incident on a surface. It is most commonly expressed in units of lux (lx) or footcandles (fc). 1 lux is equivalent to 1 lumen per square meter (lm/m^2), and 1 fc is equivalent to 1 lumen per square foot (lm/ft^2). If a point source of light (Fig. 57.11) with a uniform luminous intensity of 1 cd were surrounded by a sphere with a radius of 1 m, the illuminance on the interior surface of the sphere would be 1 lx. If the same light source were surrounded by a sphere

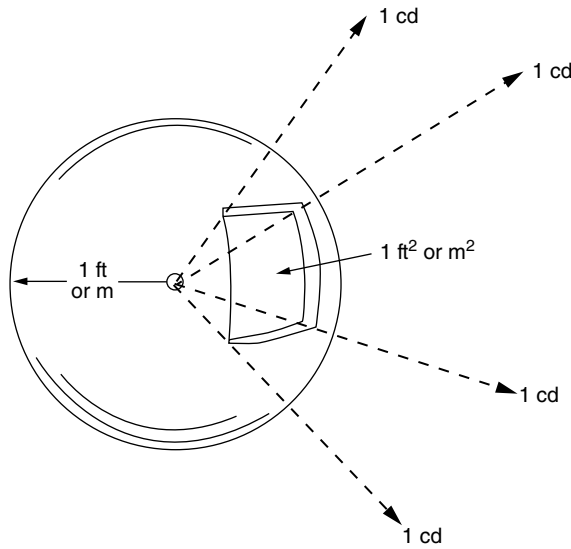


FIGURE 57.11 A point source in the center of the sphere with a uniform luminous intensity of 1 cd would produce an illuminance of 1 lx on if the sphere's radius were 1 m, and 1 fc if the sphere's radius were 1 ft.

with a radius of 1 ft, the illuminance on the interior surface of the sphere would be 1 ft. Because there are 10.76 ft² in 1 m², 1 fc is equal to 10.76 lx. Commonly, a rounded factor of 10 is used to relate fc to lx (i.e., 30 fc is \cong 300 lx).

Unlike luminous intensity and luminous flux, which are properties of the light source, illuminance is dependent upon the geometry between the light source and the surface being illuminated. For example, the 1-cd source shown in Figure 57.11 produces an illuminance of 1 lx on a surface that is 1 m away but produces only 0.25 lx on a surface that is 2 m away. The relationship between the luminous intensity of a light source and the illuminance it produces at a given distance follows the inverse-square law, where the illuminance is inversely proportional to the square of the distance. The inverse-square law is commonly written as follows:

$$E = \frac{I}{d^2} \quad (57.1)$$

where E is the illuminance in lx (or fc),

I is the luminous intensity from the source in the specific direction in cd,

and d is the distance between the light source and the surface being illuminated, in meters (or feet).

When d in Equation 57.1 is in meters, E is in lx. When d is in feet, E is in fc.

Because illuminance is a representation of the amount of light falling on a surface such as a countertop, a desk, a chalkboard, or a piece of machinery, it is the most



FIGURE 57.12 Distribution of selected illuminances on various surfaces within a space. Source: Reproduced with permission of the Lighting Research Center.

common quantity used in the specification of recommended light levels for many different applications in buildings and outdoors [1]. In offices, for example, the illuminance on desks is the common light level specification; in parking lots, the illuminance on the pavement surface is the required light level. However, illuminances (Fig. 57.12) can be horizontal (as on desks or pavement), vertical (as on paintings hung on walls or a person's face in front of a bathroom mirror), or in any other plane (such as an inclined instrument panel on a factory machine).

57.2.4.2 Luminance Luminance is defined as the amount of light directed (or reflected) from a surface, in a particular direction and within a particular solid angular cone. Luminance is most commonly expressed in units of candelas per square meter (cd/m^2), sometimes called nits. More practically, luminance represents a quantity that is somewhat analogous to the brightness of a surface. The luminance of a surface is specific to the direction from which the surface is viewed (Fig. 57.13) but does not change as the distance from the surface is changed. For example, if the luminance of a vertical wall at the end of a corridor is $100\text{cd}/\text{m}^2$ when viewed from the opposite end of the corridor, this luminance would not change as an observer moved along the corridor toward the wall because the direction of view from the wall to the observer would not change.

57.2.4.3 Reflectance Consider a black desk located underneath a luminaire that produces an illuminance of 300lx on the desktop and a white piece of paper sitting on the desk. Both the black desk and the white paper would have the same illuminance (300lx) incident upon them, but the paper would appear substantially brighter than the desk. This is because the reflectance of the paper is much higher than that of the desk. The reflectance is defined as the ratio of the luminous flux incident on a surface to the

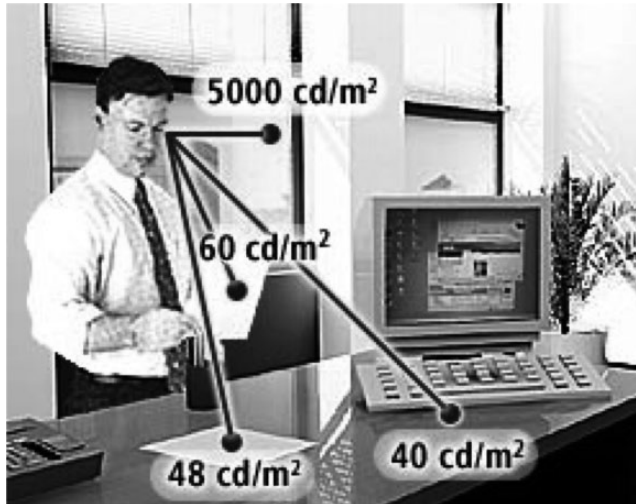


FIGURE 57.13 Surface luminances are expressed with respect to a particular direction, in the previous illustration, toward the observer's eyes. Source: Reproduced with permission of the Lighting Research Center.

amount of flux that is reflected, from the surface. Reflectances are expressed as unitless quantities from 0 (a perfectly black surface that reflects no light) to 1 (a perfectly white source that reflects all of the light that reaches it). Sometimes reflectances are given in terms of percentages from 0 to 100%. Some typical room and surface reflectances are as follows:

- White painted ceiling: 0.8
- Light finished/painted walls: 0.5
- Floors: 0.2
- White office paper: 0.8
- Asphalt pavement: 0.1

The luminance of a matte (diffuse) surface can be estimated if its reflectance and the illuminance falling on the surface are known, using Equation 57.2:

$$L = \frac{E\rho}{\pi} \quad (57.2)$$

where L is the luminance (in cd/m^2),
 E is the illuminance on the surface (in lx),
 ρ is the reflectance of the surface,
 and $\pi \approx 3.14$.

For matte surfaces, Equation 57.2 predicts the luminance in all viewing directions. For glossy or semiglossy surfaces, the relationship among luminance, illuminance, and reflectance is very complex and dependent upon the specific geometry among the light source, the surface, and the observer. Lighting calculations often make the simplifying assumption that all surfaces within an illuminated space are matte, so that their luminances can be predicted by Equation 57.2. However, it should be recognized that for highly shiny or “specular” surfaces like polished metal or glass, luminances cannot be readily estimated by this simple calculation.

57.3 BASIC PRINCIPLES OF PHOTOMETRY AND COLORIMETRY

57.3.1 Photometry

Photometry is a simple, mathematically precise system of measuring and specifying light agreed to by an international community involved with its commerce and specification. It is the basis for the illuminance and luminance quantities described in earlier sections of this chapter. In this section, the relationship between a light source’s SPD, its luminous flux, and the luminous efficiency functions used to define light is described.

57.3.1.1 Luminous Efficiency Function The luminous efficiency function is used to relate the relative effectiveness of radiant power along the visible spectrum (between about 380 and 780 nm) at creating a visual sensation, since equal amounts of radiant power at different wavelengths will not necessarily produce equal visual sensations. Several luminous efficiency functions have been developed [6], but the most common is the photopic luminous efficiency function [2], often denoted $V(\lambda)$, where V stands for visibility and λ represents the wavelength. The peak value of the photopic luminous efficiency function is at 555 nm (Fig. 57.14), where the luminous efficiency is defined as 1.0. The photopic luminous efficiency function represents the combined spectral sensitivity of the cone photoreceptors in the central portion of the human retina, and the peak spectral sensitivity of the combination of cones in this portion of the retina is 555 nm. There is also a scotopic luminous efficiency function, denoted $V'(\lambda)$, representing the spectral sensitivity of the human retina’s rod photoreceptors, which have a peak spectral sensitivity at 507 nm rather than 555 nm (Fig. 57.14).

All luminous efficiency functions are expressed in unitless quantities between 0 and 1, with the wavelength having the greatest visual sensation defined to have a value of 1 and all other wavelengths having lower values. In practice, only the photopic luminous efficiency function is used to characterize light in almost every situation [1]. This is because light levels must be very low in order to be in a state where only rod photoreceptors contribute to human vision [6]. The presence of nearly any electric light source will place an observer above the scotopic luminance range. Most interior

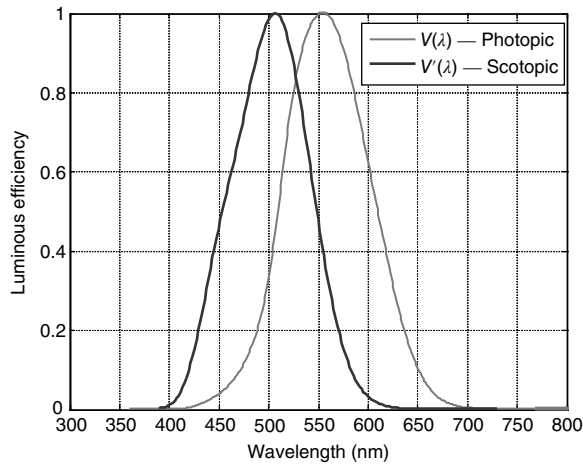


FIGURE 57.14 Photopic (right, gray) and scotopic (left, black) luminous efficiency functions. Source: Reproduced with permission of the Lighting Research Center.

lighting applications are of sufficient intensity to place the observer in the photopic luminance range, where cone photoreceptors contribute to vision. There is also a range of lighting applications, primarily outdoor nighttime applications, where a mixture of rods and cones contribute to vision (denoted the mesopic luminance range), but there is no single luminous efficiency function that can characterize mesopic luminous efficiency [7], and this special set of conditions is not discussed further in this chapter.

57.3.1.2 Calculating Luminous Flux In order to calculate the luminous flux from the SPD of a light source, several steps must be undertaken. The SPD should be expressed in terms of the radiant power (in W) produced by the light source at each wavelength in the visible spectrum, and the power at each wavelength is multiplied by the value of the luminous efficiency function at each wavelength (Fig. 57.15), integrated across the wavelength limits of the visual spectrum (e.g., 380–780 nm).

Mathematically, the calculation can be expressed as the following integral equation:

$$\Phi = k \int_{380 \text{ nm}}^{830 \text{ nm}} P(\lambda) \cdot V(\lambda) \cdot d\lambda \quad (57.3)$$

where Φ is the luminous flux (in lm),

k is a constant equal to 683 lm/W,

$P(\lambda)$ is the light source radiant power at wavelength λ ,

and $V(\lambda)$ is the photopic luminous efficiency function.

An important step in this process is the inclusion of the constant k in the aforementioned equation, which allows the conversion of specific amounts of radiant power

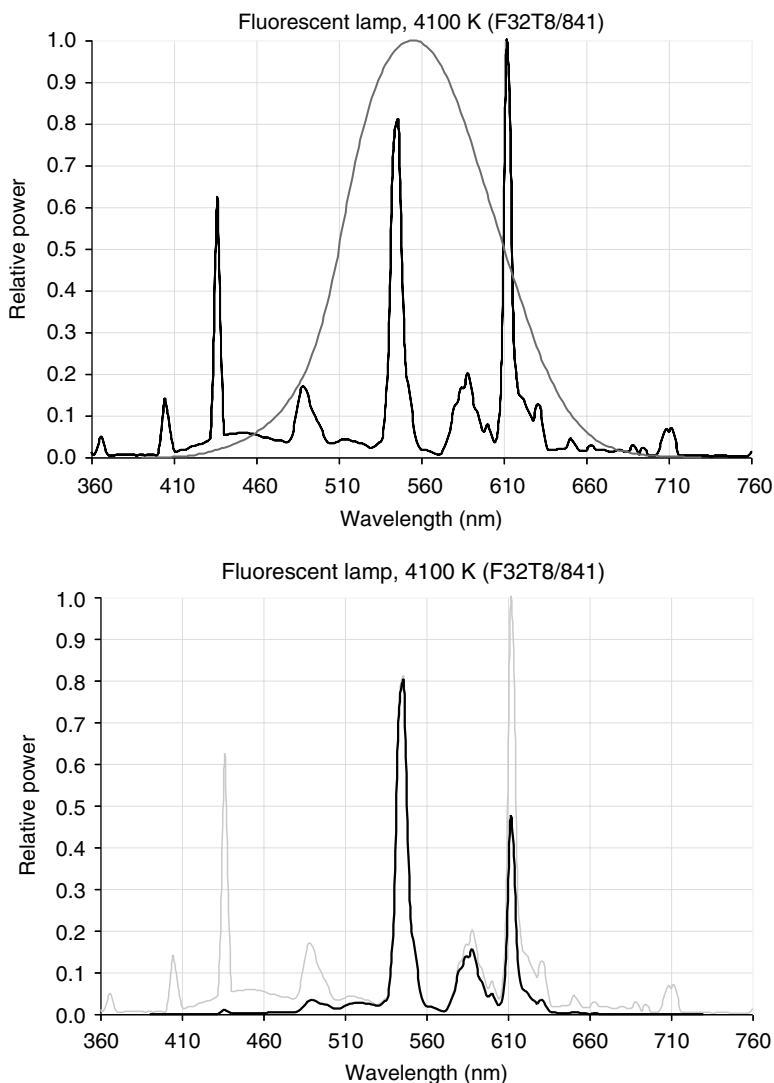


FIGURE 57.15 The light source radiant power and the luminous efficiency at each wavelength are shown at top. The products of the radiant power and luminous efficiency values at each wavelength give the quantities shown at bottom. Source: Reproduced with permission of the Lighting Research Center.

(in W) to specific amounts of luminous flux (in lm). By international convention, a light source that produces exactly 1 W of radiant power at 555 nm, the peak of the photopic luminous efficacy function, is defined to produce 683 lm, for a maximum photopic luminous efficacy of 683 lm/W. At different wavelengths, 1 W of radiant power would produce fewer than 683 lm. As an example, Figure 57.16 shows, for LED light sources having different peak wavelengths, how much relative radiant power is required

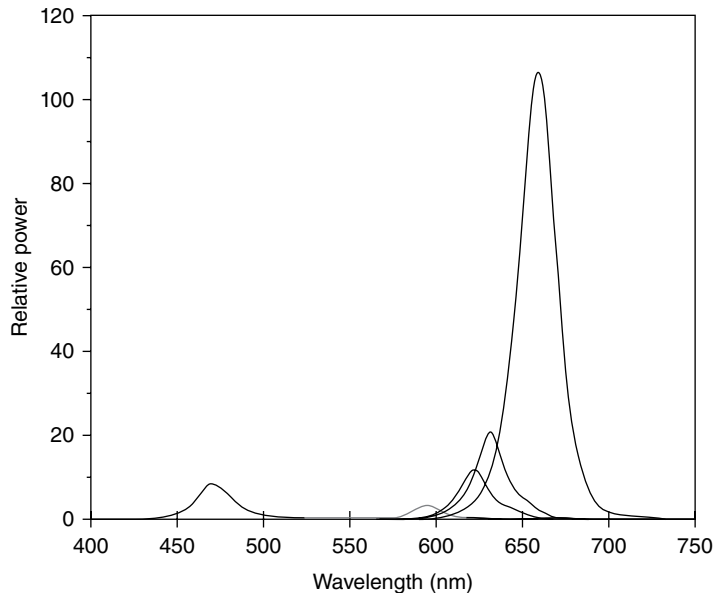


FIGURE 57.16 Relative radiant power needed for LED light sources with different peak wavelengths, in order to produce equivalent amounts of luminous flux.

to produce equivalent amounts of luminous flux. The further the spectral emission is from 555 nm, the more radiant power is required to achieve the same luminous flux.

Importantly, the value of luminous efficacy at 555 nm of 683 lm/W applies to both the photopic and scotopic luminous efficacy functions. Since the scotopic luminous efficiency function has a peak value not at 555 nm but at 507 nm, the maximum scotopic luminous efficacy is not 683 lm/W. Since the value of scotopic luminous efficiency at 555 nm is 0.402, the scotopic luminous efficacy at 507 nm, where the scotopic luminous efficiency is 1, is actually $683 \text{ lm/W} \div 0.402$, or 1700 lm/W. It is important to recognize that this difference in maximum luminous efficacies between the photopic and scotopic functions is merely a mathematical artifact and has no basis in physiology. Because the maximum scotopic luminous efficacy is 1700 lm/W, the value of k in the aforementioned equations should be 1700 lm/W, and not 683 lm/W, if a scotopic luminous flux quantity is to be calculated.

57.3.1.3 Photometric Measurement Spectral Considerations Photometric quantities such as illuminance and luminance are all related to luminous flux by geometrical relationships, such as the luminous flux density (illuminance), the luminous flux per solid angle (luminous intensity), and the density of luminous flux per solid angle (luminance). Most commonly used photometric instruments for measuring illuminance and luminance do not contain individual spectrally tuned elements at each wavelength but rather employ a material with a broad spectral response, such as selenium or silicon [1]. Silicon, for example, responds to radiant power in the UV and IR bands (Fig. 57.17)

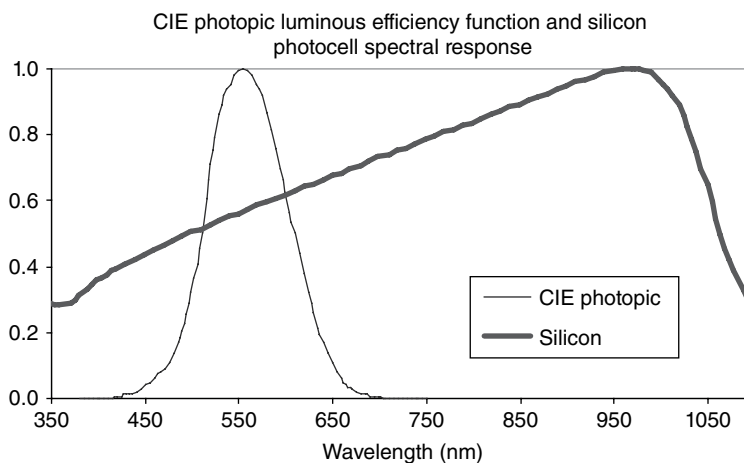


FIGURE 57.17 Spectral response of silicon (heavy curve) shown alongside the photopic luminous efficiency function commonly used to characterize light (lighter curve). Source: Reproduced with permission of the Lighting Research Center.

so a simple silicon cell alone would not be a suitable detector for a photometric instrument, since it would register a response to UV or IR radiation that would not be detected by the human eye as light.

Instead, such materials are fitted with filters that approximate the spectral response of the luminous efficiency function. No filter or combination of filters can provide a perfect match, but for broadband light sources producing “white” light, the mismatches largely cancel each other out with a reasonable estimate of the luminous quantity being measured. However, the user of such photometric instruments should be aware that they can yield quite large errors (Fig. 57.18) when measuring narrow-band (colored) sources near the extremes of the visible spectral range, such as blue LED sources.

57.3.1.4 Luminous Intensity Measurement Geometrical Considerations When measuring the luminous intensity distribution from a light source such as an LED, one could in principle measure the illuminance from the source at various angular directions and use the inverse-square law to calculate the luminous intensity using Equation 57.1. However, it is necessary to remember that the inverse-square law is strictly applicable only to pure point sources with infinitesimally small sizes. Since LEDs and all other light sources have a finite size, several constraints should be met in order to ensure reasonable accuracy. One is the so-called five-times rule [1], which states that the measurement distance should be no less than five times the maximum dimension of the light source. It can be applied to diverging sources of light (without imaging optics such as lenses) and for diffuse emitters ensures that the error in estimating luminous intensity from the illuminance should be less than 0.5%.

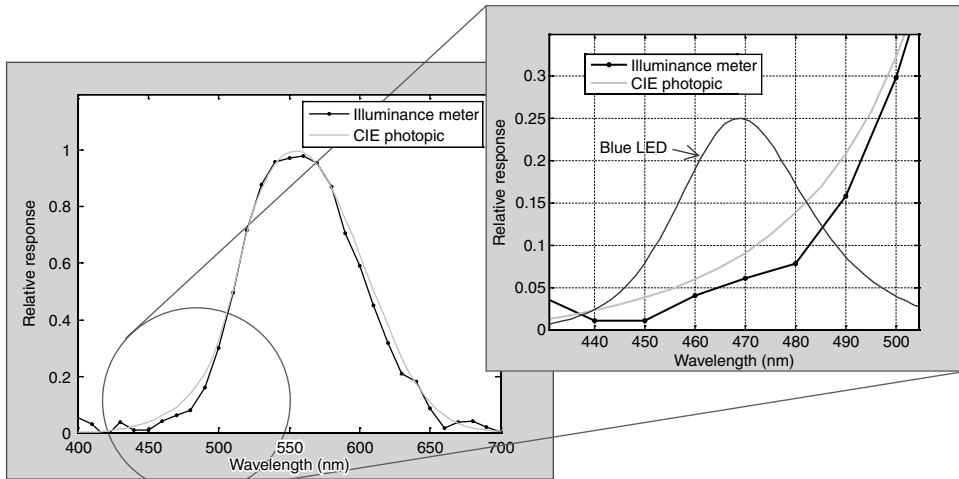


FIGURE 57.18 Left panel: a filter can provide a reasonably close match to $V(\lambda)$ for broad-band light sources. Right panel: when used to measure a narrowband source such as the blue LED, an instrument using a filter can produce substantial errors because of mismatches between the spectral response and $V(\lambda)$ in a narrow wavelength range. Source: Reproduced with permission of the Lighting Research Center.

Many luminaires are outfitted with optics, such as the epoxy capsule of a 5-mm LED source, which can help collimate the light into a narrow beam. When optics are incorporated into a light source or luminaire, the test measurement distance must usually exceed the five-times rule by a substantial amount in order to ensure that the measured values will accurately represent the actual luminous intensity from the source being measured [1].

57.3.1.5 Luminous Flux Measurements Provided illuminance measurements at various angles from a light source are made at distances sufficient to employ the inverse-square law in order to accurately estimate the luminous intensity, these values can in turn be used to estimate the total luminous flux produced by the source. For example, consider a luminaire such as a downlight that produces a radially symmetric distribution downward (Fig. 57.19). An angle of 0° is defined as the angle directly below the luminaire. One could divide the angles from 0° to 90° into four zones with angular widths of 22.5° . If luminous intensity measurements are made at the midpoints of each of these zones (i.e., at 11.25° , 33.75° , 56.25° , and 78.75°), constants (denoted Z_n) relating the luminous intensity to the luminous flux within these zones around the luminaire can be calculated. These constants are specific to the width and number of zones, but for the simple example of the four zones in Figure 57.19, the luminous intensity within a zone is multiplied by the zonal constant for that zone to estimate the luminous flux within that zone. These four products would be summed to obtain the total luminous flux produced by the downlight luminaire.

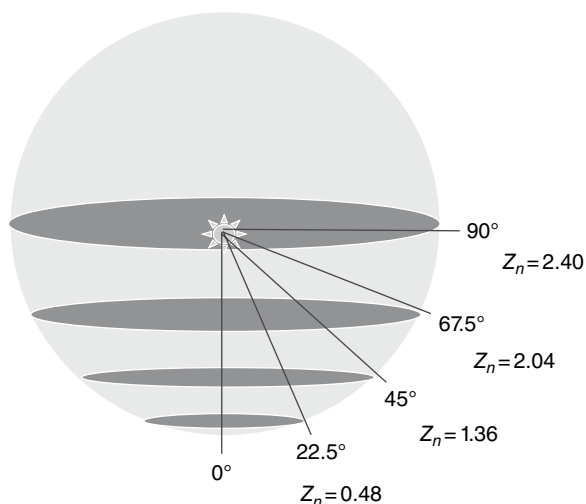


FIGURE 57.19 Illustration of the use of zonal constants (Z_n) to calculate the luminous flux produced by a light source or luminaire. Source: Reproduced with permission of the Lighting Research Center.



FIGURE 57.20 Photograph of an integrating sphere. Source: Reproduced with permission of the Lighting Research Center.

Another way to measure the luminous flux from a light source is through the use of an integrating sphere (Fig. 57.20). This is a large sphere painted matte white on the interior surface with a special high-reflectance paint. Because of multiple reflections from the white painted surface when a light source is placed inside the

sphere, the illuminance on the sphere's interior surface can be estimated by the following equation:

$$E = \frac{\rho\Phi}{4\pi r^2(1-\rho)} \quad (57.4)$$

where ρ is the reflectance of the white paint,

r is the radius of the sphere (in m),

Φ is the luminous flux from the source (in lm),

and E is the illuminance (in lx) on the interior surface of the sphere.

By rearranging terms in Equation 57.4, it is possible to solve for Φ (in lm) for a measured value of E (in lx). Of course, this method is subject to some limitations. The reflectance of paint used in spheres is not perfectly spectrally flat, so interreflections within the sphere create some spectral distortion that can affect the resulting measurement of illuminance. In addition, spheres and baffles used in the sphere will create more interreflections than would be caused within a completely empty sphere, which also influences the measured value. In conjunction with standard lamps of known calibration, however, the use of an integrating sphere can provide reasonably good accuracy for luminous flux measurements [1].

57.3.2 Colorimetry

The human retina contains three types of cone photoreceptors, which respond to light in different parts of the visible spectrum. Short-wavelength (S) cones are responsive to short wavelengths, medium-wavelength (M) cones to intermediate wavelengths, and long-wavelength (L) cones to long wavelengths (Fig. 57.21).

The three types of signals generated by these cone types offer the potential for color vision because the visual system can compare signals from different cone types and use this information to determine the likely wavelength band of a light source. For example, a blue LED producing most of its light near 460nm will elicit a relatively strong S cone signal but weak M and L cone signals. A red LED near 630nm would generate a strong L cone signal but a weaker M cone signal (and little S cone signal either). These combinations of cone inputs are denoted color channels. There are two color channels in the human visual system: a blue-yellow channel that compares signals from S cones to those from M and L cones, and a red-green channel that compares signals from M cones to those from S and L cones. If both channels are relatively balanced by the stimulation of all cone types, the resulting color appearance is white.

Interestingly, sources with different SPDs can produce the same relative blue-yellow and red-green channel signals and will appear to have the same color appearance. The illustration in Figure 57.22 shows two light sources that would both appear identical to the human eye [8].

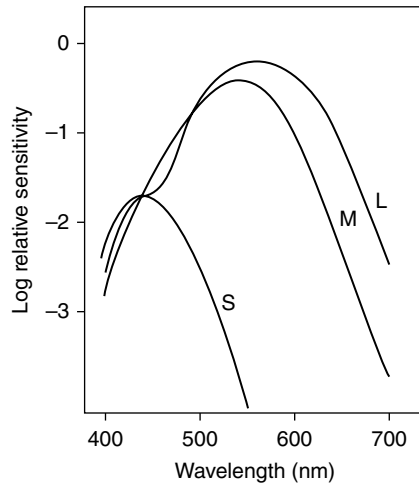


FIGURE 57.21 Spectral sensitivity of the three cone photoreceptor types in the human retina. Source: Reproduced with permission of the Lighting Research Center.

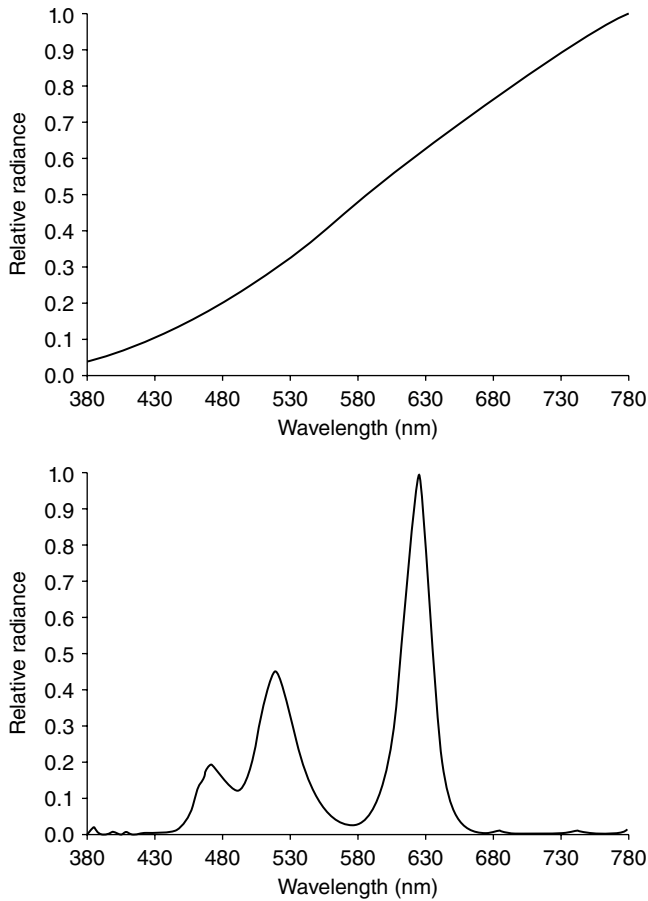


FIGURE 57.22 Spectral power distributions that will have the same color appearance to a human observer. The top panel represents the spectrum for an incandescent lamp; the bottom panel represents the spectrum for a light emitting diode source. Source: Reproduced with permission of the Lighting Research Center.

These types of SPDs are called metamers. Although the light from light sources would appear to be the same, it is important to understand that colored objects illuminated by them can and often would look very different. A very deep red object illuminated by the SPD in the left panel of Figure 57.22 would probably look red, whereas under the SPD in the right panel of Figure 57.22 it would probably appear much darker or even black. In general, light sources with SPDs having energy distributed throughout the visible spectrum are more likely to produce better color rendering (see Section 57.2.1.4 of this chapter for information on color rendering).

57.3.2.1 CIE Colorimetric System The lighting industry uses a system developed by the International Illumination Commission (CIE, for its French name, Commission Internationale de l'Éclairage) to communicate information about the color of light sources and of surfaces illuminated by different light sources. This system is not based on color appearance or the opponent channels described previously but rather on the concept of color matching: if two light sources or illuminated surfaces match each other, they can be said to be identical.

57.3.2.2 Color Matching Functions By using three primary light sources, a red, a green, and a blue source, it is theoretically possible to mix them to create an exact match to any of the individual visible wavelengths from 380 to 780 nm. Experiments with three such primary light sources, having wavelengths of 435.8 nm (blue [b]), 546.1 nm (green [g]), and 700.0 nm (red [r]), were conducted to demonstrate this [6]. Strictly, it is not possible to match all wavelengths with combinations of all three of these primaries. Wavelengths between 435.8 and 546.1 nm cannot be matched in this fashion. However, it is possible to add one of the primaries (i.e., the 700.0 nm red primary) to the individual wavelengths between 435.8 and 546.1 nm to create a match to a combination of the other two primary sources. Using arithmetical logic, adding a quantity to a term on one side of an equation can be canceled by subtracting the same quantity from the other side of the equation. The proportions of each primary needed to match all of the visible wavelengths (using negative proportions for the red primary between 435.8 and 546.1 nm) can be shown graphically; the resulting curves in Figure 57.23 are called color matching functions.

Color matching functions like the ones shown in Figure 57.23 can be generated for any three primary sources consisting of single wavelengths; in some wavelength regions, negative values would have to be used to produce the necessary color matches. To avoid the unpleasantness of color matching functions having negative values, it is possible to create color matching functions that have only positive values by using imaginary primaries, analogous to colors more saturated than any that can physically be generated. Figure 57.24 shows the three color matching functions that were standardized by the CIE in 1931 and used today to characterize color matches.

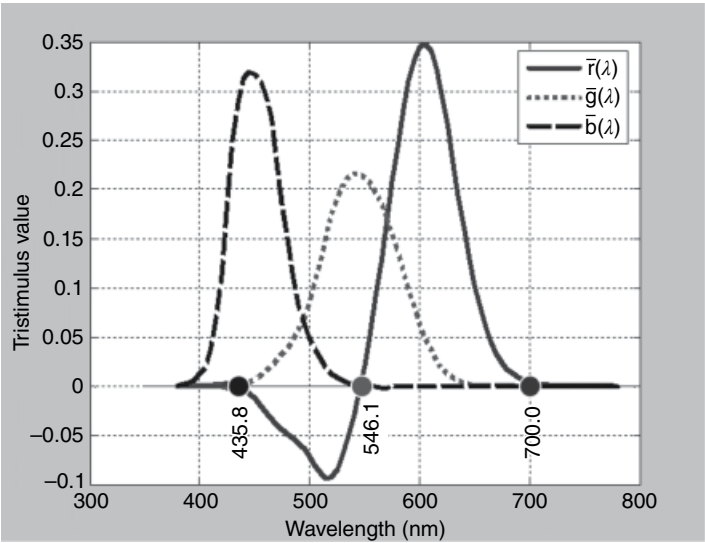


FIGURE 57.23 Color matching functions, showing the relative amounts (denoted tristimulus values) of three primary colors (b, 435.8 nm; g, 546.1 nm; and r, 700.0 nm) needed to match each wavelength in the visible spectrum.

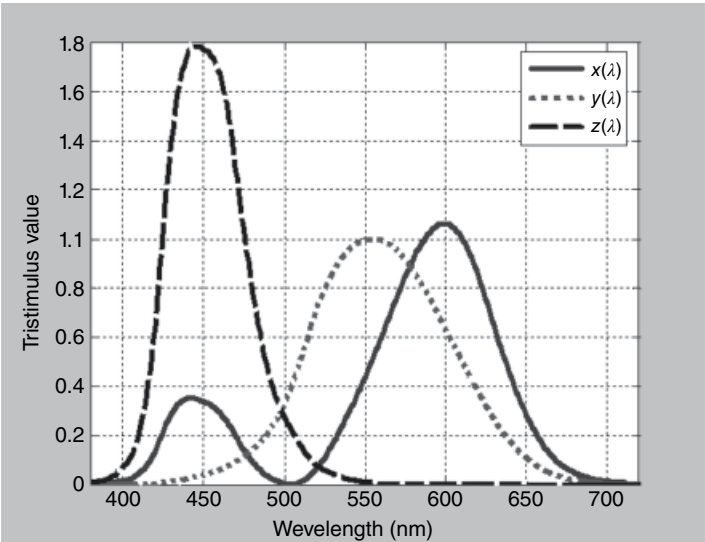


FIGURE 57.24 Color matching functions (x , y , z) based on imaginary primary color stimuli.

Each of the three functions, denoted by symbols \bar{x} , \bar{y} , and \bar{z} , can be used to calculate tristimulus values for each function, denoted by uppercase X , Y , and Z from an SPD in a similar manner used to calculate the luminous flux of an SPD using the photopic luminous efficiency function (Eq. 57.5a–c). In fact, the \bar{y} color

matching function was adjusted to match the photopic luminous efficiency function exactly:

$$\begin{aligned}
 X &= \int P(\lambda) \bar{x}(\lambda) d\lambda \\
 Y &= \int P(\lambda) \bar{y}(\lambda) d\lambda \\
 Z &= \int P(\lambda) \bar{z}(\lambda) d\lambda \\
 x &= \frac{X}{X + Y + Z} \\
 y &= \frac{Y}{X + Y + Z}
 \end{aligned} \tag{57.5a-e}$$

In Equation 57.5a–c, X , Y , and Z are the tristimulus values corresponding to the \bar{x} , \bar{y} , and \bar{z} color matching functions, respectively; $P(\lambda)$ is the SPD of the light source. The lowercase symbols x and y are the chromaticity coordinates for the SPD.

57.3.2.3 Chromaticity Coordinates From the tristimulus values X , Y , and Z , the previous equations can be used to calculate the chromaticity coordinates of the SPD in question. These coordinates, denoted by lowercase x , y , and z , indicate the relative proportion of each tristimulus value to the sum of all three values (Eq. 4.5d and e). By definition, therefore, the sum of the x , y , and z chromaticity coordinates is always 1. Because of this, once x and y are known, the value of z is fixed at $1 - x - y$, and it provides no additional information. For this reason the z chromaticity coordinate is almost never used.

The chromaticity coordinates for a particular SPD can be plotted on a set of rectangular coordinates with x and y axes. Since it is possible to calculate the chromaticity coordinates of the individual wavelengths from 380 to 780 nm, many chromaticity diagrams show these values, and they are often referred to as the spectrum locus (Fig. 57.25). The spectrum locus forms an inverted “U” shape in the chromaticity diagram. Since it is possible to mix very short wavelengths (i.e., near 380 nm) with very long wavelengths (i.e., near 780 nm), a straight line can join these extreme wavelengths where such combinations will appear as various shades of purple, and the line is called the purple boundary.

The chromaticity coordinates of any SPD that can be produced or imagined will always have chromaticity coordinates within the area bounded by the spectrum locus and the purple boundary. In addition, mixing the light from any two sources of light will result in a mixture that has chromaticity coordinates along the line connecting the two components’ chromaticity coordinates. Observing the illustration in Figure 57.25, it can be seen that it is possible to find some mixture of two wavelengths, such as 490 and 570 nm, that will produce a match to a particular mixture of two other wavelengths, such as 500 and 600 nm, because the line segments connecting each pair of wavelengths would intersect.

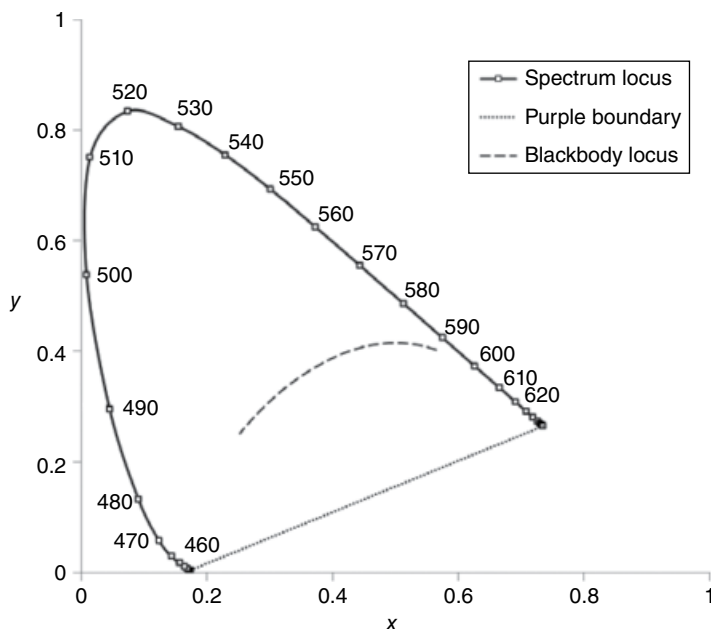


FIGURE 57.25 Chromaticity diagram showing (x, y) coordinates for several wavelengths along the spectrum locus. Also shown are the purple boundary and the blackbody locus (for a range of color temperatures between 1,667 and 25,000 K).

57.3.2.4 Color Metrics Based on Chromaticity Coordinates The chromaticity diagram and its system of chromaticity coordinates can provide many insights regarding the color properties of different light sources, and two useful color metrics can be determined from the chromaticity coordinates. One is the dominant wavelength. Selecting the chromaticity of a particular white light source as a reference (commonly, either the chromaticity of an incandescent source or, as shown in Fig. 57.26, the chromaticity of a spectrum having equal radiant power across all visible wavelengths, called an equal-energy SPD), it is possible to extend a line from the chromaticity coordinates of the reference to those of the light source in question until it reaches either the spectrum locus or the purple boundary. If the line reaches the purple boundary, then the light source in question has no dominant wavelength, but if it intersects the spectrum locus, the wavelength corresponding to the intersection point is the dominant wavelength.

Another metric is called the complementary wavelength, and it is determined by extending a line from the chromaticity coordinates of the light source in question through those of the reference source, until it intersects the spectrum locus or the purple boundary. If the line intersects the purple boundary, the source in question has no complementary wavelength, and if it intersects the spectrum locus, the complementary wavelength is determined as for the dominant wavelength. Some light sources will have both a dominant wavelength and a complementary wavelength, while others will have only one or the other.

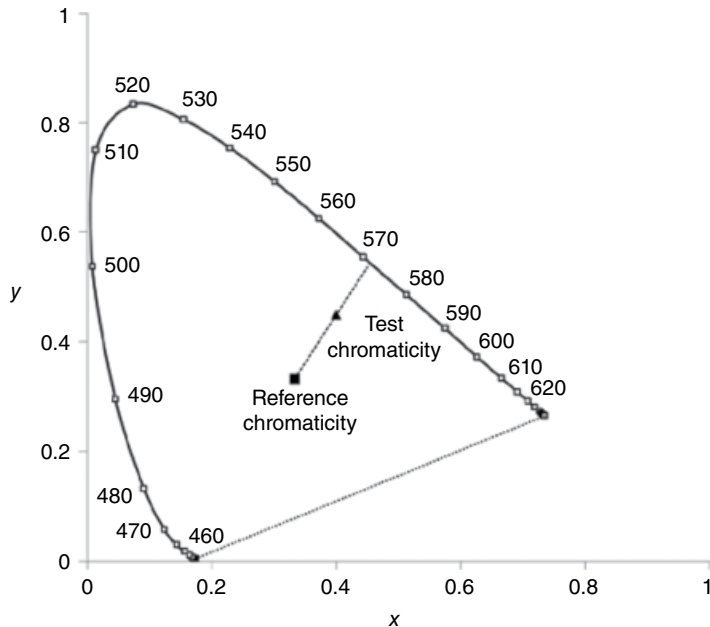


FIGURE 57.26 Illustration of the graphical determination of dominant wavelength. For the example shown here, the line segment intersecting the reference chromaticity (equal-energy SPD) and the test chromaticity intersects the spectrum locus at a dominant wavelength slightly longer than 570 nm.

Related to the dominant wavelength is another metric called the excitation purity. Using the same line as that used to determine dominant wavelength, it is determined by taking the length of the line between the reference source's chromaticity coordinates and those of the source in question and dividing that length by the length of the line between the reference source's chromaticity coordinates and the spectrum locus or purple boundary. It is generally expressed as a percentage. In Figure 57.26, the excitation purity of the test chromaticity source (using the shown reference source) is about 50% because the test chromaticity lies about halfway between the reference chromaticity and the spectrum locus.

57.3.2.5 Chromaticity Discrimination It has been stated previously that two light sources that have identical chromaticity coordinates will produce matches to the human eye. It should seem reasonable to suppose that if two light sources have chromaticity coordinates that do not match exactly, but are very close to one another, they would be judged to match most of the time. Since hardly any two light sources will ever match exactly in terms of their chromaticity, not even two lamps of the exact same type and make, because of factors such as variations in manufacturing, it is reasonable to want to know how different chromaticity coordinates can be before they will be reliably judged as different.

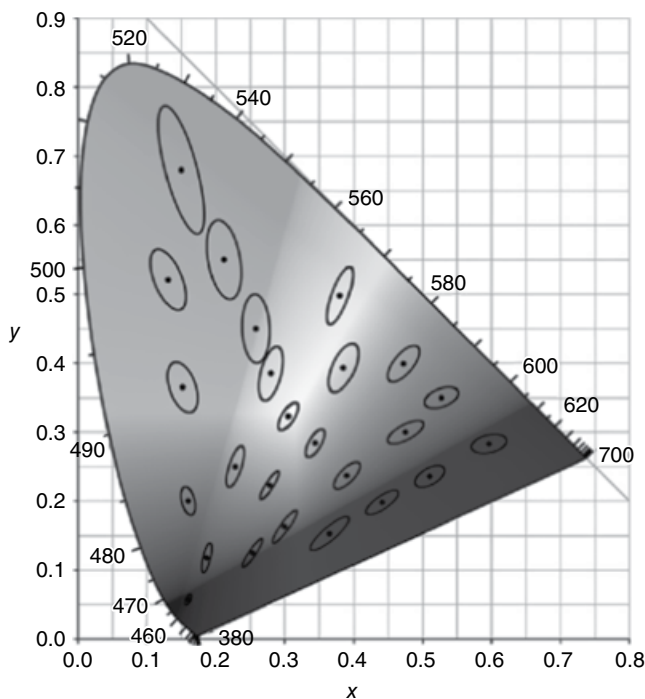


FIGURE 57.27 MacAdam [9] ellipses for various chromaticities, increased in size by a factor of 10.

So-called MacAdam ellipses [9] have been determined experimentally for different regions within the chromaticity diagram. These ellipses indicate, for a given reference source having certain chromaticity coordinates, the boundary indicating the chromaticity coordinates of light sources that most people would judge to match the appearance of the reference source. They are shown, 10 times larger than their actual size to make them easier to see, in Figure 57.27.

MacAdam ellipses are used in the specification of chromaticity tolerances for commercial lamps such as fluorescent lamps [8]. Lamps having various nominal CCTs must fall within four-step MacAdam ellipses having radii four times larger than the original MacAdam ellipses (and just under half the size of the 10-step ellipses shown in Fig. 57.27).

57.3.2.6 Uniform Chromaticity Diagram If the CIE 1931 chromaticity diagram were perceptually uniform, equal vector distances in the diagram would correspond to equal perceived color differences. The varied shapes and sizes of the MacAdam ellipses demonstrate that this is not the case. If the chromaticity diagram were perceptually uniform, the sizes of the ellipses would all be the same, and moreover, the ellipses would instead be perfect circles. Several attempts to develop more uniform chromaticity diagrams have been made, with one of the more successful being the CIE 1976 uniform

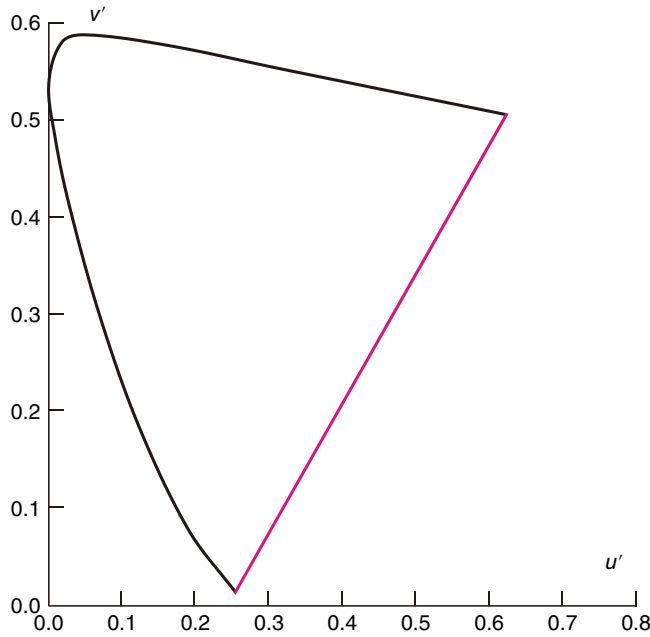


FIGURE 57.28 CIE 1976 uniform chromaticity diagram.

chromaticity diagram, using coordinates denoted u' and v' in place of x and y . The 1976 uniform chromaticity diagram was developed through a linear transformation of the x and y coordinates in the 1931 diagram:

$$u' = \frac{4x}{(-2x + 12y + 3)} \quad (57.6a)$$

$$v' = \frac{9y}{(-2x + 12y + 3)} \quad (57.6b)$$

While the 1976 CIE chromaticity diagram is more perceptually uniform than the 1931 diagram, it is not perfect in that MacAdam ellipses are not transformed into perfectly equally sized circles, but it is a marked improvement over the 1931 diagram. Nonetheless, the 1931 diagram is much more commonly used in the lighting industry to characterize the chromaticity of different light sources, and the 1976 diagram is mainly used in color research and in specific applications where color discrimination might be extremely important to predict, such as some printing processes (Fig. 57.28).

57.3.2.7 Munsell Color System Another system is used in many architectural applications [1] to indicate the color of a surface, known as the Munsell color system. Unlike the CIE chromaticity diagram, which provides a quantitative framework for

characterizing light source and surface color, the Munsell system is a color order system based on three attributes of color known as the hue, value, and chroma.

The hue is the quality identified with the color name, such as blue, green, yellow, red, or violet. Certain color combinations are also allowed, such as blue–green or yellow–red. The value indicates the lightness of the color with a value of 0 corresponding to perfect black, and a value of 10 corresponding to perfect white. The chroma is an indication of how saturated a color appears. Different hue and value combinations can have different maximum chroma values. As an example, a color with a very high value will appear like a pastel and cannot be highly saturated.

Steps along the hue, chroma, and value dimensions are judged to be approximately equal. Sets of Munsell chips containing finely gradated color chip samples along the hue, value, and chroma dimensions can be used by architects and interior designers to specify paint and finish colors in built spaces. Munsell chips are also used in some research studies as stimuli of known, repeatable colors.

57.4 INSTRUMENTATION

57.4.1 Illuminance Meters

One of the most commonly used and least expensive types of photometric instruments is the illuminance meter (Fig. 57.29). Many illuminance meters are handheld devices consisting of a sensor element, usually a silicon or selenium detector with a photopic filter having spectral characteristics similar to those illustrated in Figure 57.18. The readout displays the measured illuminance value (lx or fc). When measuring illuminance, the sensor element should be placed as close as possible and parallel to the surface on which the illuminance is being measured. The sensor element's spatial sensitivity corresponds to a cosine distribution; incident light from directly above the sensor is weighted fully, while light incident from an angle is proportional to the cosine of that angle.

Care should be taken to avoid producing shadows on the sensor element when making measurements and to avoid the possibility of reflected light from the user's clothing that can increase the illuminance at the sensor element's location. A quality illuminance meter can be purchased for between US\$100 and US\$300.

57.4.2 Luminance Meters

A number of portable luminance meters are commercially available (i.e., Fig. 57.30). Unlike illuminance meters, luminance meters require imaging optics to capture light emitted or reflected from a relatively narrow entrance angle (1° and 0.33° entrance angles are representative values). Luminance meters have view finders to allow the user to position the aperture over the part of the visual scene of interest. Spectral correction is usually achieved by using a filter over a silicon sensor element.



FIGURE 57.29 A portable illuminance meter with a detachable sensor element.



FIGURE 57.30 Portable luminance meter. Source: Reproduced with permission of the Lighting Research Center.

When using a luminance meter, it is important that the part of the visual field that is being measured is seen in clear focus through the view finder to minimize the effects of stray light from influencing the measurement. Because of the additional optics required by luminance meters, they generally have substantially higher costs than illuminance meters, on the order of US\$1000 to US\$3000.

57.4.3 Spectroradiometers

Most illuminance and luminance meters provide only photometric measurements without regard for the colorimetric properties of the sources or surfaces being measured because of the use of broadband detectors and filters that integrate responses according to the appropriate luminous efficiency function (usually, the photopic function). For this reason they cannot provide information about the color properties of the lighting conditions they are used to measure. Spectroradiometers are used to measure the spectral power at various wavelengths across the visible spectrum. They incorporate collection optics to receive the radiant power from the object being measured, usually in a manner similar to luminance meters. A monochromator with an element such as a diffraction grating or a prism disperses the light of varying wavelengths onto an array of detectors. The circuitry from the detectors processes their signals and stores the results for each wavelength band so that the SPD curve can be generated, displayed, and stored. Most spectroradiometers will also have additional processing software to calculate quantities such as the CCT, CRI, and chromaticity coordinates. Because of their complexity relative to illuminance and luminance meters, spectroradiometers are generally quite expensive, ranging from US\$5,000 to US\$50,000.

REFERENCES

1. Rea MS, ed. 2000. *IESNA Lighting Handbook: Reference and Application*, 9th ed. New York: Illuminating Engineering Society.
2. Commission Internationale de l'Éclairage. 1978. *Light as a True Visual Quantity*, No. 41. Paris: Commission Internationale de l'Éclairage.
3. Bullough JD. 2005. Research matters: What's cooler than cool? Warm! *Lighting Design and Application* 35(2): 12–14.
4. Rea MS, Freyssinier JP. 2010. *ASSIST Recommends: Recommendations for Specifying Color Properties of Light Sources for Retail Merchandising*. Troy, NY: Rensselaer Polytechnic Institute. Accessed on February 28, 2014 at <http://www.lrc.rpi.edu/programs/solidstate/assist/recommends/lightcolor.asp>.
5. Bullough JD, Gu Y, Narendran N, Taylor J. 2005. *ASSIST Recommends: LED Life for General Lighting: Definition of Life*. Troy, NY: Rensselaer Polytechnic Institute. Accessed on February 28, 2014 at <http://www.lrc.rpi.edu/programs/solidstate/assist/recommends/ledlife.asp>.
6. Wyszecki G, Stiles WS. 1982. *Color Science*, 2nd ed. New York: Wiley-Interscience.
7. Commission Internationale de l'Éclairage. 2010. *Recommended System for Mesopic Photometry Based on Visual Performance*, No. 191. Vienna: Commission Internationale de l'Éclairage.
8. Rea MS, Deng L, Wolsey R. 2004. *Lighting Answers: Light Sources and Color*. Troy, NY: Rensselaer Polytechnic Institute. Accessed on February 28, 2014 at <http://www.lrc.rpi.edu/nlpi/publicationDetails.asp?id=901>.
9. MacAdam DL. 1942. Visual sensitivities to color differences in daylight. *Journal of the Optical Society of America* 32(5): 247–273.

THE DETECTION AND MEASUREMENT OF IONIZING RADIATION

CLAIR J. SULLIVAN

Department of Nuclear, Plasma, and Radiological Engineering, University of Illinois at Urbana-Champaign, Urbana, IL, USA

58.1 INTRODUCTION

Detection of ionizing radiation can be traced by to 1895 when W.C. Rontgen made the first medical radiograph using X-rays. Since that time, many high-quality textbooks have been written on the subject [1–7]. We will not attempt to recreate those works here but rather to convey a working familiarity with the concepts behind how these detectors work and are used in everyday measurements. While there are many types of radiation such as radio-frequency emissions, infrared, etc., this chapter seeks to address specifically the detection of ionizing radiation.

The detection and measurement of ionizing radiation can be described in one sentence: *it is all about converting the incident radiation to charge (or light and then charge) and measuring that charge, whose magnitude is ideally proportional to the type, energy, and/or quantity of incident quanta*. While this statement on its surface is simple, there is much subtlety conveyed. It is important to understand what type of radiation is incident, how that type of radiation interacts with matter and ideally creates charge, and then how that charge is read out of the system and analyzed. Therefore, to truly understand radiation detection, it is necessary to understand some basic physics, statistics, materials science, and electronics design. In this chapter, we hope to provide a working knowledge of the basics necessary to understand each of these concepts.

58.2 COMMON INTERACTIONS OF IONIZING RADIATION

Ionizing radiation comes in two major categories, charged particulate radiation and uncharged radiation, with multiple different types of each. Examples of charge particulate radiation include alpha particles (which is just the nucleus of a helium atom containing two protons and two neutrons), beta particles (which is either an electron or a positron), protons, fission fragments, etc. Uncharged ionizing radiation includes photons such as gamma rays and X-rays as well as neutrons.

Most ionizing radiation is created through the radioactive decay of an unstable parent isotope. This decay occurs within some statistical time distribution and results in a daughter nucleus and the ionizing radiation particle. How quickly a given isotope decay determines is so-called radioactivity, which is mathematically described as

$$\frac{dN}{dt} = -\lambda N$$

where N is the number of nuclei, t is time, and λ is called the decay constant, which is related to the isotope's half life, $t_{1/2}$, through the relation

$$\lambda = \frac{\ln 2}{t_{1/2}}.$$

There are several units that are assigned to radioactivity, the most common being the Curie (Ci) and the becquerel (Bq). 1 Bq is equivalent to 1 decay per second or 2.703×10^{-11} Ci.

In addition to radioactive decay, ionizing radiation can be produced through other processes such as spontaneous fission, excitation resulting from an energizing source impacting a target, etc. These processes are beyond the scope of this chapter but are discussed in greater detail in many other texts [3, 7].

58.2.1 Radiation Interactions

In order to detect ionizing radiation, we must first get it to interact with a detecting medium so charge can be produced. How radiation interacts with matter is a function of the type of radiation, its energy, and what it is interacting with. The probability of an incident particle interacting with the medium can be statistically described and has been measured for many different types of materials. The measurement process is reasonably simple: a calibrated source of radiation is directed through a material of thickness x , as shown in Figure 58.1. The number of particles that make it through the material can be described by the equation

$$I = I_0 e^{-\mu x}$$

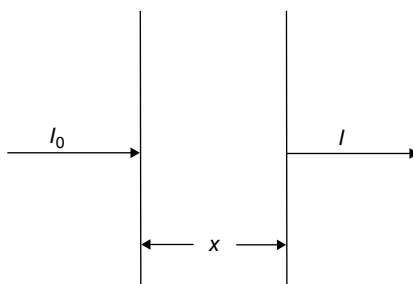


FIGURE 58.1 Example of source attenuation of initial intensity, I_0 , to measured intensity, I , through an absorber of thickness x .

where μ is called the linear attenuation coefficient. It is related to the mean free path (mfp), the average distance before an interaction takes place in an absorber, as

$$\text{mfp} = \frac{1}{\mu}.$$

It is important to note that μ , and hence the mfp, takes into account all types of interactions and is also a function of the type of radiation, energy, and what it is interacting with. For example, gamma rays can interact in one of three different ways depending on their energy: the photoelectric effect, Compton scatter, or pair production. So μ is a linear combination of the interaction probabilities of each individual type of interaction.

Another important parameter for radiation interaction is how much energy, E , is deposited as a function of path length, x . This is called the linear stopping power, which is differentially defined as

$$S = -\frac{dE}{dx}.$$

Particles with higher stopping power, such as alpha particles, deposit more energy per interaction, meaning that they lose energy quickly within an absorber. Ideally, it is this energy loss that creates the charge that will be the basis for detection.

58.3 THE MEASUREMENT OF CHARGE

There are three predominant ways that charge can be measured from a detector: the measurement of current across a resistor, voltage across a capacitor, or the mean square voltage. Generally speaking, the charge created in a radiation detector is small, necessitating a certain emphasis on charge amplification and electronics. Prior to a discussion of these various modes or the actual measurement of current or voltage, it is necessary to briefly discuss the statistics associated with radiation detection.

58.3.1 Counting Statistics

There are three predominant probability distributions that apply to radiation detection. The first and most basic of these is the binomial distribution, which is given by

$$P(x) = \frac{n!}{(n-x)!x!} p^x (1-p)^{n-x}$$

where $P(x)$ is the normalized probability of counting x successes, p is the probability of success for a single event, and n is the number of trials. It is clear then that, since x is a discrete random variable, $P(x)$, called the probability mass function (PMF), is the probability of obtaining exactly x discrete successes. This should not be confused with a cumulative probability distribution (CDF) but can lead to the CDF by considering what happens when x becomes infinitely small. Since in real measurement applications it is not possible to have infinite precision, the PMF is sufficient for our purposes.

Using this equation, it is possible to determine the mean of the distribution, μ , which can be found to be

$$\mu = \sum_{x=0}^n xP(x) = pn.$$

This fact will be useful in further simplifications described in the following. Additionally, the predicted variance of the distribution can be calculated to be

$$\sigma^2 = \sum_{x=0}^n (x - \mu)^2 P(x) = \mu(1-p).$$

To calculate $P(x)$, μ , and σ^2 , we must understand how a success, p , is defined. There are many trivial examples for probability students that involve coin flip (where success could be defined as the number of times the experimenter gets heads) or rolling a six-sided die (where success might be considered as rolling a three). In the case of coin tosses, the probability of success is obviously one half whereas for rolling the dice it is one sixth. For radiation detection, the probability of success is considered to be the probability of an individual, specific nucleus decaying. This is given by

$$p = 1 - e^{-\lambda t}$$

where λ is the decay constant for the isotope, which is generally known and is very small.

An important simplification is possible when p is small and n is large through the use of the so-called Poisson limit theorem. The product of the two approach a limiting value, or $np \rightarrow \mu$, which can be used to rewrite the binomial distribution as

$$P(x) \rightarrow \frac{\mu^x}{x!} e^{-\mu},$$

which is known as the Poissonian distribution. Since we know that p is small for radioactive decay, we can default to using the Poissonian distribution for radiation counting statistics. Like for the binomial distribution, the mean is given by $\mu = pn$. However, the predicted variance is slightly simpler:

$$\sigma^2 = \mu.$$

Further simplification is possible if we consider what happens when μ is large and take advantage of the fact that the predicted variance is equal to the mean. (Most practitioners consider “large” to be around 25–30 or when $n(1 - p) \geq 5$.) In this case, we can rewrite the Poissonian distribution as the well-known Gaussian distribution given by

$$P(x) = \frac{1}{\sqrt{2\pi\mu}} \exp\left(-\frac{(x - \mu)^2}{2\mu}\right)$$

or

$$P(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x - \mu)^2}{2\sigma^2}\right).$$

Examples of the binomial, Poissonian, and Gaussian distributions are shown in Figure 58.2.

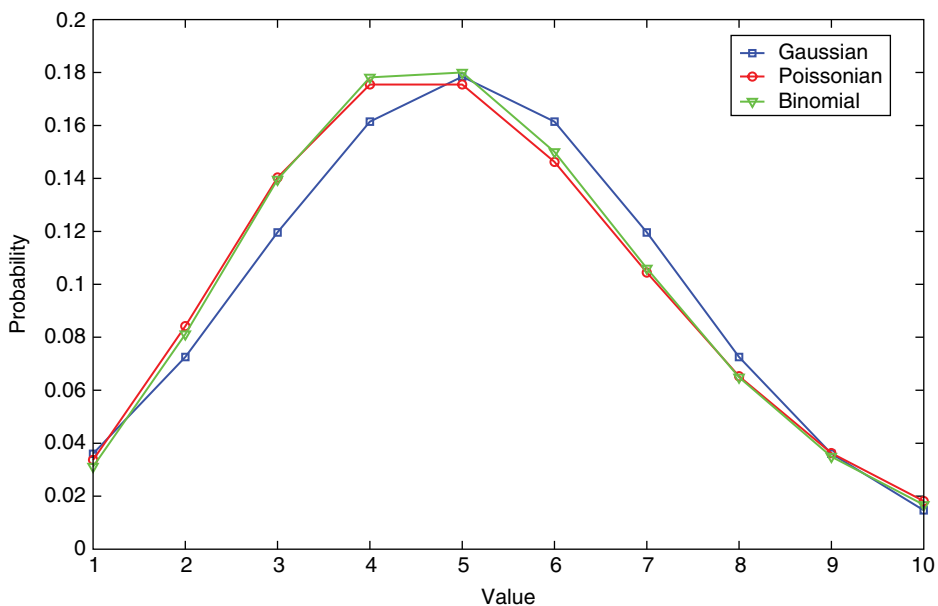


FIGURE 58.2 Examples of the binomial, Poissonian, and Gaussian distributions around a mean value of 5. Note the peak value differs based on distribution.

58.3.2 The Two Measurement Modalities

The deposition of energy from incident radiation in a detector results in the creation of charge, which must be moved through the application of an external field and then collected. This can be done as a function of time, observing individual radiation interaction events or “counts” as they are created in the detector, as a time-integrated average whereby averages of that moving charge are measured in some time window. The former case, referred to as “pulse mode operation,” measures the movement of charge within the detector on an event-by-event basis. The detector is connected to some readout electronics as shown in Figure 58.3. Assuming the time it takes for the charge to travel is significantly smaller than the RC time constant of this circuit, a signal is created representing the change in signal voltage as a function of time, $V(t)$, whose maximum is given by

$$V_{\max} = \frac{Q}{C},$$

where Q is the amount of charge created per event in the detector. This can be calculated as

$$Q = \frac{E}{W} e_0$$

where W is the mean ionization energy for the detecting material, E is the energy deposited in the interaction, and e_0 is the elementary charge of an electron (1.6×10^{-19} C). Thus it is clear that a measurement of V_{\max} is directly proportional to the energy deposited in the detector (which is hopefully proportional to the energy of the source). Because of this fact, pulse mode operation is the most common mode of operation for radiation detection.

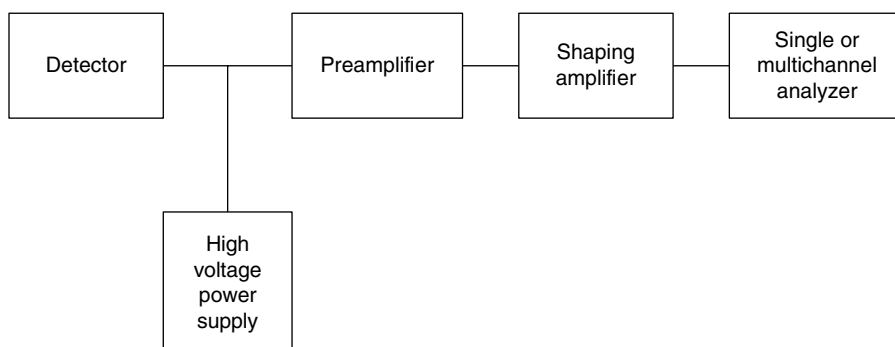


FIGURE 58.3 Standard setup for radiation measurements. This setup can be used either to count radiation events through the single channel analyzer or to collect spectra through the multichannel analyzer.

Another common method of measurement involves measuring the integrated current that is the charge moving under the application of the external field. This type of measurement is most common when an average event rate, r , must be known, such as is common in radiation dosimetry and health physics. In this case, the current is measured and averaged over some integration time. Then, r can be calculated by

$$r = \frac{\bar{I}}{Q} = \bar{I} \frac{W}{Ee_0}$$

where \bar{I} is the average current generated through the movement of charge of many incident quanta averaged over a period of time. While this measurement is not done on an event-by-event basis, the electronics necessary to measure a current signal tend to be simpler and are perfectly suitable when rate-based information (such as dose rate) is all that the user requires.

58.4 MAJOR TYPES OF DETECTORS

As previously stated, the goal of any radiation detector is to convert the incident radiation to charge through the interaction within some detecting medium. The three most common types of detecting media are gases, scintillating materials (solid, liquid, or gas), and semiconductors. (It should be noted that scintillators do not directly convert the incident radiation to charge, but they convert it to light that is later converted through charge via optical readout techniques.) Each of these detector types shall be described in detail in the following.

58.4.1 Gas Detectors

The ionization of a gas atom or molecule is the basis for the simplest type of radiation detector. The general concept behind the gas detector is that the radiation enters a volume of gas and interacts through excitation and ionization. Excitation does not directly result in the formation of charge (although charge can be created indirectly through the deexcitation process, which can sometimes release secondary radiation such as an X-ray that is later collected), so we will focus on the ionization process. In ionization, the incident ionizing radiation interacts with an atomic electron, giving it enough energy to escape the atom and become a free particle. The result of ionization is the creation of both a free electron and an ion of the gas molecule—an electron-ion pair. The minimum energy required to liberate an electron is called the ionization energy and is a function of the element as well as which shell the electron originates (outer electrons have lower ionization potentials than those in the inner shells). The ionization potentials for some key elements are provided in Table 58.1 and a photo of several commercial gas detectors is provided in Figure 58.4.

TABLE 58.1 Energies Required to Ionize the Outer Electron Shell of Select Elements^a

Element	First Ionization Energy (eV)
Ge	7.899
Si	8.151
Xe	12.13
Ar	15.759
Ne	21.564
He	24.587

^aFrom Ref. 6.



FIGURE 58.4 Examples of a variety of gas detectors of different sizes and configurations. Source: Reproduced with permission of Saint-Gobain Crystals.

In general, the radiation will likely undergo several ionization and excitation interactions along its path before it is completely absorbed. The result is that several electron-ion pairs will be created. The average amount of total charge created is determined by

$$Q = \frac{E}{W} e_0$$

TABLE 58.2 Average Values of the Energy Required to Create an Ion Pair in Some Select Gases^a

Gas	Average W Value (eV/Ion Pair)
Xe	21.5
Ar	26.2
CH ₄	28.2
O ₂	31.5
Air	34.5
N ₂	35.6
Ne	36.2
He	42.0

^aFrom Ref. 3.

where E and W , the average energy required to create an electron-ion pair, are in units of energy (usually the electron volt, eV, for radiation detection). Some values of W for common fill gases are provided in Table 58.2.

It is also important to understand the statistics of the creation of this charge. In general, we assume Poissonian statistics to govern the variation in the number of charge pairs created, N . If this was truly a Poissonian process, then we would expect the variance in the number of charge pairs created, σ_N^2 , to just be N . However, by simply assuming Poissonian statistics, we have inherently assumed that the processes that create individual charge pairs are independent. In reality, this is not quite true, so a correction factor called the Fano factor has been introduced to account for this variability:

$$\sigma_N^2 = FN$$

where F is the so-called Fano factor, which is usually less than 1 for most gases. The variance in the number of charge pairs created relates to the energy resolution of the detector, so keeping this value small is desirable.

Once the charge Q is created within the gas, all that remains is to collect it. Like charge-based radiation detectors (i.e., not scintillators), the charge pairs are separated through the creation of an electric field. This is usually done by creating two electrodes separated by both a spatial and potential difference. Detectors can be designed to use electrodes in rectilinear, cylindrical, or spherical geometries, as shown in Figure 58.5. The magnitude of the electric field, $|\vec{E}|$, is a function of this geometry and the separation of the two electrodes, d . For rectilinear coordinates, the relationship is simply

$$|\vec{E}| = \frac{V}{d}$$

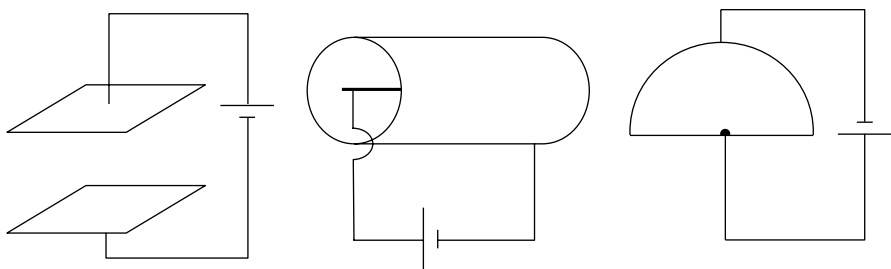


FIGURE 58.5 Examples of planar (left), cylindrical (center), and hemispherical (right) detector geometries with the anode as the high potential electrode and the cathode as the low potential electrode.

TABLE 58.3 Sample Values of Electron Mobilities for a Variety of Different Gases^a

Gas	E/p (V m/N)	
	0.6	1.5
He	8.5	15
Ar	3.9	6.2
N ₂	7	14
Air	11	17
CH ₄	96	117
Ne	12	26

^aFrom Ref. 3.

where V is the potential difference between the two electrodes. The electrode at the higher potential is called the anode whereas that of the lower potential is referred to as the cathode. In cylindrical coordinates, it can be shown that

$$|\overline{E(r)}| = \frac{V}{r \ln(b/a)}$$

where a is the radius of the inner electrode and b is the radius of the outer electrode.

It is also relevant to consider how fast the electron and ion move in the presence of the electric field. The velocity of each charged particle can be determined by

$$\overline{v_{e,i}} = \frac{\mu_{e,i} \overline{E}}{p}$$

where $\mu_{e,i}$ is the mobility of the electron and ion, respectively (of units V m/N or m² atm/V s), and p is the gas pressure. Sample values of mobility are provided in Table 58.3. The mobility of the electron is around three orders of magnitude larger

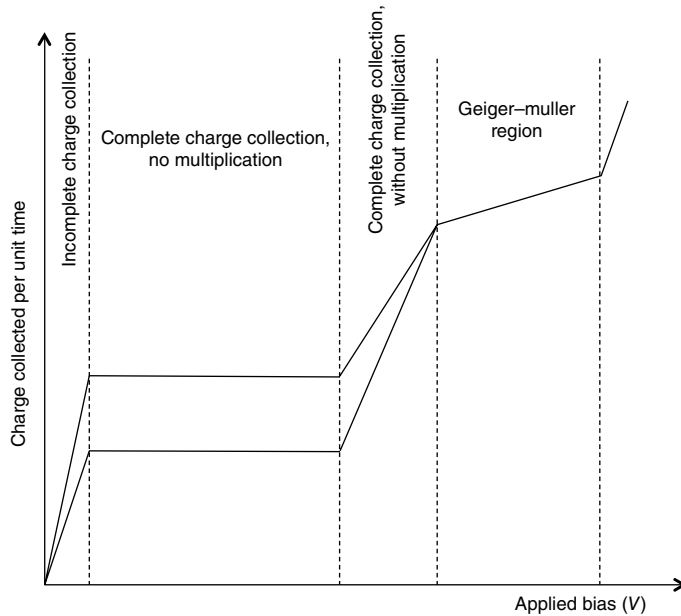


FIGURE 58.6 The four regions of gas detector operation shown for two different energy depositions. Complete charge collection without multiplication corresponds to ionization chamber operation. The next region where multiplication is added is called the proportional region. The final region, where there is no difference in collected charge with energy deposition, is the Geiger–Muller region.

than that of the ions due to the mass difference between the two particles. However, it puts significant constraints on the operation of a gas detector, if full signal generation requires the complete collection of both the electrons and the ions. Collection of the ions will take time, and this means that the count rates that can be handled will be limited. Additionally, the ions easily recombine during this time, resulting in incomplete charge collection. In the following sections, we shall present a method for overcoming this limitation.

Gas detectors can be operated either in current mode or in pulse mode. When in steady state, average values are to be measured, such as average dose rate, and current mode is typically employed due to its simplicity. However, gas detectors can also operate in pulse mode and used to perform spectroscopy. These techniques shall be discussed in detail in the following.

Beyond just variations in the overall geometry of the detector, we can describe gas detectors as being of one of three different types depending on their internal electric fields: ionization chambers, proportional counters, and Geiger–Muller (GM) tubes, as will be described in the following sections. Since the electric field is established by the applied bias, V , it is possible to measure how the number of collected ion pairs changes with V , which is shown in Figure 58.6. When examining this figure, there are four key

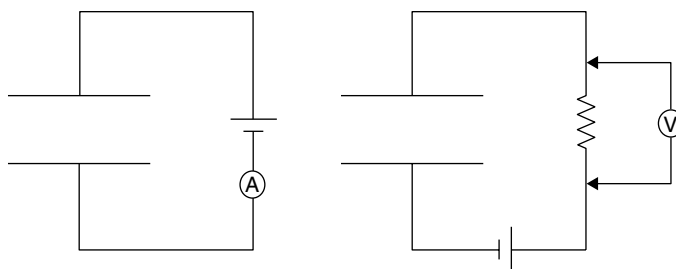


FIGURE 58.7 Circuit diagram for radiation detector operating in current (left) versus pulse (right) modes.

regimes to consider. In the first regime, the number of charge carriers collected increases with applied bias. This corresponds to the fact that the bias is increasing so a larger and larger volume of the detector becomes active. (The inactive regions are where the charge pairs are created too far from their electrodes, resulting in recombination before the charges can be collected.) The second region, which lies soundly on a plateau, is where the electric field is sufficient to collect all charges created within the detector. This is where ionization chambers operate. The third region occurs when the electric field becomes large enough to allow charge multiplication. Called the proportional region, this corresponds to the place where the output signal amplitude is proportional through a constant of multiplication to the input energy. Finally the fourth regime is another plateau where all energy-depositing events in the detector create pulses of equal amplitude. This is the GM region.

It is also important to note that most gas detectors are capable of either current or pulse mode operation, as shown in Figure 58.7, making them a very flexible choice for a variety of applications.

58.4.2 Ionization Chambers

Ionization chambers (or “ion chambers” for short) represent the simplest type of gas detector. In this configuration, the charge Q is created, moved to the electrodes by the internal electric field, and directly counted. One key parameter for an ion chamber is the applied bias at which all charges will be collected, V_s , which is called the saturation voltage. This voltage represents the minimum bias necessary to collect all charge pairs before they recombine and is evident in Figure 58.6 as where the saturation plateau begins. Increasing the applied bias further will not change the amount of charge collected and, therefore, the current amplitude assuming constant, monoenergetic irradiation because all of the charge that is generated is fully collected. Assuming the applied bias is greater than V_s , the output signal is directly proportional to the charge created through ionization interactions.

When operated in current mode, ion chambers are useful for measuring the radiation exposure rate, which has SI units of C/kg air per second. This can be related to the

traditional unit of exposure, the roentgen (R) where 1 R is $2.58 \times 10^4 \text{ C/kg}$ air. One roentgen is defined to be the exposure due to ionizing photons in 1 cm^3 of dry air at standard temperature and pressure. In the presence of a source of radiation, a detector operated at a bias of at least V_s will have a steady-state saturation current of i_s . If the ion chamber can be thought of as being air equivalent, then the exposure rate in SI units is

$$\dot{X} = \frac{i_s}{M}$$

where M is the air mass at standard temperature and pressure. (Note that it may be necessary to adjust the value of M if the detector is at a pressure or temperature other than standard.) Therefore, the measurement of the saturation current directly provides the exposure rate.

Pulse mode operation of an ion chamber, while more complicated, can provide a wealth of additional information about a radiation source, such as type of incident radiation and/or its energy. Prior to discussing practical pulse mode operation, it is necessary to consider how a signal is formed within the detectors as a function of the movement of charge. When a point charge is created at a distance d from an electrode (it doesn't matter if this charge is the electron or the ion), it induces a charge on the surface of the electrode through the method of images or mirror charges. While the actual derivation of the method of images is beyond the scope of this book (the reader is encouraged to consult [8] or [9] for more information), the result is useful. The method of images equates the solution of the potential and electric fields of a point source above a grounded conducting plane to that of two equal but opposite charges separated by the same distance, as shown in Figure 58.8. The surface charge induced on a grounded conducting surface under these conditions is

$$\sigma(x,y) = \frac{-qd}{2\pi(x^2 + y^2 + d^2)^{3/2}}$$

where q is the charge amplitude and x and y are the rectilinear coordinates of the grounded plane. Then the total induced charge on the surface can be found integrating this expression as

$$Q = \int \sigma dx dy = -q.$$

So the total induced charge is equal to the amount of charge. Note though that this integral is done over the area of an electrode. When the charge is far away, it still induces q on the surface of the electrode, but the area this is spread over is much larger since the electric flux density is smaller. As the charge gets closer to the surface, the flux density increases until the limit where the charge is at the surface of the conductor when all induced charge is concentrated in a single point directly beneath q .

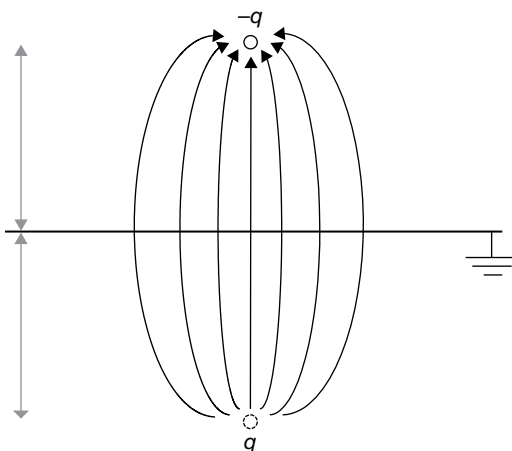


FIGURE 58.8 Illustration of the approach to use the imaginary mirror charge, q , to determine the induced charge on the surface of a grounded conductor by the real charge, $-q$.

Let us now apply this concept to the movement of electrons and ions within a gas detector. Upon their generation by ionization, each charge pair will induce a surface charge on the electrodes of the detector. They immediately begin to move because of the application of the electric field with electrons moving toward the anode and the ions moving toward the cathode. As they move closer to their respective anodes, the amount of charge they induce increases steadily until it reaches the charge value of the individual carrier ($-e_0$ for electrons and $+e_0$ for the ions) and the charge is collected on the electrode surface.

However, it is important to realize that the amount of time it takes for these charges to be collected varies significantly between the electrons and the ions due to the large mobility difference of the two. In reality, the electron is collected much sooner than the ion, so the induced signal initially increases quickly as the electrons and ions are both drifting. But once the electrons are collected, the slope of the time rate of change of induced charge decreases to reflect the fact that the only signal being induced is from the movement of the ions. This is represented schematically in Figure 58.9.

Because the ions move so slowly, this creates limitations on the rate of incident radiation that can be detected if one has to wait for all of the ions to be collected. It is therefore desirable to create a detector capable of creating signal only through the movement of electrons. One way of doing this is by choosing a charge collection time that is small relative to the collection time of the ions. In this case, the movement of the electrons from generation to collection induces a voltage signal, V_e , equal to

$$V_e = \frac{Q}{C} \frac{x}{d}$$

where x is the distance the electron must travel from generation to the anode and d is the distance between the cathode and the anode (the movement of ions is assumed to

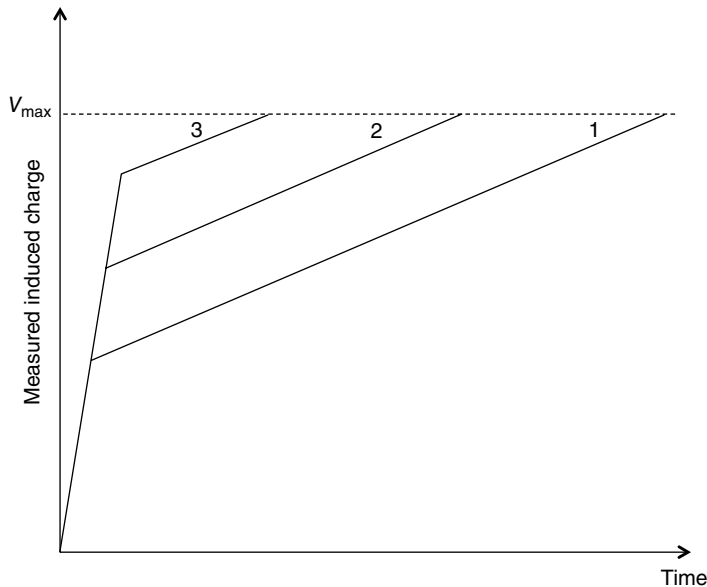


FIGURE 58.9 Pulses created through the measurement of induced charge where the electron-ion pair are created (1) close to the anode, (2) at a point midway between the cathode and anode, and (3) at a point close to the cathode. The fast rise component corresponds to the movement of both the electrons and holes. Once the electrons are collected, which happens much sooner than the ions, then the induced charge is only a function of the slow-moving ions. The measured induced charge reaches its maximal value of Q/C when both the electrons and ions are fully collected.

be negligible in this short time span). As is evident from the equation, this approach suffers from the limitation that the amplitude of the resulting signal is a function of where in the detector the ionization occurred.

The Frisch grid was invented to overcome this problem, as shown in Figure 58.10. In a “gridded” detector, a wire mesh is placed between the anode and cathode and kept at a potential that is intermediate to that of the cathode–anode potential difference. As before, the signal is measured on the anode as the induced charge from the movement of the electrons. The grid acts as a type of Faraday cage for the anode. When the electrons are generated and begin drifting in the drift region of the detector, no charge is induced on the anode. But eventually they pass through the grid and continue to drift toward the anode. Once this occurs, they induce charge on the anode until their collection, as shown in the figure. This does several things. First, the movement of the ions is not measured, so their low mobilities are not a problem. Second, the only portion of the electron movement that induces any signal is between the grid and the anode, a distance that is usually very small. The electrons still induce their full signal but over a very small distance. Therefore, the Frisch grid allows for what is called “single polarity charge sensing” with a signal amplitude that is no longer a function of the interaction location in the detector.

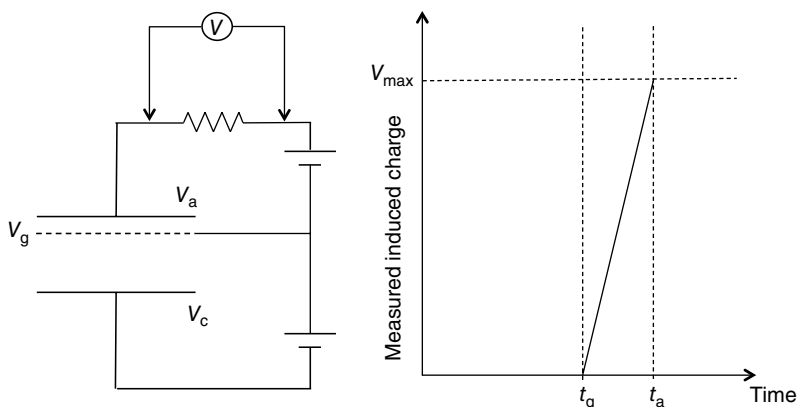


FIGURE 58.10 Schematic of a gas detector with a Frisch grid (left). The anode is held at potential V_a , the cathode at V_c , and the grid at the intermediate potential V_g . The corresponding induced charge profile shows that no charge is induced on the anode as the electron moves from creation to the grid, where it arrives at t_g (right). From there, it induces charge on the anode until it reaches its maximal value at time t_a .

58.4.3 Proportional Counters

While ionization chambers are very simple, their pulse amplitudes are very small since the only charge that can be measured is what is directly created through ionization. However, when the magnitude of the electric field in the gas detector gets large enough, it is possible that the secondary electrons created through ionization are accelerated by the field to such a point where they too can ionize gas molecules, resulting in more charge pairs being created. This process, called the Townsend avalanche, results in the multiplication in the number of charge pairs created, which results in a significantly larger signal and better overall signal-to-noise statistics. It should be noted that the ions are typically not accelerated sufficiently since the acceleration due to an electric field, a , is given by

$$\bar{a} = \frac{q\bar{E}}{m}$$

where m is the mass of the accelerated particle. So it is clear that most ions are too massive to be accelerated enough by the electric field to create further ionizations.

The minimum necessary electric field to achieve charge multiplication is a function of the gas with typical values around $5 \times 10^4 \text{ V/cm atm}$ [3]. While it is theoretically possible to achieve this magnitude of electric field in a parallel plate configuration, it is practically difficult since the necessary applied bias is so large. However, these fields can be achieved in a cylindrical geometry, as described in the

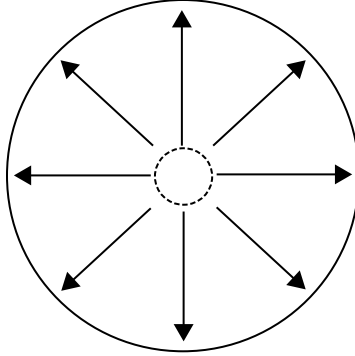


FIGURE 58.11 Illustration of the electric field inside a cylindrical detector. Note that the critical radius, r_c , where the electric field is large enough to support charge multiplication is represented by the dashed line.

previous section. In fact, if we consider how the drift velocity changes with distance, we observe

$$\bar{v} = \frac{\mu \bar{E}}{p} = \frac{\mu V}{pr \ln(b/a)}.$$

So it is evident that the electron actually increases in velocity the closer it gets to the anode, further increasing the probability of avalanche.

An example of the electric field in a cylindrical detector is shown in Figure 58.11. As is evident in the figure, the threshold for charge multiplication occurs at $r = r_c$. So an ion pair would be generated outside this volume through normal ionization and then the electron would migrate into the multiplication region as a result of the internal electric field to then be multiplied by several thousands. The net effect is a significant increase in the size of the overall signal in the detector given by

$$Q = ne_0M$$

where n electrons are initially created and M is the gas multiplication factor. Values of 10^5 – 10^6 are not uncommon for M .

It is important to realize that, despite the gas multiplication, the amount of charge that is measured is proportional to the amount of charge originally created, which is proportional to the energy deposited in the detector. This is shown in the third region of Figure 58.6. However, unlike for ionization chambers where the variance in pulse amplitude is limited to the Poissonian statistics of charge creation, in proportional counters the variance in the amount of multiplication in any given avalanche must be included. In this case, the overall statistical limit of the proportional counter can be calculated as

$$\left(\frac{\sigma_Q}{Q} \right)^2 = \frac{F}{n} + \frac{1}{n} \left(\frac{\sigma_M}{M} \right)^2$$

where n is the original number of charge carriers created through ionization and F is the Fano factor for the gas. It can be shown that this is inversely proportional to energy of the incident particle.

Due to the size of the output signal and overall ease of use, proportional counters are widely used for a variety of applications including the measurements of beta particles, low energy gamma and X-rays, and neutrons. Their use in mixed radiation fields is also common.

58.4.4 GM Detectors

GM detectors or counters have designs that are very similar to proportional counters. They are also cylindrical in geometry, although they operate at larger electric fields as indicated in the fourth region of Figure 58.6. During the acceleration in the electric field in either a proportional or a GM counter, the electron can ionize the gas molecules or it can just excite them. The excited gas will eventually deexcite through the emission of an ultraviolet photon. This photon can then trigger further ionization in the gas and more avalanches. In a proportional counter, a different gas called a quench gas is usually added in small concentrations to the mixture to absorb the UV photons. However, in a GM counter, no quench gas is added thus allowing these UV photons to create their own avalanches. Most avalanches will create at least a few photons. The result is that the UV photons create avalanches throughout the entire detector volume in what is termed the Geiger discharge.

This would be a runaway, continuous discharge if it were not for the resulting gas ions. In a GM detector, there are a significant number of ions created. Once their number reaches a critical value, the electric field they create as they move toward the cathode is enough to distort and reduce the detector's overall electric field. The reduction in electric field caused by the ions is enough to keep the electrons from being accelerated sufficiently to create an avalanche, thus terminating the discharge process. Once the ions have been collected, the detector returns to its normal operating electric field and can begin the process again.

Because the Geiger discharge continues until the ion concentration is sufficient to terminate it, the pulses created in a GM counter are all the same size since all that is required is that a specific number of ions be created. Therefore, the pulses from a GM detector cannot be used to determine the energy of the incident radiation. They are always operated as counters and are usually used to survey for the presence of radiation and to measure exposure rates.

58.4.5 Scintillators

A scintillator is any material that absorbs energy from ionizing radiation and reemits the energy in the form of light, usually in the optical or ultraviolet wavelengths. Scintillation light can be created through either fluorescence or phosphorescence, with

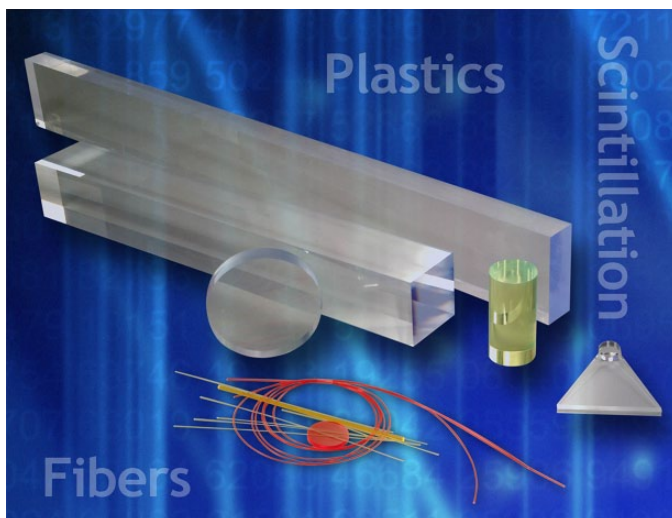


FIGURE 58.12 Example of plastic scintillators made in a variety of shapes. Source: Reproduced with permission of Saint-Gobain Crystals.

the former being the preferred method since the emission of light occurs much faster. The scintillation photons are then directly counted by a light sensor such as a photomultiplier tube (PMT) or a photodiode to create a spectrum, much in the same way that charge pairs are counted in a gas detector.

Scintillators are categorized as either organic or inorganic. Organic scintillators can be in any form—solid, liquid, or gas. They emit light through the deexcitation of an excited molecule. The most common types of organic scintillators are anthracene ($C_{14}H_{10}$), stilbene ($C_{14}H_{12}$), and polyvinyltoluene (PVT). Many plastics work well as organic scintillators. Because these types of materials can be made in many different form factors and easily produced, it is reasonably simple to mold them into desired shapes, including very large slabs and very small fibers, as shown in Figure 58.12.

Inorganic scintillators function differently than organic scintillators in that the scintillation light is emitted as the result of the deexcitation of the crystal lattice of the material. This implies that inorganic scintillators are all solids and care must be given to how they are grown. As a result, larger inorganic detectors are significantly more expensive than the comparable volume of organic scintillators. The properties of several inorganic scintillators are shown in Table 58.4 [10]. Thallium-doped sodium iodide, NaI(Tl), has long been the standard by which all inorganic scintillators are measured due to its exceptional light yield or how many scintillation photons are emitted per unit of energy deposited. However, recent research into new materials such as the lanthanum halides ($LaBr_3$ and $LaCl_3$, for example) [11–14] and elpasolite scintillators (most notably Cs_2LiYCl_6 , $Cs_2LiLaCl_6$, and $Cs_2LiLaBr_6$, called CLYC, CLLC, and CLLB, respectively) [15–17] has resulted in a significant improvement in the energy resolutions attributed to scintillators.

TABLE 58.4 Select Common Inorganic Scintillators and Their Properties^a

Scintillator	Light Yield (Photons/keV)	Light Output (%) of NaI with Bialkali PMT	Wavelength of Maximum Emission (nm)	Thickness to Stop 50% of 662 keV Photons (cm)	Density (g/cm ³)
NaI(Tl)	38	100	415	2.5	3.67
LaCl ₃ (Ce)	49	70–90	350	2.3	3.85
LaBr ₃ (Ce)	63	165	380	1.8	5.08
CsI(Na)	41	85	420	2.0	4.51
CsI(Tl)	54	45	550	2.0	4.51
BGO	8–10	20	480	1.0	7.13
CdWO ₄	12–15	30–50	475	1.0	7.9

^aData provided courtesy of Saint-Gobain Crystals.

When deciding between scintillators, it is important to consider what type of radiation is to be detected. For gamma rays, it is important that a material with high atomic number, Z , be chosen to maximize the probability of the photoelectric effect within the detector volume (which is proportional to $Z^{4.5}$). Organic scintillators have low atomic number and therefore are not usually used for gamma-ray spectroscopy. However, because of their size, they can make very large detectors that would be suitable as counters, as is often the case in portal monitoring applications. Inorganic scintillators, on the other hand, can be made with high atomic numbers, so their probability of photoelectric effect is high, but they cannot be made very large.

While the detection of neutrons is discussed in more detail in later sections, it is useful to realize that hydrogenous materials make good neutron detectors. Thus organic scintillators are frequently chosen for neutron detection.

58.4.6 Readout of Scintillation Light

Once the scintillation photons have been produced in the scintillator, they must be counted. This is usually achieved by use of a PMT. (Photodiodes can also be used, but their use is not as common as PMTs and beyond the scope of this chapter. The interested reader is encouraged to consult [3] for additional information on their operation.) The job of the PMT is to convert the scintillation light to electrons. The conversion itself happens at the surface of the PMT, which is called the photocathode. From there, the free electrons enter the tube itself and are accelerated by an electric field to a small metal dynode. If the electron is sufficiently accelerated, it has enough energy to liberate a few more electrons at this dynode. These electrons are then guided by electric field to the second dynode where they liberate a few more electrons and so on through the dynode structure until all of the electrons are finally collected at the anode. A schematic showing the internal structure of a PMT is shown in Figure 58.13.

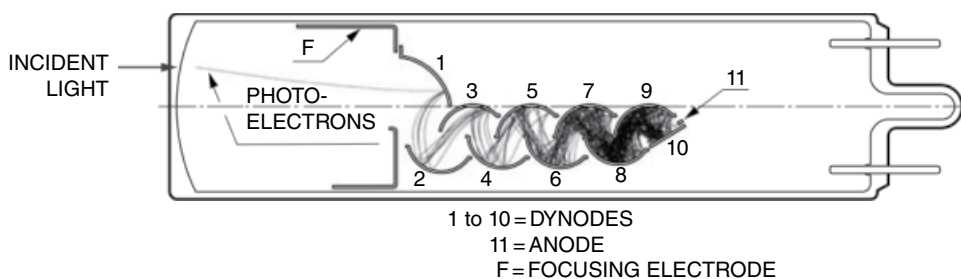


FIGURE 58.13 Schematic of a photomultiplier tube including photocathode (far left), dynodes, and anode. Source: Reproduced with permission of Hamamatsu Corporation.



FIGURE 58.14 A variety of different inorganic scintillators coupled to photomultiplier tubes. Source: Reproduced with permission of Saint-Gobain Crystals.

If an electron, on average, liberates M electrons from the surface of each dynode, then the overall gain of the PMT, G , can be approximately calculated as

$$G = M^N$$

where N is the number of dynodes. It is easy to see then that for a typical PMT with 10 stages and M on the order of 3–5, the gain can easily reach 10^6 . An example of a PMT mated with a scintillator is shown in Figure 58.14.

The selection of which PMT to use is a function of many things. First, it is ideal that the size and shape of the photocathode surface closely match the surface of the scintillator to minimize light loss. Second, the sensitivity of the photocathode is a function of wavelength, as shown in Figure 58.15. It is therefore important that the wavelength band of absorption of the PMT overlap well with the band of emission of the

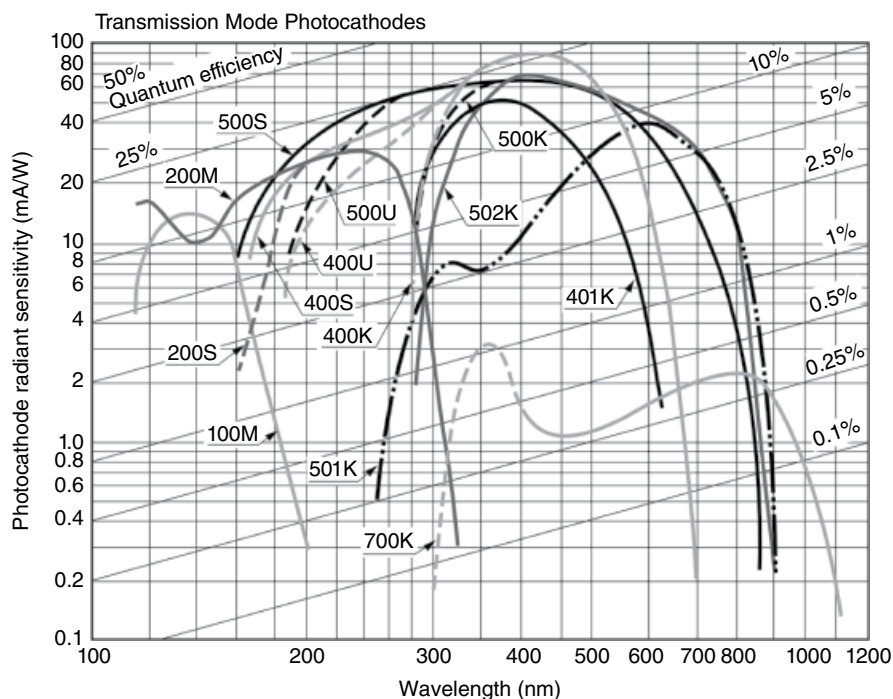


FIGURE 58.15 Sensitivity to a variety of commercial PMTs to different wavelengths of scintillation light. Source: Reproduced with permission of Hamamatsu Corporation.

scintillator. (Note that this is why scintillators like CsI(Tl) have a poor light output relative to NaI(Tl)—because the wavelength of emission is not well matched to the common bialkali PMTs used for NaI.) Lastly, the gain of the PMT is a strong function of the internal electric field uniformity. A small change in the internal electric field results in a significant change in the acceleration of the electrons and therefore seriously impacts M . The end result is that the output of the PMT has a great deal of noise. So consideration should be made on how much stability is available in the high-voltage bias supply of the PMT.

58.4.7 Semiconductors

The final type of radiation detector is the semiconductor, the physics of which is well described in many classic texts on the subject [18–20]. Semiconductors work similarly to gas detectors, where signal is created by the excitation of charge carriers to different energy bands of the crystal lattice. In many ways, this is not unlike an ionization chamber; however for semiconductors the charge carriers are electron–hole pairs. When zero energy is deposited in the semiconducting material, the electrons are bound to the atoms at energy levels determined by which shell they are bound to. This can be

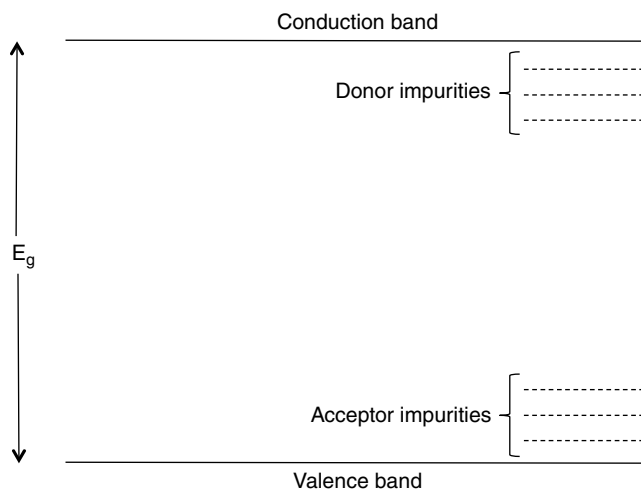


FIGURE 58.16 The band structure of a semiconductor with bandgap energy of E_g .

energetically represented by what is called the valence band. However, when energy is deposited of at least a certain minimum value, an electron from the outer shell can be liberated from the atom and is free to migrate through the material. Energetically speaking, the electron is considered to have entered the conduction band. The energy difference between the valence band and the conduction band is called the bandgap energy, E_g , and represents the required energy to create an electron–hole pair. The resulting vacancy in the valence band is a hole, which is the other charge carrier. This band structure is represented schematically in Figure 58.16.

It should be noted that the model represented in Figure 58.16 is very simplistic. In reality, semiconductor materials can be tailored by deliberately adding dopant impurities to alter the material's overall properties. These impurities can add intermediate energy levels within the bandgap, thus altering the minimum energy required to generate an electron–hole pair. The concentration of different impurity types, whether deliberately added or existing as a result of the crystal growth process, determines whether the detector is considered to be intrinsic (very low dopant concentration), n-type (high concentration of impurities with loosely bound electrons), or p-type (high concentration of impurities with holes available for electrons to occupy at much lower levels). In practice, it is ideal to combine different impurity types in different locations to form a p-n junction or diode or a p-i-n junction detector. For further details on the process of doping semiconductors and the physics resulting from the addition of impurities, the reader is encouraged to consult [19].

There are many ways to generate charge carriers in semiconductors. One common, problematic way is through thermal generation. For gas detectors, the typical ionization energies are on the order of tens of eV. However, for semiconductors, it is more common to have E_g on the order of 1–3 eV. This implies that it is possible for thermal

energy to create charge pairs. The probability of thermal generation at temperature T of a charge pair is proportional to

$$p(T) \propto T^{3/2} \exp\left(-\frac{E_g}{2kT}\right).$$

Thus it is clear that for detectors with smaller bandgap energies such as Ge require cooling in order to minimize thermal generation of charge carriers.

Ideally, the charge carrier generation will occur due to the deposition of energy from incident radiation. Contacts are applied to the surface of the semiconductor to create the anode and cathode through the application of a voltage bias. This results in the collection of the charges in a way that is similar to ionization chambers. Once the electron-hole pair are created, they are collected in the normal way by the application of an electric field between an anode and cathode. Similar to gas detectors, the drift velocity of each charge carrier can be calculated as

$$\overline{v_{e,h}} = \mu_{e,h} \bar{E}$$

where $\mu_{e,h}$ is the mobility of the electrons and holes, respectively. Unlike gas detectors where the mobility of the ions is orders of magnitude less than that of the electrons, for semiconductors the mobilities tend to be close to the same order of magnitude. Several common semiconductor radiation detectors and their properties are presented in Table 58.5.

For the detection and measurement of ionizing radiation, semiconductor devices have many benefits. One key benefit is their superior energy resolution. Assuming trapping of the charge carriers is small and the readout electronics do not contribute much noise to the overall system, the noise is just a function of charge carrier statistics. Assuming Poissonian statistics and including the Fano factor, F , the variance in the number of charge carriers, N , created during the deposition of E energy is

$$\sigma_N^2 = FN = F \frac{E}{W}$$

TABLE 58.5 Select Physical Properties of Some Common Semiconductor^a

Parameter	Si	Ge	GaAs	CdZnTe	CdTe	HgI ₂	TlBr
Density (g/cm ³)	2.33	5.33	5.32	5.78	5.85	6.4	7.56
Average atomic number	14	32	31.5	49.1	50	62	58
Bandgap (eV)	1.12	0.67	1.43	1.572	1.44	2.15	2.68
Electron $\mu\tau$ product (cm ² /V)	>1	>1	8×10^{-5}	4×10^{-3}	3×10^{-3}	3×10^{-4}	2×10^{-6}
Hole $\mu\tau$ product (cm ² /V)	~1	>1	4×10^{-6}	1.2×10^{-4}	2×10^{-4}	4×10^{-5}	2×10^{-6}
Resistivity (Ω cm)	<10 ⁴	50	10 ⁷	3×10^{10}	10 ⁹	10 ¹³	10 ¹²

^aFrom Ref. 21.

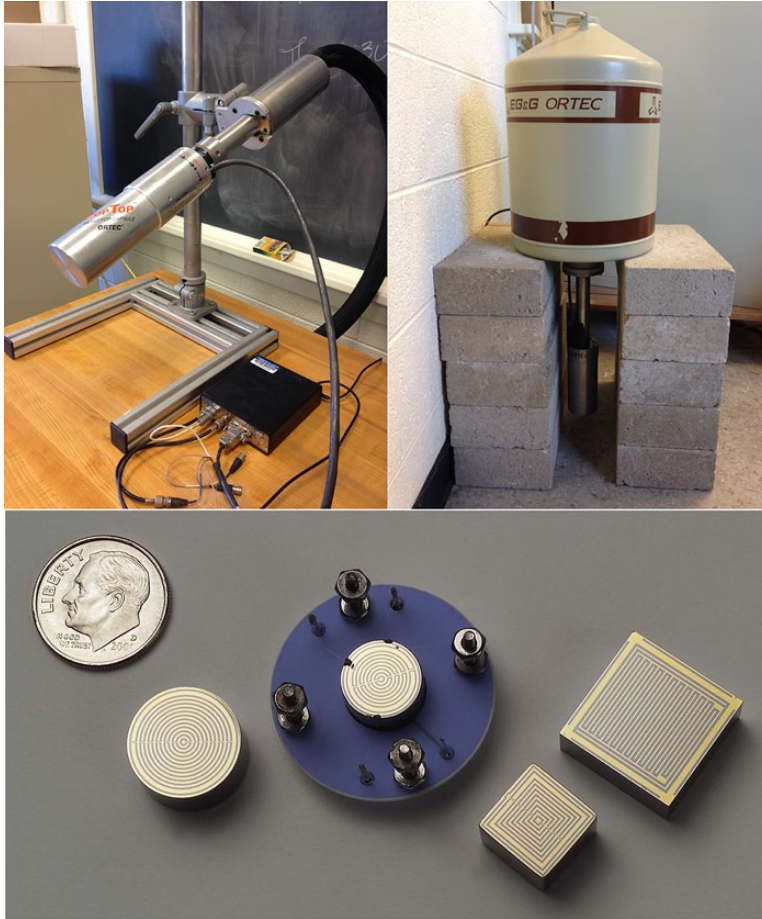


FIGURE 58.17 Examples of different types of high-purity germanium (HPGe, top row) and CdZnTe (bottom row) semiconductor devices.

where W is the average energy to generate an electron–hole pair in the medium and E is the energy deposited in the interaction. For most common semiconductors, F is on the order of 0.1 and W is around 1 eV/charge carrier [22]. Therefore, σ_N^2 is much smaller for semiconductor radiation detectors than any other detector.

Semiconductors, because of their excellent energy resolution and high density, are widely used in a number of different types of applications. They are an obvious choice for spectroscopy, both of gamma rays and charged particles. However, they also can be used as imagers, primarily in the medical industry. Most recent research has focused on identifying new materials appropriate for detectors as well as sophisticated readout techniques that mimic the Frisch grid for gas detectors [21, 23–27]. Several commercial and research-grade semiconductors are shown in Figure 58.17. A comparison of the energy spectra possible with semiconductors is shown in Figure 58.18.

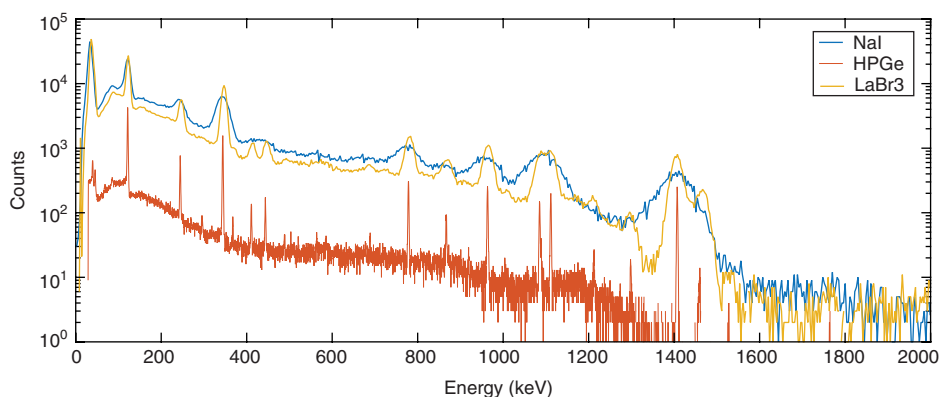


FIGURE 58.18 Comparison of the energy spectrum of Eu-152 obtained by a HPGe (top), LaBr₃ (middle), and NaI (bottom) spectrometers, illustrating the difference between high, medium, and low resolution.

58.5 NEUTRON DETECTION

Unlike other forms of radiation, neutrons do not directly ionize atoms, so any method for detecting them is based on indirect processes where the neutron creates either a charged particle or a photon that is subsequently detected. Neutrons can interact in one of six different ways that fall into the broader categories of scattering or absorption, as shown in Figure 58.19. An individual scatter cannot create a secondary particle that will generate ionization; however in most materials, a neutron will scatter many times before being absorbed.

When considering any neutron interaction, it is necessary to understand the energy of the neutron since the probability of any of these interactions (called the cross section) is strongly a function of the neutron energy. Unlike other types of radiation, the energy of a neutron source is specified based on a continuous distribution rather than a discrete energy, and all interaction probabilities are statistically derived from this. Neutrons can be thought of as an ideal gas at a given temperature, which is statistically described by the Maxwell–Boltzmann distribution shown in Figure 58.20:

$$p(E) = \frac{2\pi}{(\pi kT)^{3/2}} e^{-\frac{E}{kT}} E^{1/2}$$

where T is the temperature of the gas. Based on this distribution, the average energy of the neutron source can be calculated as

$$\bar{E} = \frac{3}{2} kT = 5.227 \times 10^{-15} \text{ v}^2$$

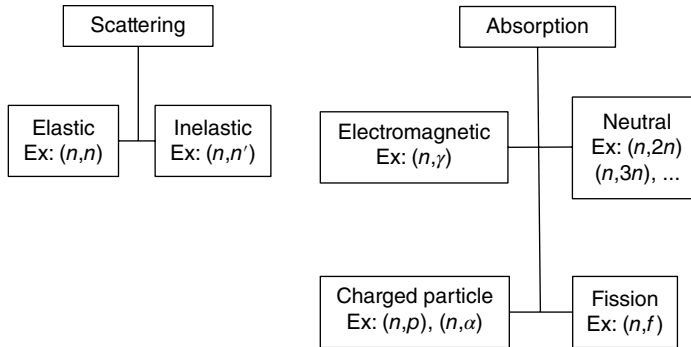


FIGURE 58.19 The six different neutron interaction mechanisms grouped into the categories of scattering versus absorption.

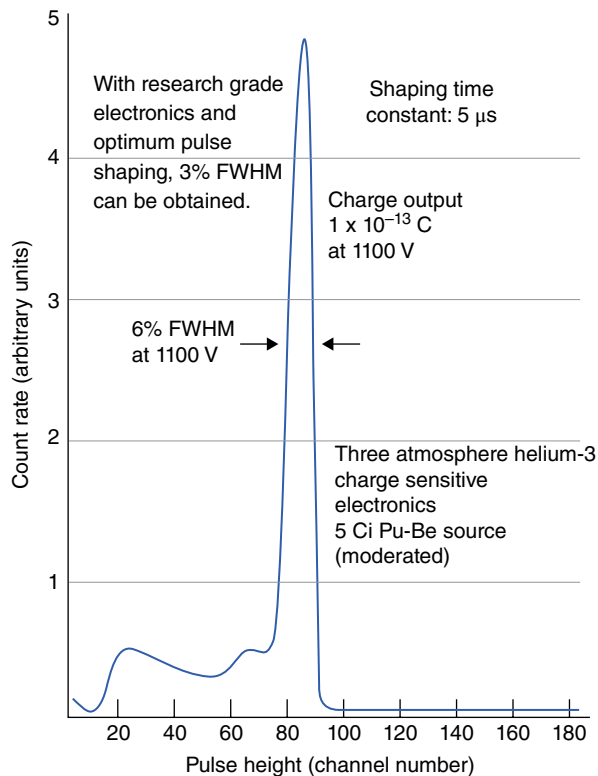


FIGURE 58.20 Sample pulse height spectrum with varying gamma-ray noise from a ^3He proportional counter, showing the peak at full Q energy deposition and the wall effect associated with partial charge collection when the proton or triton deposit a portion of their energy in the wall of the tube. Source: Reproduced with permission of GE Reuter-Stokes.

where v is the velocity of the neutron. Because of this temperature or velocity dependence on the energy distribution, neutrons are given terms like “cold,” “thermal,” “fast,” “slow,” etc. to describe their energies. For the sake of this text, we shall refer to thermal neutrons as those with \bar{E} less than around 0.5 eV whereas a fast neutron will have \bar{E} greater than 1 eV. (The intermediate region is considered to be epithermal but will not be further discussed in this chapter since the most important differences in neutron detection occur between the thermal and fast regimes.)

When an absorption interaction occurs, two daughter products are created traveling in directions opposite each other, and a certain amount of energy, Q , is liberated in the reaction. This energy is shared among the daughter products with the exact value imparted to the daughter determined by the conservation of energy and momentum. (In other words, lighter daughter products will receive more energy and heavier ones.) Note that there is no relationship between Q and the energy of the incident neutron. So a measurement of Q does not allow for determination of the energy of the neutron. Neutron spectroscopy is beyond the scope of this text, but the interested reader is encouraged to consult [7] for additional information.

58.5.1 Thermal Neutron Detection

At thermal energies there are only a few absorption reactions with a high enough cross section to allow for a reasonable detector efficiency. These are summarized in Table 58.6. The materials summarized in this table can be incorporated into any of the aforementioned types of detectors (gas, scintillator, semiconductor).

The most common type of thermal neutron detectors are proportional counters incorporating ^3He , ^{10}B , or ^6Li . The most common fill gases are ^3He and $^{10}\text{BF}_3$. In these detectors, the thermal neutron is captured through the process outlined in Table 58.6. From there, the daughter products ideally deposit all of their energy within the gas, resulting in Q energy to be measured. However, depending on where the absorption occurs, it is possible that some value less than Q will be deposited in the detector if one of the daughter products hits the wall of the detector. The deficit in energy deposited in the gas is a function of how far from the wall the daughter product is created—the closer the product is to the wall, the less energy it will deposit. The result is a

TABLE 58.6 Thermal Neutron Reaction Data^a

Reaction	Cross Section for 0.025 eV Neutrons (Barns)	Q -Value (MeV)
$^{10}\text{B}(n,\alpha)^7\text{Li}$	3840	2.792 (6%)
$^{10}\text{B}(n,\alpha)^7\text{Li}^*$	3840	2.310 (94%)
$^3\text{He}(n,p)^3\text{H}$	5400	0.764
$^6\text{Li}(n,\alpha)^3\text{H}$	937	4.78

^a From Ref. 7.

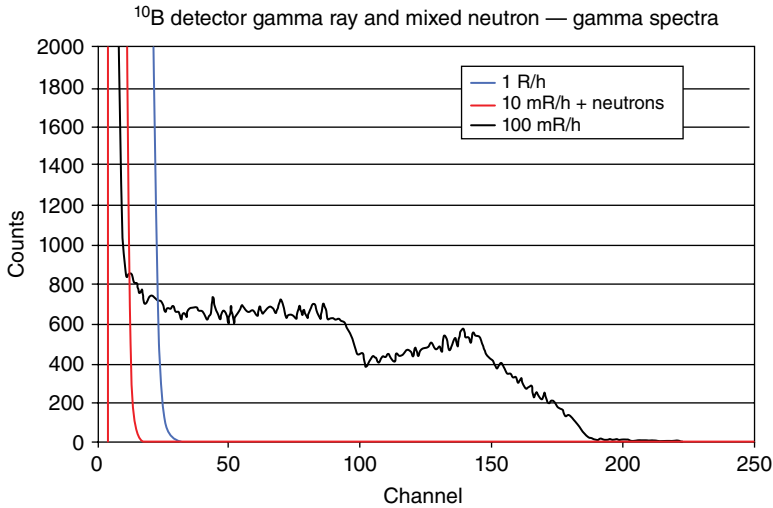


FIGURE 58.21 Example of a spectrum from a ^{10}B -lined proportional counters in the presence of increasing gamma-ray background. Note the two plateaus correspond to the wall effect from the two reaction products, ^7Li and an alpha particle, from left to right. Source: Reproduced with permission of GE Reuter-Stokes.

continuum of possible values of deposited energy for that product ranging from near the product's full value to close to zero. This is schematically represented in Figure 58.21 with the corresponding features in the spectrum present.

It is also possible to deposit the neutron-absorbing material on the wall of the detector; ^{10}B -lined proportional counters are a good example of this. This design has the advantage that the aforementioned gases, while good at absorbing neutrons, are not well suited for use in proportional counters due to low W values or poor multiplication properties. Lined proportional counters, on the other hand, are designed to have the neutron interact in the wall and then have one of the reaction products enter the gas volume where a more suitable gas is used for proportionality. In this case, it is not possible to have the full Q value deposited in the detector since only one of the two reaction products will enter the gas. Additionally, depending on how much wall thickness must be penetrated by that product before it reaches the gas, the amount of energy the product has available to deposit will range from nearly zero (if the absorption occurs far from the gas) to nearly the full product energy (if the absorption occurs very near the surface of the lining). An example spectrum illustrating this is shown in Figure 58.22.

In addition to incorporating ^3He , ^{10}B , or ^6Li into a variety of different detector configurations, the process of fission also can be used for detection through what is called a fission chamber. In these types of detectors, the wall of the detector is lined with a fissionable material, usually ^{235}U or ^{239}Pu , and a thermal neutron causes a fission in this lining. The result of this reaction are fission products, many of which carry a significant charge to them. Further, the Q value associated with fission can be extremely

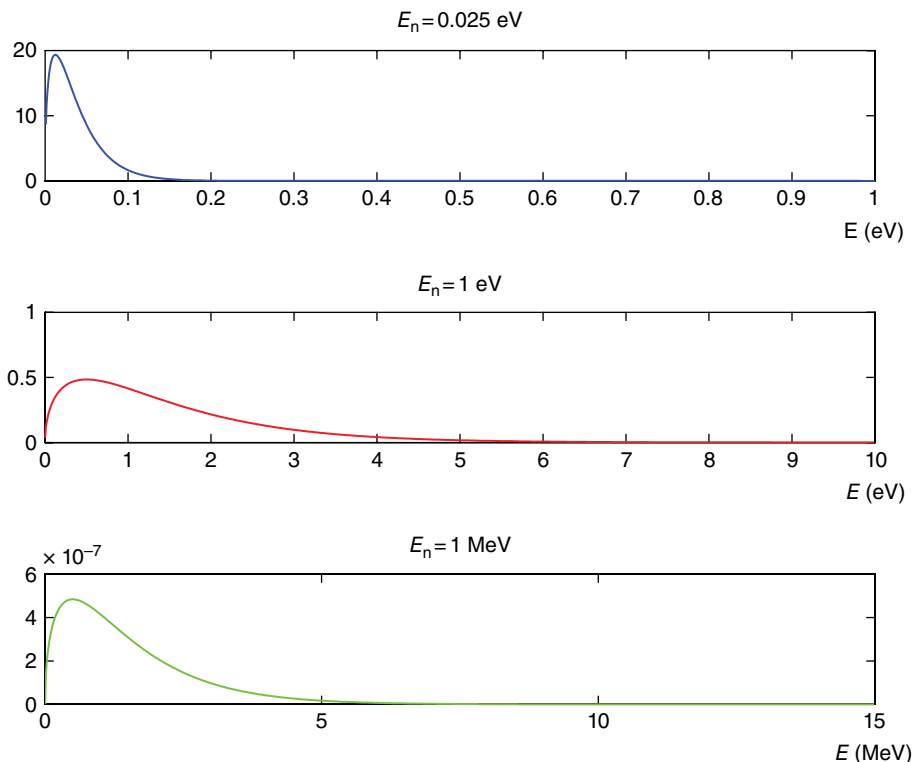


FIGURE 58.22 Example probability distribution function of the Maxwell–Boltzmann distribution for neutron sources of different energies. Note the variations in the axes values.

large—150 MeV or greater. So each fission product carries with it a great deal of energy that can be deposited in the detector. This is advantageous because of the impact on the spectrum of gamma rays. This is shown in Figures 58.21 and 58.22. With increasing gamma-ray dose rates, the lower region of the spectrum begins to overwhelm the region associated with neutron counts. This makes it difficult to set a lower energy threshold for neutron counting. However, if we consider that the peak in the spectrum is associated with full Q deposition, then it is clear that having a larger Q value results in improved threshold setting capabilities to discriminate the noise associated with gamma rays. Additionally, because the Q value is so large, these detectors can be operated as ionization chambers since the resulting charge from ionization is so large that no gas multiplication is required.

58.5.2 Fast Neutron Detection

There are a few options when fast neutrons must be detected. One option is to surround one of the aforementioned thermal neutron detectors with sufficient moderator to slow a fast neutron down to the thermal range and then detect the slower neutrons. However,

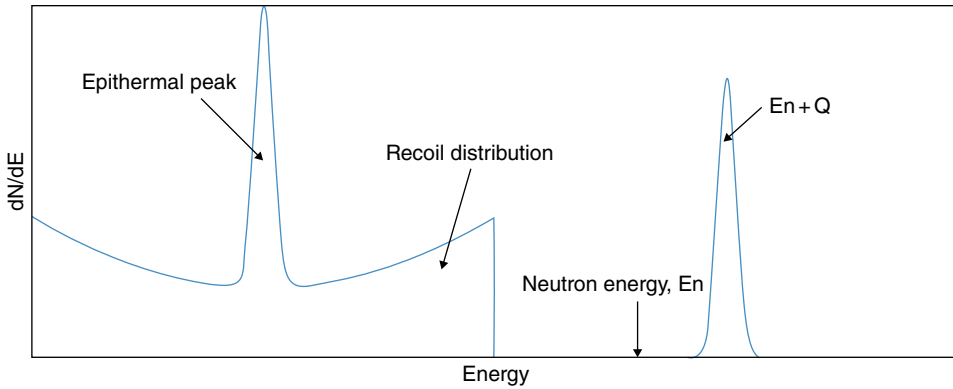


FIGURE 58.23 Sample fast neutron spectrum showing full energy deposition at $E_T = E_n + Q$ and epithermal peak at Q .

there are options for the detection of fast neutrons more directly. Note that like thermal neutrons, fast neutrons are not directly detected but are measured through their reaction products. Some of the aforementioned neutron absorption mechanisms can still be employed for fast neutron detection, albeit with significantly smaller cross sections. However, a new mechanism for fast neutron detection exists based on the increased importance of elastic scattering at fast energies.

When the energy of the neutron is comparable to or greater than the Q value of the material, it is possible to measure the energy of the incident neutron by measuring the energies of the reaction products. Hence, the information on the energy of the neutron is not lost, as it is in thermal neutron detection.

In scattering off of a light nuclei, the energy of the recoil nucleus, E_R , is given by

$$E_R = \frac{4A}{(1+A)^2} E_n \cos^2 \theta$$

where A is the mass of the target nucleus, E_n is the energy of the incident neutron, and θ is the scatter angle in the lab reference frame. There are clearly maxima in E_R depending on A and θ , which governs how to properly choose detecting material—materials with small mass have a higher amount of energy imparted to the recoil nuclei, which is ideal for detection.

The most common detectors in this energy range are proportional counters made from hydrogen, ^3He , or methane or organic scintillators, either as solid plastic or as liquid. Each different type of detector functions as described in previous sections, but they are measuring the energy deposited by the reaction products or recoil nucleus. Since the energy distribution of the recoil nuclei can be predicted, we can draw the expected energy spectrum for a fast neutron detector. When the full energy of all of the reaction products is deposited, this results in a peak in the spectrum at energy $E_n + Q$, as shown in Figure 58.23. When a portion of the recoil nucleus is measured, a

continuum is present representing the multiple possible values of energy deposited by the recoil nucleus. Lastly, a peak is present at Q called the “epithermal peak.” This corresponds to the deposition of energy by a thermal neutron, which liberates Q . Based on this, it is clear that the energy of the neutron can be inferred from the full energy peak where the total energy deposited is

$$E_T = E_n + Q.$$

However, it is important to note that this is the energy of the neutron when it reaches the detector. Neutrons are moderated by many things, especially including interactions with the environment between the source and the detector. So while a neutron source itself might be a fast source, by the time the neutron reaches the detector, it could have easily thermalized, especially if the detector is located a significant distance from the source.

58.6 CONCLUDING REMARKS

Many different types of radiation detectors have been presented in this chapter. It can be confusing to determine what type of detector to choose for a particular application. Ultimately the user needs to determine what type of radiation they are attempting to detect and what type of data they hope to obtain from the detector, be it an average dose or high-resolution spectroscopy. Detector selection is usually based on a number of different factors in addition to these, including the overall detection efficiency, resolution (if spectroscopic data is required), the detector’s physical size and operating constraints (e.g., does the detector require cryogenic cooling, does it need to operate as a handheld device with its own power supply, etc.), and cost. However, at their core, all radiation detectors perform the same function: they convert the incident radiation to charge when that radiation deposits energy in the detector. Accurate measurement of that radiation requires accurate quantification of the charge created. While current research in radiation detection focuses on creating new materials for converting deposited energy to charge or new readout and data processing algorithms, the fundamental concept of charge measurement is still the same.

REFERENCES

1. H. Cember, *Introduction to Health Physics*, 4th ed. New York: McGraw-Hill Medical, 2009.
2. P. W. Frame, “A history of radiation detection instrumentation,” *Health Physics*, vol. 88, no. 6, pp. 613–637, June 2005.
3. G. F. Knoll, *Radiation Detection and Measurement*, 4th ed. Hoboken, NJ: John Wiley & Sons, Inc., 2010.

4. C. Leroy, *Principles of Radiation Interaction in Matter and Detection*, 3rd ed. Singapore: World Scientific, 2012.
5. W. J. Price, *Nuclear Radiation Detection*. New York: McGraw-Hill, 1958.
6. D. Reilly, N. Ensslin, H. Smith, and S. Kreiner, *Passive Nondestructive Assay Manual*. Springfield, VA: National Technical Information Service, 1990.
7. N. Tsoulfanidis, *Measurement and Detection of Radiation*, 2nd ed. Washington, DC: Taylor & Francis, 1995.
8. D. J. Griffiths, *Introduction to Electrodynamics*, 4th ed. Englewood Cliffs, NJ: Prentice Hall, 2013.
9. J. D. Jackson, *Classical Electrodynamics*, 3rd ed. New York: John Wiley & Sons, Inc., 1999.
10. S. Derenzo, M. Boswell, M. Weber, and K. Brennan, "Scintillation Properties." [Online]. Available: <http://scintillator.lbl.gov/> (Accessed: February 14, 2014).
11. D. Alexiev, L. Mo, D. A. Prokopovich, M. L. Smith, and M. Matuchova, "Comparison of $\text{LaBr}_3\text{:Ce}$ and $\text{LaCl}_3\text{:Ce}$ with NaI(Tl) and cadmium zinc telluride (CZT) detectors," *IEEE Transactions on Nuclear Science*, vol. 55, no. 3, pp. 1174–1177, June 2008.
12. R. González, J. M. Pérez, O. Vela, and E. de Burgos, "Performance comparison of a large volume CZT semiconductor detector and a $\text{LaBr}_3\text{(Ce)}$ scintillator detector," *IEEE Transactions on Nuclear Science*, vol. 53, no. 4, pp. 2409–2415, August 2006.
13. W. M. Higgins, J. Glodo, E. Van Loef, M. Klugerman, T. Gupta, L. Cirignano, P. Wong, and K. S. Shah, "Bridgman growth of $\text{LaBr}_3\text{:Ce}$ and $\text{LaCl}_3\text{:Ce}$ crystals for high-resolution gamma-ray spectrometers," *Journal of Crystal Growth*, vol. 287, no. 2, pp. 239–242, January 2006.
14. K. S. Shah, J. Glodo, M. Klugerman, L. Cirignano, W. W. Moses, S. E. Derenzo, and M. J. Weber, " $\text{LaCl}_3\text{:Ce}$ scintillator for γ -ray detection," *Nuclear Instruments & Methods in Physics Research Section A*, vol. 505, no. 1/2, p. 76, June 2003.
15. B. S. Budden, L. C. Stonehill, J. R. Terry, A. V. Klimenko, and J. O. Perry, "Characterization and investigation of the thermal dependence of $\text{Cs}_2\text{LiYCl}_6\text{:Ce}^{3+}$ (CLYC) waveforms," *IEEE Transactions on Nuclear Science*, vol. 60, no. 2, pp. 946–951, April 2013.
16. J. Glodo, R. Hawrami, and K. S. Shah, "Development of $\text{Cs}_2\text{LiYCl}_6$ scintillator," *Journal of Crystal Growth*, vol. 379, pp. 73–78, September 2013.
17. J. Glodo, E. van Loef, R. Hawrami, W. M. Higgins, A. Churilov, U. Shirwadkar, and K. S. Shah, "Selected properties of $\text{Cs}_2\text{LiYCl}_6$, $\text{Cs}_2\text{LiLaCl}_6$, and $\text{Cs}_2\text{LiLaBr}_6$ scintillators," *IEEE Transactions on Nuclear Science*, vol. 58, no. 1, pp. 333–338, February 2011.
18. C. Kittel, *Introduction to Solid State Physics*, 3rd ed. New York: John Wiley & Sons, Inc., 1966.
19. G. Lutz, *Semiconductor Radiation Detectors Device Physics*. Berlin: Springer, 2007.
20. S. M. Sze, *Physics of Semiconductor Devices*, 3rd ed. Hoboken, NJ: Wiley-Interscience, 2007.
21. A. Owens and A. Peacock, "Compound semiconductor radiation detectors," *Nuclear Instruments and Methods in Physics Research Section A*, vol. 531, no. 1–2, pp. 18–37, September 2004.

22. M. Harrison, "Fano factor and nonuniformities affecting charge transport in semiconductors," *Physical Review B*, vol. 77, no. 19, 2008.
23. F. Zhang, Z. He, D. Xu, G. F. Knoll, D. K. Wehe, and J. E. Berry, "Improved resolution for 3-D position sensitive CdZnTe spectrometers," *IEEE Transactions on Nuclear Science*, vol. 51, no. 5, pp. 2427–2431, October 2004.
24. Z. He and B. W. Sturm, "Characteristics of depth-sensing coplanar grid CdZnTe detectors," *Nuclear Instruments & Methods in Physics Research Section A*, vol. 554, no. 1–3, pp. 291–299, December 2005.
25. A. Owens, *Compound Semiconductor Radiation Detectors*. Boca Raton, FL: Taylor & Francis, 2012.
26. P. J. Sellin, "Recent advances in compound semiconductor radiation detectors," *Nuclear Instruments and Methods in Physics Research Section A*, vol. 513, no. 1–2, pp. 332–339, November 2003.
27. Y. Zhu, S. E. Anderson, and Z. He, "Sub-pixel position sensing for pixelated, 3-D position sensitive, wide band-gap, semiconductor, gamma-ray detectors," *IEEE Transactions on Nuclear Science*, vol. 58, no. 3, pp. 1400–1409, June 2011.

MEASURING TIME AND COMPARING CLOCKS

JUDAH LEVINE

Time and Frequency Division and JILA, NIST and the University of Colorado, Boulder, CO, USA

59.1 INTRODUCTION

Time and time interval have played important roles in all societies since antiquity. The original definitions were based on astronomy: the solar day and year and the lunar month were widely used as measures of both time and time interval. As I will show in Section 59.15, the strong connection between astronomy and time persists even today, when both time and time interval are measured by means of clocks. I will begin by describing a generic clock, and I will then discuss various means of comparing these devices and characterizing their performance using a combination of deterministic and stochastic parameters. I will conclude with a short description of calibrating them in terms of international standards of time and frequency.

59.2 A GENERIC CLOCK

All clocks consist of two components: a device that produces or observes a series of periodic events and a counter that counts the number of events and possibly also interpolates between consecutive events to improve the resolution of the measurement. The choice of which periodic event to use as the reference period for the clock plays a

fundamental role in determining its performance so that it is natural to characterize a particular clock design based on an evaluation of the reference period that it uses to drive its counter.

In addition to the two components discussed in the previous paragraph, real clocks and time scales have a time origin that is derived from some external consideration. As a practical matter, the time origin is generally chosen to be sufficiently far in the past so that most epochs of interest have positive times with respect to the origin.

In addition to a time origin, real time scales are used to construct a calendar—an algorithm that assigns names to clock readings. These considerations are very important but are mostly outside of the scope of this discussion. Although I will not discuss the methods used to implement a calendar, I will discuss the methods that are currently used to define Coordinated Universal Time (UTC) and the discussions that are currently underway (as of 2016) about possibly modifying the definition of this time scale.

Two distinct parameters are important in characterizing the frequency reference of any clock: (i) The accuracy of the reference period—how closely does the period conform to the definition of the second. (ii) The stability of the reference period over both the short and long terms. (Stability is a necessary prerequisite for an accurate device but is not sufficient, and it is quite common for real oscillators to have a stability that is significantly better than the accuracy.) A number of methods have been developed for characterizing the stability of periodic events, and I will briefly describe the tools that implement these methods in the next section. I will discuss the question of accuracy in a subsequent section.

59.3 CHARACTERIZING THE STABILITY OF CLOCKS AND OSCILLATORS

The methods that are used for these purposes fall into two general classes: methods that characterize the worst-case performance of a clock and methods that characterize the average performance using statistical parameters derived from a root-mean-square (RMS) calculation. In both cases, the analysis is based on a finite-length data set.

A worst-case analysis is usually sensitive both to the length of the data set that is analyzed and to the exact interval of the observation because large glitches usually occur sooner or later, and a data set either includes a large glitch or it doesn't. We might expect that the results of a worst-case analysis would show a large variation from sample to sample for this reason.

A statistical analysis, on the other hand, assumes implicitly that the data are stationary so that neither the interval of observation nor the length of the data set is important in principle. A statistical analysis tends to attenuate the effect of a glitch, since even a large glitch may have only a small impact on an ensemble-average value. More generally, a statistical analysis is not suitable if the data are not stationary, since

ensemble-average values will exist in a formal sense but will not be very useful in understanding the performance of the actual device.

In order to characterize the stability of a device under test, we can imagine that it is compared to a second clock that is perfect. That is, the perfect clock produces “ticks” that are exactly uniform in time. The interval between ticks of the perfect clock is τ so that the ticks occur at times $0, \tau, k\tau, (k+1)\tau$, etc., where k is some integer. (As I have discussed in Section 59.2, the times are relative to some origin that is defined outside of the measurement process, and the time of the first tick is 0 with respect to that origin. This is not a limitation in the current discussion, since the origin is simply an additive constant that is not important when discussing time stability.) The time of the clock under test is read each time the standard clock emits a tick, and the time differences are x_k, x_{k+1}, \dots , where x_k is a short-hand notation for the time-difference reading at time $k\tau$, etc. In general, the units of time are seconds and fractions of a second. The “frequency” of a clock is the fractional frequency difference between the device under test and a perfect clock operating at the same nominal frequency. For example, the frequency of the device under test, f , which generates a signal at a physical frequency of F , is

$$f = \frac{F - F_o}{F_o}, \quad (59.1)$$

where F_o is the output frequency of the perfect device used in the comparison. With this definition, the frequency of a clock is a dimensionless parameter, and the time difference after a time τ has changed by $f\tau$.

59.3.1 Worst-Case Analysis

Analyzing the time-difference data from the perspective of a worst-case analysis sounds easy. We simply look for the largest absolute value of x in the data set, assuming that the device under test and the standard reference device were set to the same time at the start of the measurement. A statistic that realizes this idea is the maximum time-interval error (MTIE) [1], which is usually calculated as the difference between the largest and smallest time differences in the ensemble of measurements. (With this definition, a device that has an arbitrarily large and constant time difference has an MTIE value of 0, because MTIE is a measure of the evolution of the time difference, not the magnitude of the time difference itself. In this respect, the MTIE statistic is really a measure of the frequency offset between the device under test and the standard reference.) It is commonly used to characterize the oscillators in telecommunications networks.

The MTIE statistic depends both on frequency accuracy and frequency stability, since a clock with a frequency offset with respect to the reference device will eventually produce unbounded time differences even if the frequency offset is absolutely

stable. As we mentioned before, the results of a worst-case analysis can show large variations from one set of data to another one so that the MTIE statistic is normally defined as the largest MTIE value over all data sets of some fixed length within some larger time interval, T [2]. In an extreme case, the value of T is the life of the device so that a device that exhibited one glitch after years of faithful service might not satisfy a specification based on MTIE in this case. This form of the definition is therefore unambiguous but not very useful.

An alternative view is to consider the MTIE value obtained from a single data set to be an estimate of the underlying “true” value of MTIE, which is characterized by the standard statistical parameters of a mean and a standard deviation. The standard statistical machinery can then be used to provide an estimate of the probability that the observed value is (or is not) consistent with a mean value for MTIE that is specified by some required level of performance. Since this analysis method characterizes MTIE as a statistical parameter, it usually requires some ancillary assumption about handling measurements that are not consistent with the mean and standard deviation of the distribution of the remainder of the observations. In other words, is a large outlier (i) treated as an error that should be ignored, (ii) accepted as a low-probability event that is consistent with the mean and standard deviation deduced from previous data, or (iii) an indication that the mean or the standard deviation should be updated by including this new observation? One solution is to completely reject data that differ from the mean by more than four standard deviations and to provisionally accept data that differ by more than three but less than four standard deviations. The data in the provisional category are compared to subsequent values, and all of these newer data may be used to provide an update to the mean or to the standard deviation as appropriate. The specific algorithm that is used is based more on administrative considerations and experience than on a rigorous statistical analysis, since errors are generally not statistical events by definition.

The evolution of the time-difference data that are the input to an MTIE calculation is sensitive both to the frequency stability of the device under test and to the sampling interval, τ , and the data become increasingly insensitive to fluctuations in the frequency of the device under test with respect to the reference device that are much shorter than the sampling period. A fluctuation in the offset frequency whose period is an exact sub-multiple of the sampling period has no effect on MTIE. Therefore, the sampling period must be short enough so that these shorter-period fluctuations either are not important in the application supported by the device or are known to be small a priori. Since the length of a data set is limited in practice, this consideration implies a trade-off between the shortest and longest frequency fluctuations that can be estimated from a set of measurements. However, a measurement that uses a longer sampling interval to detect longer-period fluctuations must guarantee (by means of digital or analog filtering) that the shorter-period fluctuations are not aliased by the longer sampling period.

Another way of addressing the aliasing problem is to use a very rapid sampling period (so as to minimize or eliminate the impact of aliasing) but to acquire these

measurements in blocks separated by dead time in which the clock is not observed. This method will also have a potential aliasing problem for frequency fluctuations that are synchronous with the sum of the sample period and the dead time. This problem can be addressed by varying the dead time in a pseudorandom way, but this complicates the analysis somewhat, since the blocks are no longer equally spaced in time.

I will now describe the statistical estimates of time and frequency that are commonly used to characterize clocks and oscillators outside of the telecommunications domain. I will limit my discussion to the original Allan variance, since its significance can be explained intuitively. The more complicated versions of the Allan variance and the characterizations in the frequency domain using Fourier analysis are described in the literature. (see Ref. [3].)

59.3.2 Statistical Analysis and the Allan Variance

The statistical analysis starts with the same time differences that we discussed in the previous sections. The average frequency of the clock under test with respect to our perfect clock over the time interval τ between measurements is estimated as

$$y_k = \frac{x_k - x_{k-1}}{\tau}. \quad (59.2)$$

The numerator and denominator on the right-hand side of Equation 59.2 have the units of time so that the frequency defined by this equation is a dimensionless quantity. If the device under test had a frequency that was constant with respect to the perfect clock, then Equation 59.2 would give the same result for any value of k . (Note that the device under test need not have the *same* frequency as the perfect clock. We would get the same result for every value of k even if the frequency difference was *any* constant value.)

Real event generators are not perfect, and it is useful to characterize their performance by means of the estimator

$$y_{k+1} - y_k = \frac{x_{k+1} - 2x_k + x_{k-1}}{\tau}. \quad (59.3)$$

Equation 59.3 gives the difference in the frequency of the device under test between two consecutive, equal measurement intervals with no intervening dead time. This estimator provides an estimate of frequency stability—not frequency accuracy. From the perspective of the measurement at time τ_k , this statistic is an estimate of the time difference that will be observed at the next measurement time with index $k+1$ based on the evolution of the time difference in the time interval ending at index k . Its magnitude is not sensitive to a constant time difference or frequency difference between the

device under test and the perfect reference device. (Compare this to MTIE, which was discussed in the previous section and which is sensitive to a constant frequency difference but not to a constant time difference.)

The final step in the definition of the estimator is to assert (or to hope) that the variations estimated by Equation 59.3 are stationary. That is, the computation does not depend in a systematic way on the value of the index k —any choice of k would produce a value that is consistent (in a statistical sense) with the result for any other choice of k . Then, the RMS of Equation 59.3 has a well-defined value and that RMS value has an associated, well-defined, standard deviation. When various normalizing constants are added, the mean square value of Equation 59.3 estimated over all possible values of k is the two-sample or Allan variance for an averaging time of τ , and the RMS value is the two-sample or Allan deviation for that averaging time. If there are N time-difference data with indices 1, 2, ..., N , then the Allan variance at averaging time τ is defined as the average of the $N-2$ calculations as

$$\sigma_y^2(\tau) = \frac{1}{2(N-2)\tau^2} \sum_{k=2}^{N-1} (x_{k+1} - 2x_k + x_{k-1})^2, \quad (59.4)$$

and the Allan deviation is the square root of this value. Since the time-difference data are equally spaced in time, we can use the same data to compute the estimate of the Allan variance for any multiple of the sampling interval, τ :

$$\sigma_y^2(m\tau) = \frac{1}{2(N-2m)(m\tau)^2} \sum_{k=2}^{N-2m+1} (x_{k+2m-1} - 2x_{k+m-1} + x_{k-1})^2. \quad (59.5)$$

The normalization is defined so that the Allan variance has the same value as the classical variance in the case of a random time-difference noise process, which we will discuss later. This type of process is often called “white phase noise.” The number of terms that contribute to the sum in Equation 59.5 decreases as m is made larger so that the estimates for large values of m are likely to exhibit more variation from one set of observations to another one. The maximum value of m that is used in Equation 59.5 is often limited to $N/3$ for this reason.

It is important to emphasize that the two-sample Allan deviation is a measure of frequency *stability*—not frequency accuracy. Frequency stability is obviously a very desirable quality for a clock, and many real-world devices are characterized in this way. It is clearly not sufficient to characterize devices that are used to provide standards of time, time interval, or frequency.

A very powerful technique is to examine the dependence of the Allan variance on the averaging time, τ , because this dependence can provide insight into the noise processes that drive the magnitude of the Allan deviation. To take a simple example, suppose that the device under test has a true constant frequency offset with respect to the perfect device used for the calibration. If this constant fractional frequency offset

is f , then the measured time differences at times $k\tau$ in the absence of any noise processes would be

$$x_k = fk\tau + x_0, \quad (59.6)$$

where x_0 is the time difference between the device under test and the perfect clock when $k=0$. Then Equation 59.2 would estimate the average frequency as

$$y_k = \frac{x_k - x_{k-1}}{\tau} = f, \quad (59.7)$$

which is independent of k so that the estimate of the Allan variance is 0 for all averaging times. As expected, the Allan variance is 0 for a clock with a constant offset frequency, and it provides no information on the magnitude of this frequency.

A more interesting example is to suppose that the device under test had a constant frequency offset as in the previous example but that the time-difference measurements are affected by a random noise process that might originate in the measurement hardware and not in the clock itself. The time-difference measurements in this case would be given by

$$x_k = fk\tau + x_0 + \varepsilon_k, \quad (59.8)$$

where the noise contribution is characterized by a zero-mean signal with a well-defined variance:

$$\begin{aligned} \langle \varepsilon_k \rangle &= 0 \\ \langle \varepsilon_j \varepsilon_k \rangle &= \sigma^2 \delta(j - k) \end{aligned} \quad (59.9)$$

and δ is the Dirac delta function, which is one if its argument is zero and zero otherwise. The deterministic contributions to the summation in Equation 59.4 or 59.5 cancel as in the previous example, and what is left is a sum that is proportional to the variance of the noise process but independent of the summation index, k . The Allan variance therefore decreases as the reciprocal of the square of the sampling interval, and this conclusion does not depend on the magnitudes of either the deterministic or stochastic contributions, provided only that the data satisfy Equations 59.8 and 59.9. A plot of the logarithm of the Allan variance as a function of the logarithm of the measurement interval would have a slope of -2 . If the frequency of the oscillator was not exactly constant but varied with the index k , then the summations in Equation 59.4 or 59.5 will contain a contribution from the deterministic terms in the time differences that is some function of the interval between the measurements. The log-log plot will have a slope that is greater than -2 , and the exact value of the slope will depend on the details of the noise process that is driving the frequency fluctuations. In general, the slope of the log-log plot of the Allan variance as a function of the measurement interval is an

indicator of the type of noise process that is contributing to the measured time differences. The details of this relationship and the usefulness of other variances that are related to the simple Allan variance discussed here are described in the literature [3].

The frequency dispersion estimated by the Allan deviation generates a corresponding time dispersion. The time dispersion is usually called $\sigma_x(\tau)$, and it is formally defined in terms of the modified Allan variance, which is described in Ref. [3].

We can provide a simple estimate of the time dispersion in the case of the simple white phase noise example that we considered in the previous paragraph. If we correct the measured time differences by the deterministic Equation 59.6, the residual time dispersion for any measurement is simply ε_k , a random process defined by Equations 59.8 and 59.9. If we substitute Equations 59.8 and 59.9 into Equation 59.4, then the summation is simply a sum of $N-2$ identical terms and

$$\sigma_y^2(\tau) = 3 \frac{\langle \varepsilon_k^2 \rangle}{\tau^2} \quad (59.10)$$

or

$$\sqrt{\langle \varepsilon_k^2 \rangle} = \frac{\sigma_y(\tau)\tau}{\sqrt{3}}, \quad (59.11)$$

which relates the statistical RMS time dispersion to the Allan deviation for the case of white phase noise. Since the Allan deviation for white phase noise varies as the reciprocal of the measurement interval, the estimate of the RMS time dispersion does not depend on the interval between measurements. This is not a surprising result, since it follows directly from the assumption of the statistics of the noise process defined in Equation 59.9. A more conservative estimate is to take the RMS time dispersion as simply the product of the time interval between two measurements and the Allan deviation at that time interval, and I will generally use this more conservative estimate in the following discussion, since the constant in the denominator of Equation 59.11 is valid only for white phase noise.

59.3.3 Limitations of the Statistics

Both MTIE and the Allan variance (including a number of variants that we have not discussed) are measures of stability, but they define stability in different ways. Neither statistic is sensitive to a constant time offset, but the value of MTIE is sensitive to a constant frequency offset, whereas the value of the Allan variance is not. Therefore, the Allan variance is a measure of the predictability of the future time difference of a clock based on its past performance, and arbitrarily large *constant* frequency offsets do not degrade the prediction.

However, a clock with some other *deterministic (but not constant)* frequency offset produces time-difference values that are just as well determined as a clock with only a constant frequency offset, but the Allan variance treats the two very differently. The frequencies of many types of oscillators (e.g., hydrogen masers) can be approximated

as varying linearly with time, and the Allan variance of the time-difference measurements from such devices (which have a quadratic dependence on the time) does not give a realistic estimate of their stability if by stability we mean how well can a future time difference be estimated based on previous performance. For this reason, it is common to estimate and remove a quadratic function of the time from these data before they are analyzed to compute the Allan variance. This process of “prewhitening” the data is well known in the statistical literature and is often implemented using a Kalman filter [4] and in power-spectral analysis [5].

Many oscillators have frequencies whose fluctuations can be more easily (and intuitively) characterized in the Fourier frequency domain rather than as a stationary process in the time domain, which is a basic assumption of the Allan variance. For example, an oscillator whose frequency was sensitive to ambient temperature could be expected to exhibit a diurnal frequency variation if it was operated in an environment without tight control of the ambient temperature. The Allan variance calculation will model these time differences as a stationary noise process, and it usually reports a large value at a time interval corresponding to roughly one-half of the period of the driving process. This is not too difficult to interpret correctly if there is only one such contribution, but it can be ambiguous if there are several “bright lines” in the power spectrum of the time differences, and prewhitening the data is particularly important in this case.

Finally, it is important to keep in mind that the calculation of the Allan variance is based on a particular method for averaging the time differences. Therefore, while the *slope* of a log–log plot of the Allan deviation as a function of averaging time provides insight into the underlying noise processes that drive the time differences, the *value* of the Allan deviation at any averaging time is useful as an indicator of the performance to be expected from the device only if the clock is used in a manner that is consistent with the averaging procedure that is part of the definition. For example, the definition of the Allan variance that we have used is based on data that are equally spaced in time with no dead time between the measurements. There are other statistics that are related to the simple Allan variance we have discussed (the modified Allan variance (Mod Avar), as an example, and its close relative the time variance (TVAR)), and these statistics have more complex averaging procedures, which are less likely to be used in a real application. Although the simple Allan variance has some limitations in identifying the underlying noise type in some circumstances, its definition is often closer to the way a device will actually be used, and it is often the preferred analysis tool for this reason.

59.4 CHARACTERISTICS OF DIFFERENT TYPES OF OSCILLATORS

The frequency stability of an oscillator can be realized either actively, where the discriminator is actively oscillating at a resonant frequency derived from its characteristics, or passively, where the frequency of an oscillator is locked to the resonant response of a passive discriminator.

A quartz crystal oscillator is generally an active device, because the frequency is determined by the mechanical resonance in the quartz that is excited by an external power source. Atomic frequency standards fall into both categories. Atomic clocks that use a transition in cesium or rubidium as the frequency reference generally fall into the second category, where the discriminator is passive and is interrogated by a separate oscillator that is locked to the peak of the transition probability of the clock transition. Hydrogen masers can be either active or passive. In both types of devices, the actual output frequency is generally a function both of the resonant frequency of the discriminator and the method that is used to interrogate it. The parameters of the interrogation method may have a dependence on ambient temperature or may vary with time so that even nominally identical oscillators generally have different output frequencies. (A primary frequency standard is designed so that these perturbing influences are minimized, and the residual perturbations are estimated by means of various ancillary measurements.)

The vast majority of oscillators currently in use are stabilized using a mechanical resonance in a quartz crystal. Newer devices are stabilized using a microelectromechanical (MEMS) device [6]. The quartz crystals used as the frequency reference in inexpensive wristwatches can have a frequency accuracy of 1 ppm and a stability of order 10^{-7} . (A frequency accuracy of 1 ppm translates into a time dispersion of order 0.09 s/day.) These devices are generally sensitive to the ambient temperature so that substantially better performance can be realized using active temperature control. The moderate stability of the frequency is exploited in many control applications and in the operation of Internet time servers, as we will discuss later. In spite of some very clever techniques that have been developed to improve the long-term stability of the frequency of these devices, none of the methods can totally eliminate the sensitivity of the frequency to environmental perturbations, to stochastic frequency fluctuations that are hard to model, and to a dependence on the details of manufacture that are hard to replicate.

Atomic frequency standards use an atomic or molecular transition as the frequency discriminator in an attempt to address the limitations of the frequency stability of mechanical devices that I discussed in the previous paragraph. The atoms can be used in a passive or active configuration. In the passive configuration, the atoms are prepared in the lower state of the clock transition and are illuminated by the output from a separate variable-frequency oscillator. The frequency of the oscillator is locked to the maximum in the rate of the clock transition. There are a number of different methods for preparing the atoms in the lower state and for detecting the transition to the upper state. The details are described in the literature [7].

In the active configuration, the atoms are pumped into the upper state, and the radiation emitted when they decay to the lower state stimulates other atoms to decay by stimulated emission. The oscillating frequency is determined by the atomic transition frequency and by the properties of the cavity that is used to trap the radiation and provide the ambient field that induces stimulated emission. Most lasers and some hydrogen

masers work this way. The cavity of an active hydrogen maser is generally tuned to improve the stability of the output frequency [8].

In both active and passive hydrogen masers, the output signal is generated by an oscillator (typically a quartz crystal device) whose frequency is locked to the transition frequency of the atoms. For sufficiently short averaging times (typically less than about 0.1 s), the stability of the output frequency is determined by the properties of the quartz oscillator and by the process used to generate the output ticks. The spectrum is generally white phase noise. At longer averaging times (greater than about 10 s), the stability of a maser is generally limited by the thermal noise in the oscillating field in the cavity and in the control loop that is used to lock the output oscillator. This is usually white frequency noise. Other types of atomic standards are generally operated in the passive configuration and have similar statistics.

59.5 COMPARING CLOCKS AND OSCILLATORS

Comparing the times of two clocks is a simple process in principle, but the process becomes more complicated as the resolution of the measurement increases. I will discuss two classes of comparisons. In the first situation, the devices to be compared are in the same laboratory so that the signals from both clocks are available locally, and we don't have to consider the characteristics of a transmission network. In the second configuration, one of the clocks is at a remote location so that the statistics of the time comparison must include the effects of the transmission network.

The simplest time comparison is simply a one-time measurement. We read the time difference between the two clocks, using a time-interval counter, for example, and we use the measurement to adjust the time of one of them either by making an adjustment to its physical output or by noting its time offset for future administrative corrections. The implication of the process is that the reference clock is much more accurate than the device whose time difference we are measuring, and we need not consider the possibilities that the reference clock is broken or that the time comparison had a significant measurement error. Furthermore, this simple "set it and forget it" process does not provide any insight into the statistics of the device under test so that we have no way of knowing how rapidly its time will diverge from the correct time after it has been set. Thus, we have no way of knowing how often to repeat the measurement process in order to have the device being calibrated maintain some specified level of accuracy. These considerations lead to a more sophisticated measurement program.

To simplify matters, we will again assume that the clock under test is being compared to a second clock that is so much more accurate that it can be considered as perfect from the perspective of the measurement process. We will model the time differences of the clock under test by a combination of deterministic and stochastic parameters. The deterministic time differences are generally specified in terms of three parameters, the initial time difference, the frequency offset, and the frequency aging.

This formulation leads to a quadratic relationship with constant parameters whose independent variable is the elapsed time since the start of the measurement. This formulation is not adequate for most real devices and measurement processes.

In the first place, the frequency offset and frequency aging are not manifest constant parameters for any real device, and treating them as constants is neither adequate nor optimum. In addition, many applications depend on real-time estimates of the parameters, and it becomes increasingly cumbersome to reevaluate a static quadratic form of the modeled time differences each time a new data point is measured. This type of analysis also requires that we save all of the measurements since the start of the experiment. Therefore, most analyses use an iterative form of the estimate, in which the current time difference is modeled based on the parameters estimated from the previous measurements. The previous measurements themselves are not needed.

In this method, we estimate the time difference at time t_k in terms of the previous data as

$$\hat{x}_k = x_{k-1} + y_{k-1}\Delta t + \frac{1}{2}d_{k-1}(\Delta t)^2, \quad (59.12)$$

where x_k , y_k , and d_k are the time difference, frequency offset, and frequency aging at time t_k and $\Delta t = t_k - t_{k-1}$. It is generally easiest to use measurements that are equally spaced in time, but the formulation of Equation 59.12 is valid whether or not this is the case.

The measurement process measures the time difference at time t_k and returns the measured value X_k . The difference between the value we predicted and the value we observed is

$$\delta_k = X_k - \hat{x}_k. \quad (59.13)$$

In the absence of any noise contributions and assuming that the initial time difference, the frequency offset, and the frequency aging are perfectly constant values, Equation 59.12 has only three parameters so that the time differences in Equation 59.13 will converge to zero after three measurement cycles even if we are totally ignorant of the initial values of these parameters. But now we return to the real world, where the measurements have a stochastic component and the clock parameters are not absolutely constant. The goal of a real-world measurement process is to partition the time differences obtained in Equation 59.13 into a deterministic portion, which we use to update our estimates of the parameters in Equation 59.12, and a stochastic contribution, which we attenuate by averaging or ignoring completely.

The measurement process in the previous paragraph cannot succeed in the most general case because there is only one observable (Eq. 59.13) and multiple parameters that must be determined. The process can succeed in practice because the deterministic parameters in Equation 59.12 change slowly with time so that it is possible to treat them as substantially constant over short averaging times. Stated another way, the

process of modeling the time differences by means of Equation 59.12 will succeed if and only if the time interval between the measurements is short enough to validate this assumption. In the next section, I discuss this requirement quantitatively.

59.6 NOISE MODELS

As I mentioned in the previous section, the limitations of the method come from our ignorance of the partition of the measured data into stochastic and deterministic components. To get insight into the characterization of the noise, it is useful to model an oscillator as a passive resonant system (such as an atomic transition) that is interrogated by a separate oscillator. The frequency of the oscillator is locked to the peak in the resonance response, and the output of the oscillator is used to generate the “ticks” that are used to drive the time display. Most oscillators that are stabilized using a transition in cesium or rubidium are configured this way. Lasers and most crystal oscillators cannot be modeled so easily because the oscillation frequency depends on a complicated combination of the gain of an amplifier and the phase shift in a resonant feedback loop. Nevertheless, even these types of oscillators can be reasonably well characterized using the machinery that I will discuss in the following sections.

59.6.1 White Phase Noise

A common method for generating the ticks is to generate an output pulse each time the sine wave of the oscillator passes through zero with a positive slope. Real zero-crossing detectors have some uncertainty in the exact trigger point, and this uncertainty is often represented as an equivalent noise voltage at the input to the circuit. I will designate this noise voltage as V_n . This noise contribution is modeled as a zero-mean random process that is unrelated to the true input signal. If the output of the oscillator has amplitude A and an angular frequency ω , then the noise voltage, whose amplitude is much smaller than the amplitude of the signal, introduces a time jitter in the determination of the time of the zero-crossing whose amplitude is

$$\Delta t = \frac{V_n}{A\omega}. \quad (59.14)$$

This fluctuation in the times of the output pulses is not associated with any frequency fluctuations in the oscillator itself. The noise voltage is inherent to the discriminator and has nothing to do with the input signal so that the magnitude of the time fluctuation for any measurement is unrelated to the impact for any other measurement. As we showed in the previous section, the Allan deviation for this type of noise varies as the reciprocal of the interval between measurements. Since the fluctuation does not arise in the oscillator itself, correcting the apparent frequency jitter caused by this noise by

steering the parameters of the oscillator (e.g., its frequency) is not the optimum strategy. I will discuss this point in more detail later. For now we note that the time jitter defined by Equation 59.14 has a mean of zero. From the physical perspective, there is an underlying “true” time difference that can be estimated by averaging multiple measurements. The standard deviation of the estimate decreases as more measurements are performed, and the mean value at any time is an unbiased estimate of the true time difference.

59.6.2 White Frequency Noise

We next consider the control loop that locks the oscillator frequency to the peak of the resonance response. A common method of detecting the peak in the response is to lock the oscillator to the zero of the first derivative of the resonance response function. The first derivative signal is generated by applying a small modulation to the oscillator frequency and synchronously detecting the amplitude of the response of the resonant system at the frequency of the applied frequency dither. An alternate method locks the oscillator to a frequency that is between two frequencies above and below the resonance where the upper and lower frequencies are measured as the values where the response has fallen to some specified fraction of the peak. (This is often implemented by means of a zero-mean, bipolar, square wave dither of the oscillator frequency.) The input to the discriminator is the difference between the response at the higher frequency and the response at the lower one. Both methods introduce a deterministic dither in the frequency output of the oscillator, which must be removed before the signal is used.

In either method, the discriminator locks the oscillator to a point where some voltage goes through zero. When the frequency is locked, the magnitude of the voltage specifies how much the frequency differs from the desired lock point, and the sign of the voltage specifies whether the frequency output is higher or lower than the desired operating point.

The response of this discriminator is limited by the same noise problems that I discussed in the previous section, but now it is the frequency of the oscillator rather than the output of a pulse that is affected by the equivalent noise at the input of the discriminator. The same argument as in the previous section shows that the result is a random frequency modulation. As in the previous section, the noise contribution is assumed to be a zero-mean random process so that frequency jitter is about a “true” value. This noise is identified by the slope of -0.5 in the log–log plot of the Allan deviation.

These random frequency fluctuations are integrated to produce time dispersion. Since the frequency fluctuations are characterized by a random process with a mean of zero, the impact on the time differences is a random walk with a variable step size, and so white frequency fluctuations are often described as a random walk in time (or phase, which is the same thing). Unlike the white phase noise case discussed in the previous section, the impact of white frequency noise on the measured time differences depends on the averaging time through Equation 59.12.

Since the impact of white frequency noise on the measured time differences is a function of the averaging time whereas white phase noise is independent of it, at least in principle it is possible to distinguish between the two by an appropriate choice of the averaging time; white frequency noise can be neglected at very short averaging times, and white phase noise becomes negligible as the averaging time is increased.

59.6.3 Long-Period Effects: Frequency Aging

The effects I have discussed in the previous sections are modeled as being driven by stochastic noise processes that are assumed to have zero means. That is, there is an underlying “true” time difference in the white phase noise domain, and there is an underlying “true” frequency offset in the white frequency noise domain. These models are useful because the behavior of many oscillators is well characterized in these terms.

Although there are oscillators that also exhibit a stochastic frequency aging that can also be modeled as a random zero-mean process just as we did previously for time and for frequency, the frequency aging of many oscillators is often a combination of a random function that is only approximately characterized as a zero-mean process combined with an approximately constant aging value.

There are a number of processes that can produce frequency aging that varies only very slowly and is approximately constant over many measurement cycles. For example, the transition frequency that is used as the reference frequency in an atomic clock is generally sensitive to external electric and magnetic fields (the DC Stark and Zeeman effects, respectively), to collisions between the atoms, to frequency shifts that result from the interaction with the probing field (the AC Stark effect), and to many other effects. These perturbing influences may have long-period variations that are translated into long-term frequency aging of the oscillator. The mechanical properties of quartz crystals, which determine the resonant frequency of a quartz crystal oscillator, often have similar long-period changes. Stochastic frequency aging may be caused by more rapid variation in any of these parameters and in other effects whose quantitative driving term admittance is not known.

From Equation 59.12, we can see that a constant frequency aging would produce time dispersion proportional to the square of the time interval between the measurements. Based on the relationship between time dispersion and Allan deviation, the Allan deviation for this type of aging would have a slope of +1 on a log–log plot of Allan deviation with respect to averaging time. The plot of the log–log plot of the Allan deviation with respect to averaging time has a slope of +0.5 for stochastic frequency aging.

Since the time dispersion due to frequency aging varies as the square of the interval between measurements, it is often possible to partition the measured variation in the time differences into three domains—a very short domain where white phase noise dominates the variance, an intermediate domain where frequency noise is the main contributor, and a long-period domain where frequency aging dominates.

59.6.4 Flicker Noise

In addition to the white time, frequency, and aging processes that I discussed in the previous sections, there is another contribution to the variance of time-difference measurements that cannot be easily characterized using a simple model analogous to the ones presented in the previous sections. I will characterize this type of noise process by contrasting it to the white phase noise and white frequency noise processes that I discussed before.

If a time-difference measurement process can be characterized as being limited by pure white phase noise, then there exists an underlying “true” time difference between the two devices. The measurements scatter about the true time difference, but the distribution of the measurements (or at worst the mean of a group of them) can always be characterized by a simple Gaussian distribution with only two parameters: a mean and a standard deviation. We can improve our estimate of the mean time difference by averaging more and more observations, and this improvement can continue forever in principle. There is no optimum averaging time in this simple situation—the more data we are prepared to average, the better our estimate of the mean time difference will be.

The situation is fundamentally different for a measurement in which one of the contributing clocks is dominated by zero-mean white frequency noise. Now it is the frequency that can be characterized (at least approximately) by a single parameter—the standard deviation.

Suppose we measure the time differences between a perfect device and the device under test, where the device under test has the same nominal frequency as the perfect device, but its frequency stability is degraded by white frequency noise. If the time difference between the two devices at some epoch is $X(t)$, then, since the deterministic frequency difference is zero, we would estimate the time difference a short time in the future as

$$X(t + \tau) = X(t) + y(t)\tau, \quad (59.15)$$

where $y(t)$ is the instantaneous value of the white frequency noise of the device under test, and

$$\langle y(t) \rangle = 0, \quad (59.16)$$

$$\langle y^2(t) \rangle = \varepsilon^2. \quad (59.17)$$

Since $y(t)$ has a mean of zero by assumption, Equation 59.15 predicts that the time difference at the next instant will be distributed uniformly about the current value of X , and the mean value of $X(t + \tau)$ is clearly $X(t)$. In other words, for a clock whose performance is dominated by zero-mean white frequency noise, the optimum prediction

of the next measurement is exactly the current measurement with no averaging. Note that this does not mean that our prediction is that

$$X(t + \tau) = X(t) \quad (59.18)$$

but rather that

$$\langle X(t + \tau) - X(t) \rangle = 0, \quad (59.19)$$

which is a much weaker statement because it does not mean that our prediction will be correct but only that it will be unbiased on the average. This is, of course, the best that we can do under the circumstances. The frequencies in consecutive time intervals are uncorrelated with each other by definition, and no amount of past history will help us to predict what will happen next. The point is that this is the opposite extreme from the discussion earlier for white phase noise, where the optimum estimate of the time difference was obtained with infinite averaging of older data.

Clearly, there must be an intermediate case between white phase noise and white frequency noise, where some amount averaging would be the optimum strategy, and this domain is called the “flicker” domain. Physically speaking, the oscillator frequency has a finite memory in this domain. Although the frequency of the oscillator is still distributed uniformly about a mean value of zero, consecutive values of the frequency are not independent of each other, and time differences over sufficiently short times are correlated. Both the frequency and the time differences have a short-term “smoothness” that is not characteristic of a simple random variable, and this smoothness is often mistaken for a pseudodeterministic variation. Flicker phase noise is intermediate between white phase noise and white frequency noise, and we would therefore expect that the Allan deviation of the time-difference data would have a dependence on averaging time that is midway between white and random walk processes. In fact, the simple Allan variance that we have discussed cannot distinguish between white phase noise and flicker phase noise. The more complicated modified Allan variance (Ref. [3]) is needed for this purpose, and the slope of this variance for flicker phase noise is indeed midway between the slopes for white and random walk processes.

The same kind of discussion can be used to define a flicker frequency noise that is midway between white frequency noise and white aging (or random walk of frequency). The underlying physical effect is the same, except that now it is the frequency aging that has short-period correlations. We could think of a flicker process as resulting from a very large number of very small jumps—not much happens in the short term because the individual jumps are very small, but the integral of them eventually produces a significant effect. The memory of the process is then related to this integration time.

The slopes of the log–log plot of both the Allan deviation and the modified Allan deviation are zero for flicker frequency noise. In other words, the estimate of the frequency does not improve with longer averaging times. The Allan deviation for these averaging times is often called the “flicker floor” of the device for this reason.

Data that are dominated by flicker-type processes are difficult to analyze. They appear deterministic over short periods of time, and there is a temptation to try to treat them as white noise combined with a deterministic signal—a strategy that fails once the coherence time is reached. On the other hand, they are not quite noise either—the correlation between consecutive measurements provides useful information over relatively short time intervals, and short data sets can be well characterized using standard statistical measures. However, the variance at longer periods is much larger than the magnitude expected based on the short-period standard deviation.

The finite-length averages that we have discussed can be realized using a sliding window on the input data set. This is simple in principle but requires that previous input data be stored; an alternative way of realizing essentially the same transfer function is to use a recursive filter on the output values. For example, suppose that the optimum averaging time is T . Then, if Y_{k-1} is an estimate of some parameter at time t_{k-1} and if y_k is the new data point received at time t_k , then we would estimate the update to Y at time t_k by means of

$$Y_k = \frac{wY_{k-1} + y_k}{w + 1}, \quad (59.20)$$

where w is a dimensionless parameter given by

$$w = \frac{t_k - t_{k-1}}{T}. \quad (59.21)$$

The time interval between measurements is chosen so that $w \leq 1$ —there is at least one measurement in the optimum averaging time. This method is often used in time scales, since these algorithms are commonly implemented recursively. Both the recursive and nonrecursive methods could be used to realize averages with more complicated transfer functions. These methods are often used in the analysis of data in the time domain [9].

59.7 MEASURING TOOLS AND METHODS

The oscillators that we have discussed before generally produce a sine wave output at some convenient frequency such as 5 MHz. (This frequency may also be divided down internally to produce output pulses at a rate of 1 Hz. Crystals designed for wristwatches and some computer clocks often operate at 32,768 Hz—an exact power of 2, which simplifies the design of these 1 Hz dividers.) A simple quartz crystal oscillator might operate directly at the desired output frequency; atomic standards relate these output

signals to the frequency appropriate to the reference transition by standard techniques of frequency multiplication and division. The measurement system thus operates at a single frequency independent of the type of oscillator that is being evaluated. The choice of this frequency involves the usual trade-off between resolution, which tends to increase as the frequency is made higher, and the problems caused by delays and offsets within the measurement hardware, which tend to be less serious as the frequency is made lower.

Measuring instruments generally have some kind of discriminator at the front end—a circuit that defines an event as occurring when the input signal crosses a specified reference voltage in a specified direction. Examples are 1 V with a positive slope in the case of a pulse or zero volts with a positive slope in the case of a sine wave signal. The trigger point is chosen at (or near) a point of maximum slope so as to minimize the variation in the trigger point due to the finite rise time of the waveform.

The simplest method of measuring the time difference between two clocks is to open a gate when an event is triggered by the first device and to close it on a subsequent event from the second one [10]. The gate could be closed on the very next event in the simplest case, or the N th following one could be used, which would measure the average time interval over N events. The gate connects a known high-frequency oscillator to a counter, and the time interval between the two events is thus measured in units of the period of this oscillator. The resolution of this method depends on the frequency of this oscillator and the speed of the counter, while the accuracy depends on a number of parameters including the latency in the gate hardware and any variations in the rise time of the input waveforms. The resolution can be improved by adding an analog interpolator to the digital counter, and a number of commercial devices use this method to achieve subnanosecond resolution without the need for a reference oscillator whose frequency would have to be at least 1 GHz to realize this resolution without interpolation.

In addition to the limit on the resolution of time-difference measurements that results from the maximum rate of the oscillator that drives the time-interval counter, time-difference measurements using fast pulses have additional problems. Reflections from imperfectly terminated cables may distort the edge of a sharp pulse, and long cables may have enough shunt capacitance to round the rise time by a significant amount. In addition to distorting the waveforms and affecting the trigger point of the discriminators, these reflections can alter the effective load impedance seen by the oscillator and pull it off frequency. Isolation and driver amplifiers are usually required to minimize the mutual interactions and complicated reflections that can occur when several devices must be connected to the same oscillator, and the delays through these amplifiers must be measured. These problems can be addressed by careful design, but it is quite difficult to construct a direct time-difference measurement system whose measurement noise does not degrade the time stability of a top-quality oscillator, and other methods have been developed for this reason. Averaging a number of closely spaced time-difference measurements is usually not of much help because these effects

tend to be slowly varying systematic offsets, which change only slowly in time, and so have a mean that is not zero over short times.

Many measurement techniques are based on some form of heterodyne system. The sine wave output of the oscillator under test can be mixed with a second reference oscillator that has the same nominal frequency, and the much lower difference frequency can then be analyzed in a number of different ways. If the reference oscillator is loosely locked to the device under test, for example, then the variations in the phase of the beat frequency can be used to study the fast fluctuations in the frequency of the device under test. The error signal in the lock loop provides information on the longer-period fluctuations. The distinction between “fast” and “slow” would be set by the time constant of the lock loop. As usual, we assume that any fluctuations in the reference oscillator are small enough to be ignored.

This technique can be used to compare two oscillators by mixing a third reference oscillator with each of them and then analyzing the two difference frequencies using the time-interval counter discussed earlier. In the “dual-mixer” version of this idea developed at NIST [11], this third frequency is not an independent oscillator but is derived from one of the input signals using a frequency synthesizer. The difference frequency has a nominal value of 10 Hz in this case. The time interval counter runs with an input frequency of 10 MHz and can therefore resolve a time interval of 10^{-6} of a cycle. Since the heterodyne process preserves the phase difference between the two signals, a phase measurement of 10^{-6} of a cycle is equivalent to a time-interval measurement with a resolution of 0.2 ps at the 5 MHz input frequency. It would be very difficult to realize this resolution with a system that measured the time difference directly at the 5 MHz input frequency or by measurements of the 1 Hz pulses derived from this reference frequency by a process of digital division.

All of these heterodyne methods share a common advantage: the effects of the inevitable time delays in the measurement system are made less significant by performing the measurement at a lower frequency where they make a much smaller fractional contribution to the periods of the signals under test. Furthermore, the resolution of the final time-interval counter is increased by the ratio of the input frequencies to the output difference frequencies (5 MHz to 10 Hz in the NIST system). This method does not obviate the need for careful design of the front-end electronics—the increased resolution of the back-end measurement system places a heavier burden on the high-frequency portions of the circuits and the transmission systems. As an example, the stability of the NIST system is only a few ps—about a factor of 10 or 20 poorer than its resolution. Some of the factors that degrade the stability of a channel in a dual-mixer system are common to all of the channels in a single chassis so that the differential stability of a pair of channels (which is what drives an estimate of the Allan variance) can be better than the stability of each channel alone.

Heterodyne methods are well suited to evaluating the frequency stability of an oscillator, but they often have problems in measuring absolute time differences because they usually have an integer ambiguity offset—an unknown integer number

of cycles of the input frequencies between the cycles that trigger the measurement system and the cycles that produce the 1 Hz output pulses that are the output “on-time” signals. For example, the time difference between two clocks measured using the NIST dual-mixer system is offset with respect to measurements made using a system based on the 1 Hz pulse hardware by an arbitrary number of periods of the 5 MHz input frequency (i.e., some multiple of 200 ns). To further complicate the problem, this offset generally changes if the power is interrupted or if the system stops for any other reason.

The offset between the two measurement systems must be measured initially, but it is not too difficult to recover it after a power failure, since the time step must be an exact multiple of 200 ns. Using the last known time difference and frequency offset, the current time can be predicted using a simple linear extrapolation. This prediction is then compared with the current measurement, and the integer number of cycles is set (in the software of the measurement system) so that the prediction and the measurement agree. This constant is then used to correct all subsequent measurements. The lack of closure in this method is proportional to the frequency dispersion of the clock multiplied by the time interval since the last measurement cycle, and the procedure will unambiguously determine the proper integer if this time dispersion is significantly less than 200 ns. This criterion is easily satisfied for a rubidium standard if the time interval is less than a few hours; the corresponding time interval for cesium devices is generally at least a day.

59.8 MEASUREMENT STRATEGIES

In the previous section, I discussed a number of methods that can be used to measure the time differences between two devices. All measurement techniques have some residual noise that appears as jitter in the time-difference measurements. I will assume for now that this jitter can be characterized as white phase noise. That is, it is a pure random process, and the impact on any measurement can be fully characterized by a distribution with a mean of zero and a standard deviation that is a property of the measurement system and does not depend on temperature, aging, or any of the other limitations that affect most real-world systems. Specifically, the noise of the measurement process can be characterized by relationships similar to Equation 59.9, previously.

With this assumption, the optimum strategy for estimating the time difference between a device under test and a second device that we think of as perfect (or at least very much better than the device under test) is to make repeated measurement of the time difference and to average the results. This number of measurements that can be combined into a single average will be limited by the assumption that the measurements differ only by the white phase noise of the measurement process, perhaps combined with the white phase noise of the oscillator itself as described earlier.

I will model the time differences between the device under test and our perfect reference device in terms of an initial time difference, a frequency offset, y , and a frequency aging, d :

$$x = x_0 + yt + \frac{1}{2}dt^2, \quad (59.22)$$

and my initial goal is to determine the time difference, x_0 , in the presence of the white phase noise of the measurement process, which has a standard deviation ε_m .

Equation 59.22 is not very useful as a tool to estimate the time difference directly. For example, it cannot be used to estimate the parameters x_0 , y , and d by applying standard least squares to an ensemble of measurements, because the parameters y and d are not constants but change slowly with time and have both stochastic and deterministic contributions. The least squares analysis will provide numerical estimates in a formal sense, but the estimates are neither physically significant nor statistically stationary, since Equation 59.22 is trying to fit a single quadratic relationship to an ensemble of data in which the parameters of the quadratic vary with the value of the independent variable. A simple least squares analysis does not have the flexibility to provide a robust estimate of parameters that are themselves statistical variables. The only exception might be for a very short data set where the principal contribution to the time dispersion is the white phase noise of the measurement process. The frequency offset and frequency aging can be taken to be approximately constant in this situation. However, we will use the iterative form of this relationship, Equation 59.12, in the more general case.

Since the noise of the measurement process is a random function that depends only on the characteristics of the measurement device and not on time or on any external perturbations, I can average the measured time differences to attenuate the effect of the measurement noise. I can continue to do this as long as the assumption that the time differences are randomly distributed about a “true” value as a result of the measurement noise alone. The duration of my average is time T , where T is given by

$$yT + \frac{1}{2}dT^2 \ll \varepsilon_m. \quad (59.23)$$

That is, the average can continue as long as the deterministic evolution of the time differences is much less than the noise of the measurement process so that the measurements are extracted from an ensemble of measurements that have the same time difference within the measurement uncertainty. An averaging time that satisfies this constraint will usually satisfy the weaker constraint that is driven by the noise in the frequency of the oscillator:

$$\sigma_y(T)T \ll \varepsilon_m, \quad (59.24)$$

where $\sigma_y(T)$ is the two-sample Allan deviation of the device under test for an averaging time of T . Equation 59.24 is a weaker condition because the frequency stability of most oscillators is generally better than the frequency accuracy.

These principles can be illustrated with a numerical example. Suppose that we wish to characterize a rubidium oscillator with a time-difference system that has a measurement noise of 1 ns (10^{-9} s). Based on the generic type of the device, we might estimate that the oscillator has a deterministic fractional frequency offset of 5×10^{-11} . (Recall that fractional frequency offsets are dimensionless.) The deterministic frequency aging is about 4×10^{-18} /s (about 1×10^{-11} /month). If we consider the limit imposed by Equation 59.23, the first term requires that $T \ll 20$ s. The second term is negligible for that value of T so that the requirement of Equation 59.23 is primarily driven by a constant frequency offset with no deterministic aging. This is a common result, which we will discuss in greater detail as follows. The two-sample Allan deviation of a rubidium standard for an averaging time of about 20 s is of order 10^{-12} so that Equation 59.24 is also easily satisfied with this averaging time.

Thus, if we know only generic values of the parameters that characterize the device, we can average the time differences for something less than 20 s. If we decide to use 16 measurements of the 1 Hz pulses from the device, we could average the 16 measurements, and the standard deviation of the measurements would be improved by a factor of 4. This is about the best we can do without knowing more about the device. Note that we cannot make a robust estimate of the frequency yet because our measurements are dominated by the white phase noise of the measurement process.

If we average 16 measurements, the uncertainty in the time difference has been reduced to about 0.25 ns or 250 ps. If we continue to make time-difference measurements, we can no longer average them directly, since the distribution becomes increasingly driven by the deterministic frequency offset of the device. We enter the domain where the deterministic frequency offset is making a contribution to the time differences, and the measurements are now limited by a combination of the white frequency noise of the device and white phase noise of the measurement. Thus, the measurement strategy changes from simply averaging the measured time differences to estimating the average frequency offset (the first derivative of the measurements) as well. This intermediate case is difficult to handle with this simple method, since we do not know how to partition the variance of the time-difference data into a contribution of the phase noise of the measurement process and the frequency noise of the clock itself.

One way to handle this ambiguity is to reverse the inequality in Equation 59.24. For example, if we made a time-difference measurement every 10^4 s, the white phase noise of each of the measurements would still be only 1 ns, but the contribution of the frequency noise of the oscillator would now contribute about 10 ns. We have now moved into the complementary domain, where the variance of the time differences is dominated by the frequency noise of the oscillator and the contribution of the noise of the measurement process is small enough to be ignored. The contribution of the deterministic frequency aging to the time differences is of order 0.5×10^{-18} /s $\times 10^8$ s² = 0.05 ns so

that we can make the reasonable assumption that *all* of the observed variance can be modeled as white frequency noise of the oscillator. (The ensemble algorithm that is used at NIST and other national laboratories often operates in this measurement domain where the data can be modeled as pure white frequency noise to a good approximation.) The optimum strategy in this domain is to average the offset frequency estimates obtained from the first differences of the time-difference data divided by the interval between these values.

This approach can continue as long as the measurements are in the white frequency noise domain, and the strategy will give an increasingly accurate estimate of the deterministic offset frequency of the device. We can derive the lower bound of this domain from the white phase noise of the measurement process as we have done before, but we don't have enough information to specify the upper end of this domain uniquely. We can make a rough estimate by comparing the time dispersion resulting from the frequency fluctuations to the time dispersion driven by the deterministic frequency aging and defining the upper limit of the averaging time as the interval when these two contributions are equal. This upper limit to the averaging time is defined by

$$\sigma_y(T)T = \frac{1}{2}dT^2 \quad (59.25)$$

or

$$T = \frac{2\sigma_y(T)}{d}. \quad (59.26)$$

Equation 59.26 highlights a fundamental problem with the model that we are using. The two-sample Allan deviation of a typical rubidium standard is generally not less than 10^{-12} for intermediate averaging times and tends to increase at longer averaging times because nonwhite frequency fluctuations usually become important there. On the other hand, the deterministic frequency aging, d , is of order $10^{-18}/\text{s}$ so that the averaging time predicted by Equation 59.26 (where the contribution of deterministic frequency aging to the frequency fluctuations is equal to the stochastic contribution) is generally longer than 10^6 s. In other words, it is difficult to estimate the deterministic frequency aging of a device because the aging is masked by the stochastic frequency fluctuations for moderate averaging times. Furthermore, nonstatistical considerations are often important at longer averaging times, which makes the numerator of Equation 59.26 larger and the problem of estimating the deterministic aging more difficult. On the other hand, the frequency aging contributes to the time differences as the square of the time interval so that it will almost always dominate the time dispersion at sufficiently long times.

The conclusions of the preceding discussion are more general than the specific case that I discussed, and the result is that it is very difficult to characterize the deterministic component of the long-term performance of any oscillator. (A time scale, which is an

ensemble of oscillators, is not immune to this problem.) The only difference among the different types of oscillators is where the “long-term” time domain begins. For example, the deterministic frequency fluctuation of a cesium standard is masked by the stochastic frequency fluctuations for almost any averaging time out to the life of the device (years), and cesium standards are generally modeled with no deterministic frequency aging for this reason. A hydrogen maser, on the other hand, has a deterministic frequency aging of order $10^{-16}/\text{day}$ ($\sim 10^{-21}/\text{s}$), but its stochastic frequency fluctuations are small enough so that the frequency aging must be included in any model of the time differences. The frequency aging of a rubidium standard is often of order $10^{-11}/\text{month}$ ($\sim 3.9 \times 10^{-18}/\text{s}$) and can be ignored only for short averaging times.

The basis of this discussion is that it is possible to find measurement domains where only one type of noise process dominates the variance of the measured time differences. This idea is widely used in modeling oscillators that are members of a time scale ensemble. For example, the AT1 algorithm used to estimate the average time of an ensemble of clocks at the NIST laboratory in Boulder, Colorado, is designed based on this principle [12]. The measurement system used at NIST has a measurement noise on the order of 1 ps and a time interval between measurements of 720 s, and the ensemble algorithm is based on the premise that the variance of the time differences can be modeled as white frequency noise. The hydrogen masers that are members of the ensemble have a constant frequency aging that is determined outside of the ensemble algorithm. This parameter is treated as a constant by the algorithm because it is difficult to compute a statistically robust estimate of the aging because of the problems discussed previously.

The algorithm used by the International Bureau of Weights and Measures (the BIPM in French) to compute International Atomic Time (TAI) and UTC is also based on these principles. The measurement noise of the time differences of the clocks located at the various national timing laboratories is much larger than the local time-difference measurements at NIST so that the interval between measurements has to be increased proportionally to guarantee that the contribution of the measurement noise to the time differences is smaller than the contribution of the frequency variations of the clocks. However, this increase in the time interval between measurements increases the impact of the frequency aging of the masers that contribute to TAI and UTC, and the algorithm used by the BIPM has been modified to recognize this effect [13]. (see also [14].) As we would expect from the previous discussion, including an explicit frequency aging term improves the long-term stability of the time scale by bringing the model closer to the actual behavior of the clocks that are members of the ensemble.

59.9 THE KALMAN ESTIMATOR

The preceding discussion depended on the assumption that the time interval between measurements was a free parameter that could be adjusted at will based on statistical considerations. In each case, it was chosen so that the variance of the time-difference

measurements could be modeled as arising primarily from a single source. However, there can be systems where the time interval between measurements is constrained by other factors to values that do not support this simplifying assumption. The variance of the time-difference measurements must be apportioned to more than one source in these configurations, but there is no way to do this using the machinery that we have developed so far. The Kalman estimator is one way of partitioning the variance in these situations, and I will discuss the general method in this section.

The Kalman estimator starts from the same recursive relationship for the time differences that we presented in Equation 59.12, but it adds two additional recursive relationships describing the evolution of the offset frequency and the frequency aging. The “Kalman state” of the clock is characterized by the values of these three parameters at any instant. The three equations that characterize the evolution of the state as a function of time are

$$\begin{aligned}x_k &= x_{k-1} + y_{k-1}\Delta t + \frac{1}{2}d_{k-1}(\Delta t)^2 + \xi \\y_k &= y_{k-1} + d_{k-1}(\Delta t) + \eta \\d_k &= d_{k-1} + \zeta.\end{aligned}\tag{59.27}$$

The time and offset frequency components are defined recursively with a deterministic contribution and a stochastic contribution (ξ and η , respectively). The frequency aging might have an initial constant value but usually is assumed to start at zero with only a stochastic variation, ζ . In each case, the stochastic contribution to the corresponding parameter is assumed to be a noise process that has a mean of zero and a variance that is initially known from other considerations, at least to first order. The noise contributions are assumed to be uncorrelated both in time and with each other. In other words, all three noise parameters satisfy relationships of the form

$$\begin{aligned}\langle \xi(t) \rangle &= 0 \\ \langle \xi(t)\xi(t') \rangle &= \xi^2 \delta(t-t') \\ \langle \xi(t)\eta(t) \rangle &= 0\end{aligned}\tag{59.28}$$

for all combinations of the parameters and for all t and t' .

The deterministic terms in Equation 59.27 describe how the state of the system evolves between measurements. The Kalman formalism can support measurements of any of the components of the state, but the most common arrangement is a measurement of the time difference of the clock with respect to a “perfect” reference as we discussed previously. In general, the measurement of the state component (the time difference in our discussion) will not agree with the value predicted from the previous state values by Equation 59.27, and the Kalman formalism provides a method of assigning the causes of this residual partially to updating the values of the deterministic parameters

and partially to the noise parameters. The details of how to do this are in the literature [15]. Practical realizations of the Kalman algorithm applied to estimating the parameters of clocks often have the same difficulty in estimating the frequency aging term that we discussed earlier for basically the same reason—it is difficult to calculate a robust estimate of the frequency aging in the presence of stochastic frequency noise that is generally larger even at moderately large averaging times.

In addition, a Kalman algorithm can be no better than the accuracy of the model Equations 59.27 and 59.28, which are used to define the evolution of the clock state. The assumption that the stochastic inputs to the frequency and frequency aging are random, uncorrelated, zero-mean processes is often not an accurate description of the true state of affairs. Even when the model equations accurately describe the steady-state evolution of the state parameters, Kalman algorithms often have start-up transients that can be troublesome in some real-time applications.

59.10 TRANSMITTING TIME AND FREQUENCY INFORMATION

I now consider the problem of comparing the time difference between two clocks that are not at the same location so that the time difference must be measured by means of a channel that connects the two devices. There is no difference between this configuration and the one we described previously in principle, but there are a number of practical differences that make this type of measurement significantly more complicated.

The first issue that I will consider is the transmission delay that is introduced by the channel. The delay is at least $3\text{ }\mu\text{s/km}$ so that it must be measured in all but the simplest measurement programs of time differences. (If the goal of the measurement program is a comparison of the frequency difference between the two devices, then the magnitude of the channel delay is not important, provided only that it remains constant to the level required by the measurement process. Although this sounds like an easier requirement to satisfy, designing a channel that satisfies this requirement and verifying that it does so is often not significantly easier than going the whole way and measuring the delay itself.)

There are many applications where the delay is small enough to be ignored. For example, there are a very large number of wall clocks, wristwatches, and some process-control devices that are calibrated and set on time by means of the radio signals transmitted by the NIST radio station WWVB in Fort Collins, Colorado. The transmission delay, which can be on the order of milliseconds, is simply ignored in these devices, since the required accuracy is generally only on the order of 1 s. Simple devices that are synchronized using the signals from the Global Positioning System (GPS) often work the same way—the unmodeled portion of the transmission delay (e.g., due to the refractivity of the ionosphere and the troposphere) is of order 65 ns in this case, but it is still much smaller than the required accuracy, which may be only to the nearest millisecond or even only to the nearest second. I will not consider these

applications here, and I will focus on the applications where some measurement of the channel delay is needed to satisfy the demands of the application.

I will describe three methods that are currently used to estimate the channel delay: (i) modeling the delay by means of ancillary parameters and measurements, (ii) the common-view method and its “melting-pot” variant, and (iii) two-way methods. I will also discuss the “two-color” method that is often used as an adjunct to one of the other methods when at least some of the path is through a medium that is dispersive. That is, the speed of the signal is a function of the frequency used to transmit the message. I will first discuss the general characteristics and assumptions of each of the methods, and I will then illustrate them with a more detailed discussion where the method is applied to a specific system.

The assumption that is implicit in this discussion is that the channel delay is not an absolute, unvarying parameter that can be measured once at the beginning by a process that may be complicated but needs to be done only once. There are some simple channels that satisfy this requirement—a coaxial cable between two parts of a building, for example.

A measurement of the delay of a long coaxial cable will often have some engineering complexity because coaxial cables are dispersive and have a frequency-dependent attenuation. In general, the attenuation, which is a result of the series inductance and shunt capacitance of the cable, increases with increasing frequency. Therefore, the rise time of a pulse transmitted on such a cable, which is a function of the high-frequency components of the signal, is increased as the signal propagates through the cable. The delay measurement is often further complicated by small impedance mismatches at the end of the cable, which cause reflections that interfere with the primary pulse used to measure the delay. These reflections can be exploited in the measurement process by leaving the far end of the cable unterminated and measuring the round-trip travel time of a pulse sent from the near end and reflected back from the open remote end. (The measurement of the travel time is not sensitive to the details of the signal that is transmitted along the cable, and a measurement that transmits a pseudorandom code instead of a pulse can have some technical advantages because it can be less affected by the attenuation of the high-frequency components of the test signal.) Whichever method is used, the delay is normally considered to be a constant that is a characteristic of the cable so that the measurement is normally a one-time effort. I will not consider this situation in detail, and I will focus on measurements where the delay cannot be measured once as a calibration constant.

59.10.1 Modeling the Delay

This method is based on the assumption that the channel delay can be estimated by means of some parameters that are known or measured from some ancillary measurement. For example, the geometrical path delay between a satellite and a receiver on the ground is estimated as a function of the position of the receiver, which

has been determined by some means outside of the scope of the timing measurement, and the position of the satellite, which is transmitted by the satellite in real time.

There are only a few situations where the channel delay estimated from a model is sufficiently accurate to satisfy the requirements of an application. In most cases, the delay estimate has significant uncertainties, even if the model of the delay is well known. For example, the delay through the troposphere is a known function of the pressure, the temperature, and the partial pressure of water vapor, but these parameters are likely to vary along the path so that end-point estimates may not be good enough. This limitation is bad enough for a nearly vertical path from a ground station to a satellite, but the uncertainties become much larger for a lower-elevation path between two ground stations or from signals from a satellite that is near the horizon.

59.10.2 The Common-View Method

This method depends on the fact that there are two (or more) receivers that are equally distant from a transmitter. Since the two path lengths are the same, any signal sent from the transmitter arrives at the same time at both receivers. Each receiver measures the time difference between its local clock and the received signal, and these two measurements are subtracted. The result is the time difference between the clocks at the two receivers. In this simple arrangement, the accuracy of the time difference does not depend on the characteristics of the transmitted signal or the path delay.

In the real world, it is difficult to configure the two receivers so that they are exactly equally distant from the transmitter, and some means must be used to estimate the portion that is not common to the two paths. This estimate is not as demanding as estimating the full path delay so that the common-view method attenuates any errors in the estimate of the path delay.

There are also a number of subtle effects that we must consider when the two path delays of a common-view measurement are not exactly equal. If a single signal is used to measure the time difference, then the signal does not arrive at the two receivers at the same time, since the path delays are somewhat different. Therefore, any fluctuation in the characteristics of the receiver clocks during this time difference must be evaluated. On the other hand, if the measurement is made at the same instant as measured by the receiver clocks, then the signals that are measured by the two receivers did not originate from the satellite at the same instant of time. Therefore, the fluctuations in the characteristics of the satellite clock during this time interval must be evaluated.

Finally, a common-view measurement algorithm does not support casual associations among the receivers. The stations that participate in the measurement process must agree on the source to be observed and the time of the observation. There must also be a channel between the two receivers to transmit the measured time differences. On the other hand, there needs to be no relationship between the receivers and the transmitter, and the transmitter need not even know that it is being used as part of a common-view measurement process. Signals from commercial analog television

stations have been used as common-view transmitters, and the zero-crossings of the mains voltage can also be used to compare clocks with an uncertainty on the order of a fraction of a millisecond.

59.10.3 The “Melting-Pot” Version of Common View

The previous discussion of common view focused on a number of cooperating receivers, where each one measured the time difference between a physical signal and the local clock. However, there can be some situations where there is no single transmitter that can be observed by the receivers at the same epoch. For example, if the common-view method is implemented by means of signals from a GPS navigation satellite, then receivers on the surface of the Earth that are sufficiently far apart cannot receive signals from any one satellite at the same time.

However, determining the position of a receiver by means of signals from the GPS satellites depends on the fact that the clock in each satellite has a known offset in time and in frequency from a system-average time that is computed on the ground and transmitted up to the satellites. Each satellite broadcasts an estimate of the offset between its internal clock and this GPS system time scale. By means of this information, two stations that observe two different satellites can nevertheless compute the time difference between the local clock and GPS system time, rather than computing a time difference between the physical signal transmitted by the satellite and the time of the local clock. The common-view time difference in this case is not with respect to a physical transmitter but rather with respect to the computed paper time scale, GPS system time.

In general, a receiver may be able to compute the time difference between its clock and GPS system time by means of the signals from several satellites, which explains the origin of the term “melting-pot” method. All of these measurements should yield the same time difference in principle, but this is not the case in practice for a number of reasons.

In the first place, the path delays between the receivers and the various satellites are not even approximately equal so that any error in computing the path delays is not attenuated as it is in the common-view method described in the preceding text. In addition, the method depends on the accuracy of the offset between each of the satellite clocks and the system time. As a practical matter, the full advantage of the melting-pot method is realized only when the orbits of the satellites and the characteristics of the onboard clocks have been determined using postprocessing—the values broadcast by the satellites in real time are usually not sufficiently accurate to be useful for this method.

On the other hand, the melting-pot method can usually use the observations from several satellites at the same time so that the random phase noise of the measurement process can be attenuated by averaging the data from the multiple satellites. Therefore, a comparison between the simple two-way method and the melting-pot version

depends on a comparison between the noise of the measurement process, which would favor a melting-pot measurement using multiple satellites, and the uncertainties and residual errors in the orbital parameters of the satellites and the offset between the clock in each satellite and GPS system time. The accuracy of the melting-pot method improves as more accurate solutions for the orbits and satellite clocks become available [16].

59.10.4 Two-Way Methods

There are a number of different implementations of the two-way method, but all of them estimate the one-way delay between a transmitter and a receiver as one-half of the round-trip delay, which is measured as part of the message exchange. The accuracy of the two-way method depends on the symmetry of the delays between the two end points. The accuracy does not depend on the magnitude of the delay itself; although the magnitude of the delay can be calculated from the data in the message exchange, the accuracy of the time difference does not require this computation.

There are generally two aspects of the delay asymmetry that must be considered. The first is a static asymmetry—a difference in the delays between the end points in the opposite directions. In general, this type of asymmetry cannot be detected from the data exchange, and it places a limit on the accuracy that can be realized with any two-way implementation. The second type of asymmetry is a fluctuation in the symmetry that has a mean of zero. In other words, the channel delay is symmetric on the average, but this does not guarantee the symmetry of any single exchange of data. The impact of this type of fluctuating asymmetry can be estimated with enough data. As I will show in more detail later, a smaller measured round-trip delay is generally associated with a smaller time offset due to any possible asymmetry.

The transmitter and receiver at the end points of the path are often sources of asymmetry. These hardware delays are often sensitive to the ambient temperature. However, the admittances to temperature fluctuations may be different at the two end stations, and the temperatures at the two end points may be quite different. Finally, there are often components of the measurement process that are outside of the two-way measurement loop, and any delays originating in these components must be measured on every message exchange or measured once and stabilized.

59.10.5 The Two-Color Method

Suppose that there is a portion of the path that has an index of refraction that is significantly different from the vacuum value of one. This difference of the index of refraction relative to its vacuum value is the *refractivity* of the path. Suppose also that the refractivity is dispersive. That is, it depends on the frequency that is used to transmit the message. If the length of the path is measured using the transit time of an electromagnetic signal, the refractivity will increase the transit time so that the effect

of the refractivity will be to make the path length appear too long. If the length of the true geometric path is D , then the measured length will be L , where L is given by

$$L = nD = D + (n - 1)D. \quad (59.29)$$

I will now consider the special case where the refractivity can be expressed as a product of two functions: $F(p)G(f)$. That is,

$$n - 1 = F(p)G(f). \quad (59.30)$$

The first function, F , includes parameters that characterize the transmission medium, including any dependence of these parameters on the environment such as the ambient temperature, relative humidity, etc. The second function, G , describes the dispersive characteristics of the path. Both of these functions can be arbitrarily complex and non-linear—the only requirement is that the separation be complete. The function F cannot depend on the frequency that is used to transmit the signal, and the function G cannot depend on the parameters that describe the characteristics of the path.

If I measure the apparent length of the path using two frequencies, f_1 and f_2 , I will obtain two different values for the apparent length because the index is dispersive. These two measured values are L_1 and L_2 , respectively. Since the geometrical path length is the same for the measurements at the two frequencies, the difference between the two measurements can be used to solve for the value of the function $F(p)$:

$$\begin{aligned} L_1 - L_2 &= F(p)(G(f_1) - G(f_2))D \\ F(p) &= \frac{L_1 - L_2}{D} \frac{1}{(G(f_1) - G(f_2))} \\ n_1 - 1 &= F(p)G(f_1) = \frac{L_1 - L_2}{D} \frac{G(f_1)}{G(f_1) - G(f_2)}. \end{aligned} \quad (59.31)$$

If I substitute the last relationship of Equation 59.31 into Equation 59.29 evaluated for frequency f_1 , I obtain

$$L_1 = D + (n_1 - 1)D = D + (L_1 - L_2) \frac{G(f_1)}{G(f_1) - G(f_2)}, \quad (59.32)$$

which allows me to find the geometrical path length, D , in terms of L_1 , the length measured using frequency f_1 , and the difference in the lengths measured at the two frequencies f_1 and f_2 multiplied by a known function of the two frequencies. If I call this function of the two frequencies H , then

$$\begin{aligned} H(f_1, f_2) &= \frac{G(f_1)}{G(f_1) - G(f_2)} \\ D &= (1 - H)L_1 + HL_2. \end{aligned} \quad (59.33)$$

The details of the function G are not important, provided only that it is known and that the medium is dispersive. (The denominator of the fraction on the right-hand side of Eq. 59.31 or 59.32 is zero for a nondispersive medium. The difference in the apparent lengths in Eq. 59.31 will also be zero in this case.) Note that the second term on the right-hand side of Equation 59.32, which is the correction to the geometrical length D due to the dispersive medium, does not depend on L , the extent of that medium, but only on the apparent difference in this length for the measurements at the two frequencies. Thus this relationship is equally valid if only a portion of a geometric path is dispersive, and the correction term specifies the apparent change in the length of only that portion of the path. Note also that I do not have to know anything about the function $F(p)$ —only that the separation into two terms expressed by Equation 59.30 represents the dispersion.

The measurements of both L_1 and L_2 will have some uncertainty in general so that the two-color determination of the geometrical length, D , will have an uncertainty that is greater than it would have been if the medium were nondispersive so that a measurement at one frequency would have been adequate. The magnitude of the degradation depends on the details of the function H , and I will discuss this point again when I describe measurements using navigation satellites such as those of the GPS.

59.11 EXAMPLES OF THE MEASUREMENT STRATEGIES

In the following sections I will describe systems that use the various measurement strategies that I have outlined earlier. I will begin by describing the characteristics of the satellites of the GPS, since the data transmitted by these satellites are widely used for timing applications in the one-way, common-view, and melting-pot modes. The Russian GLONASS system, the European Galileo system, and the Chinese BeiDou system are different in detail, but the following discussion describes the general features of all of them. In general, the differences between the systems are hidden from the general user and are a concern only of the receiver designer.

59.11.1 The Navigation Satellites of the GPS

The GPS system uses at least 24 satellites in nearly circular orbits whose radius is about 26,600 km. (The number of satellites in the constellation that are active at any time is generally >24 .) The orbital period of these satellites is very close to 12 h, and the entire constellation returns to the same point in the sky (relative to an observer on the Earth) every sidereal day (very nearly 23 h 56 m).

The satellite transmissions are derived from a single oscillator operating at a nominal frequency of 10.23 MHz as measured by an observer on the Earth. In traveling from the satellite to an Earth-based observer, the signal frequency from every satellite is modified by two effects that are common to all of them—a redshift due to the second-order

Doppler effect and a blueshift due to the difference in gravitational potential between the satellite and the observer. These two effects produce a net fractional blueshift of about 4.4×10^{-10} ($38 \mu\text{s/day}$), and the proper frequencies of the oscillators on all of the satellites are adjusted downward to compensate for this effect, which is a property of the orbit and is therefore common to all of them. In addition to these common offsets, there are two other effects—the first-order Doppler shift and a frequency offset due to the eccentricity of the orbit, which vary with time and from satellite to satellite. The receiver computes and applies the corrections for these effects.

The primary oscillator is multiplied by 154 to generate the $L1$ carrier at 1575.42 MHz and by 120 to generate the $L2$ carrier at 1227.6 MHz. (The newer GPS satellites will transmit signals on additional frequencies, and the Galileo, GLONASS, and BeiDou systems transmit signals at slightly different frequencies.) These two carriers are modulated by three signals: the precision “P” code, a pseudorandom code with a chipping rate of 10.23 MHz and a repetition period of 1 week; the “clear access” or coarse acquisition “C/A” code with a chipping rate of 1.023 MHz and a repetition rate of 1 ms; and a navigation message transmitted at 50 bits/s. The codes are derived from the same 10.23 MHz primary oscillator. Under normal operating conditions, the C/A is present only on the $L1$ carrier. Many timing receivers process only the C/A code. Although the P code is normally encrypted with an encryption key that is not available to unclassified users, many receivers can operate in a “semi-codeless” mode where the P code data can be decoded with some increase in the noise of the process.

Each GPS satellite transmits at the same nominal frequencies but uses a unique pair of C/A and P codes. The codes are constructed to have very small cross-correlation at any lag and a very small autocorrelation at any nonzero lag (code division multiple access (CDMA)). The receiver identifies the source of the signal and the time of its transmission by constructing local copies of the codes and by looking for peaks in the cross-correlation between the local codes and the received versions. Since there are only 1023 C/A code chips, it is feasible to find the peak in the cross-correlation between the local and received copies using an exhaustive brute-force method. When this procedure succeeds, it locks the local clock to the time broadcast by the satellite modulo 1 ms, the repetition rate of the whole C/A code. The procedure locks the local clock to the satellite time with a time offset due to the transmission delay (about 65 ms) and allows the receiver to begin searching for the 50 bits/s navigation message.

The navigation message contains an estimate of the time and frequency offsets of each satellite clock with respect to the composite GPS time, which is computed using a weighted average of the clocks in the satellites and in the tracking stations. This composite clock is in turn steered to UTC(USNO), which is in turn steered to UTC as computed by the BIPM. The time difference between GPS system time and UTC(USNO) is guaranteed to be less than 100 ns (modulo 1 s), and the estimate of this offset, which is transmitted as part of the navigation message, is guaranteed to be

accurate to 25 ns (also modulo 1 s). In practice, the performance of the system has almost always substantially exceeded its design requirements.

The UTC time scale includes leap seconds, which are added as needed to keep UTC with ± 0.9 s of UT1, a time scale based on the position of the Earth in space. The GPS time scale does not incorporate additional leap seconds beyond the 19 that were defined at its inception; the time differs from UTC by an integral number of additional leap seconds as a result. This integer-second difference, GPS time—UTC, is currently (December, 2015) 17 s and will increase as additional leap seconds are added to UTC. The number of leap seconds between GPS time and UTC is transmitted as part of the navigation message but is not used in the definition of GPS time itself. Advance notice of a future leap second in UTC is also transmitted in the navigation message.

Most modern receivers can observe several satellites simultaneously and can compute the time differences between the local clock and GPS system time using all of them at same time. In each case, the time of the local clock that maximizes the cross-correlation with the signal from each satellite is the *pseudorange*—the raw time difference between the local and satellite clocks. (It is related to the geometrical time of flight with an additional time offset since the clock in the receiver is generally not synchronized to GPS system time.)

Using the contents of the navigation message, the receiver corrects the pseudorange for the travel time from the satellite to the receiver, for the offset of the satellite clock from satellite system time, etc. (If the receiver can process both the $L1$ and $L2$ frequencies, then the receiver can also estimate the additional delay through the ionosphere due to its refractivity by applying the two-color method described before to the difference in the pseudoranges observed using the $L1$ and $L2$ frequencies. If the receiver can process only the $L1$ signal, then it usually corrects for the ionospheric delay using a parameter transmitted in the navigation message.) The result is an estimate of the time difference between the local clock and GPS system time.

In principle, the time difference between the local clock and GPS system time should not depend on the specific satellite whose data are used for the computation. In practice, the time differences computed using the data from different satellites will differ because of the noise in the measurement processes and because of errors in the broadcast ephemerides and the parameters of the satellite clocks. The group of time differences with respect to GPS system time, computed from the different satellites, forms a “redundant array of independent measurements” (RAIM), and some analysis methods compare these time differences in an attempt to detect a bad satellite. These “T-RAIM” algorithms succeed when the time difference computed using the data from one satellite differs from the mean of the differences computed from the other satellites by a statistically significant amount. The T-RAIM algorithm can be used in the one-way, common-view, or melting-pot algorithms that I discuss in the next sections. The same idea is used in the Network Time Protocol (NTP) to identify a bad server. I will discuss this point in greater detail as follows.

59.11.2 The One-Way Method of Time Transfer: Modeling the Delay

This method is most often used with time transfer by means of signals from navigation satellites because they are the only systems that transmit enough information to support an accurate delay estimate. The estimate of the transit time of a message from a navigation satellite to a receiver on the ground can be divided into a number of components that are increasingly difficult to estimate.

The largest single estimate to the propagation delay is the delay resulting from the geometric path length. The magnitude of this delay depends somewhat on the position of the satellite in the sky but is typically approximately 65 ms. The path length is computed from the position of the satellite, which is estimated from the orbital parameters transmitted as part of the navigation message, and the position of the ground receiver. In pure timing applications, the position of the ground receiver is assumed to be known from other data, and I will assume that this is the case in the current discussion. (If the position of the receiver is not known *a priori*, it can be estimated by computing the distances from the receiver to multiple satellites and solving for the four unknowns: the three Cartesian coordinates of the position of the receiver and the time offset of its clock with respect to satellite system time.) In a real-time application the accuracy of the estimate of the geometric path delay is limited by any uncertainty in the position of the receiver (the vertical coordinate usually has the largest uncertainty) and by errors in the broadcast ephemeris parameters. These combined uncertainties are generally on the order of a few meters, which is equivalent to an uncertainty in the time delay of about 10 ns or less. Thus the uncertainty in the correction is much smaller than the magnitude of the correction itself.

The additional delay due to the passage of the signal through the ionosphere adds approximately 65 ns to the geometric delay. A receiver that can process both of the frequencies transmitted by a satellite can estimate the effect of the ionosphere using the two-color method that I have described previously. Simpler, single-frequency receivers can use an estimate of the effect of the ionosphere that is broadcast by the satellite as part of the navigation message. This is a globally averaged prediction and is therefore less likely to be accurate at any specific location.

The additional delay due to the passage of the signal through the lower atmosphere (the troposphere) is much smaller than the additional delay through the ionosphere, but there is no easy way of estimating it because the refractivity does not depend on the carrier frequency so that the two-color method cannot be used. Some more sophisticated analyses estimate this delay by means of local measurements of atmospheric pressure, temperature, and water vapor content, but these data are not always available. Even when they are available at a site, these parameters often have significant azimuthal variation, which is generally not easily estimated. (Boulder, Colorado, is potentially particularly bad in this respect, since the mountains to the west and the plains to the east would be expected to have quite different temperature profiles.) There are also models of the refractivity of the troposphere, which estimate this parameter as a function of the day of the year and, possibly, the coordinates of the receiving station.

The magnitude of this delay is typically on the order of 6 ns at the zenith, and it increases for satellites at lower elevation by a factor that is roughly proportional to the increase in the slant path through the troposphere relative to the zenith path length. The increase in the slant path delay relative to the zenith delay is usually estimated as proportional to the reciprocal of the sine of the elevation angle.

If an analysis assumes the slant path model of the variation of the delay, it is possible to solve for the zenith delay by observing the apparent variation in the time difference estimates obtained from satellites at very different elevation angles. This estimate does not work as well as we might like because the slant path model is only an approximation, because the tropospheric refractivity often has a significant azimuthal variation, and because the variation from one satellite to another is also affected by measurement noise and by errors in the broadcast ephemerides or any error in the coordinates of the receiver.

Many time-difference measurements ignore the effect of the troposphere altogether. This introduces a systematic error of order 10 ns in the time-difference estimates; as I mentioned in the preceding text, the magnitude of this error depends on the elevation of the satellites that are being observed.

The final contributions to the model of the delay are effects that are local to the receiver: the delay through the hardware and the motion of the station due to the Earth tides and other geophysical effects. The delay through the receiver hardware is normally assumed to be a constant that varies only very slowly over periods of years. The delay is often dominated by the delay through the antenna and the cable from the antenna to the receiver, and a value on the order of 100 ns is typical. (Delays through coaxial cables are of order 5 ns/m.)

It is possible to calibrate the delay through a receiver using a special signal generator that mimics the signals from the real satellite constellation. This type of equipment is not widely available, and most timing laboratories perform a differential calibration in which the delay through the receiver under test is compared to the delay of a “standard” receiver. This method is obviously not adequate for a one-way measurement but is widely used because most timing laboratories use the satellites in common view, which I discuss in the following section.

The motion of the station and other geophysical effects contribute 1–2 ns to the overall delay. The magnitude of these effects can be calculated and included in more complicated postprocessed analyses but are generally ignored for real-time applications.

59.11.3 The Common-View Method

I have already described most of the important features of the common-view method. It is most often used with signals from the navigation satellites, but it is more general than this and can be used with other sources as well. Signals from LORAN transmitters and even from television stations have been used in this way. There have even been some experiments to use the zero-crossings of the power line in common view within a building or over a small area.

The method has two principal limitations:

1. It is difficult in practice to configure the measurements so that the receivers are all equidistant from the source. Therefore, some correction is almost always necessary to model the differential delay. The differential delay is much smaller than the delay itself in most configurations so that the required accuracy of the model of the delay is correspondingly easier to satisfy. However, ignoring the difference in the path delays is often not sufficiently accurate.
2. The common-view method (and its melting-pot variant) cannot provide any help in mitigating the effects of delays that are local effects at a site. The differential effects of the ionosphere can be significant and are usually estimated using the two-color method. The differential effects of the troposphere cannot be estimated in this way and are often ignored. Ignoring the differential effects of the troposphere is often justified because the total contribution is relatively small and the differential contribution is correspondingly smaller.

Multipath is a more serious local effect that is often too large to ignore. The effect is caused by copies of the signal that reach the antenna after they have been reflected from some nearby object. These signals always travel a longer distance than the primary one, and they arrive later than the primary signal as a result. A simple omnidirectional antenna typically responds to these signals, and the receiver computes a correlation that is a complicated sum of the direct and reflected signals.

The multipath effect is a complicated function of the position of the satellite with respect to the antenna and the local reflectors, and it is therefore periodic with the orbital period of the satellite. From this perspective, the orbital periods of the GPS satellites are all very close to one sidereal day (23 h 56 m) so that the multipath reflections have this periodicity. They can usually be estimated by comparing the time differences measured from any satellite at the same sidereal time on consecutive days.

The BIPM has exploited this sidereal-day periodicity in defining the common-view tracking schedules that are used by timing laboratories and National Metrology Institutes to compare time scales and to facilitate the computation of TAI and UTC. The observation time for each satellite is advanced by 4 min every day in the BIPM schedule so that every satellite returns to the same point in the sky on every track each day relative to the antenna and to any multipath reflectors. Thus, the multipath environment is a constant for each track, although it generally varies from track to track. This has the advantage of converting the varying effects of multipath to systematic offsets that are approximately constant for each track. The assumption of a sidereal-day periodicity is not exact so that the offset due to the multipath contribution changes slowly with time. These long-period effects can be hard to distinguish from the contributions due to the random walk of frequency and frequency aging that I discussed previously.

Locating the antenna far away from reflecting surfaces can help minimize the impact of multipath reflections; adding choke rings and a ground plane to an antenna, which attenuate signals arriving from the side or from below, can also help.

Another strategy to mitigate the impact of multipath is to exploit the sidereal-day periodicity and compute the average frequency of the local clock with respect to the GPS system time as an average over a sidereal day. The multipath contribution cancels in the sidereal-day time difference so that the frequency estimated in this way is almost insensitive to multipath effects.

59.11.4 Two-Way Time Protocols

In the following sections, I will describe three two-way time protocols that are commonly used to compare clocks at remote locations. The list is intended to be descriptive rather than exhaustive. For example, I do not discuss time transmission using optical fibers because this method is generally too expensive to be used for long distances and because the underlying physics is basically the same as the other methods that I do describe. I also do not discuss the Precise Time Protocol (PTP, often called IEEE 1588) in any detail for much the same reasons. Its capabilities are very similar to a hardware-assisted version of NTP, and it is generally not well suited to long-distance time comparisons because it assumes that the delay is nearly constant so that it does not have to be measured on every message exchange.

59.11.4.1 The NTP The NTP is widely used to transmit time and compare clocks that are linked together by a channel that is based on a packet-switched network such as the Internet. The NTP message format is based on the User Datagram Protocol [17] (UDP). The UDP message exchange is not sensitive to the details of the physical hardware that is used to transmit the packets. However, as with all two-way protocols, the accuracy of the NTP message exchange depends on the symmetry of the inbound and outbound delays, and this symmetry is often limited by the characteristics of the physical layer used to transmit the messages. In the following discussion I will focus on the time-difference accuracy of the message exchange; I will defer the question of how often a system should initiate such an exchange (the “polling interval”) to a later section, and I will discuss only briefly the question of how a client system should discipline its local clock based on the exchange of messages with a server.

The protocol is initiated when station “A” sends a request for time information to station “B.” The two stations might have a client–server relationship, in which the client intends to adjust its clock based on the results of the exchange, or it could be a peer-to-peer exchange in which two systems exchange timing information with the goal of setting the times of both systems to agree with each other.

The message is sent at time T_{1a} as measured by the clock on system A. The transmission delay from station A to station B is δ_{ab} so that when the message arrives at station

B, the time at station A is $T_{1a} + \delta_{ab}$. The time of arrival at station B, measured by the local clock at that station, is T_{2b} , and the time difference between stations A and B is

$$(\Delta T)_{ab} = (T_{1a} + \delta_{ab}) - T_{2b}. \quad (59.34)$$

The B system responds by sending a message back to A. The message leaves the B system at time T_{3b} and arrives back at the A system at time $T_{4a} + \delta_{ba}$, and the time at the A system at that instant is T_{4a} .

The total round-trip transit time is measured at station A as the time that has elapsed during the message exchange as measured by the clock on station A, less the time between when station B received the request and when it replied, as measured by the clock on station B:

$$\Theta = \delta_{ab} + \delta_{ba} = (T_{4a} - T_{1a}) - (T_{3b} - T_{2b}). \quad (59.35)$$

We now assert that the path delay is symmetric so that the inbound and outbound delays are equal. Then the path delay from A to B, δ_{ab} , is simply one-half of the expression on the right-hand side of Equation 59.35. If we substitute one-half of the right-hand side of Equation 59.35 into Equation 59.34, the time difference between stations A and B is

$$(\Delta T)_{ab}^s = \frac{T_{1a} + T_{4a}}{2} - \frac{T_{2b} + T_{3b}}{2}. \quad (59.36)$$

The superscript s indicates that the time difference is computed using a symmetric path delay. If the path delay is not symmetric, then the inbound and outbound delays are not equal. We can parameterize this asymmetry as

$$\delta_{ab} = (0.5 + \varepsilon)\Theta. \quad (59.37)$$

The asymmetry parameter ε can take values from +0.5 to -0.5. The positive limit indicates that the path delay from A to B dominates the round-trip delay and the delay in the other direction is negligibly small, while the negative limit specifies the inverse: the delay from A to B is negligible compared to the reverse delay from B to A.

If we substitute Equation 59.37 into Equation 59.34, then the first term on the right-hand side of Equation 59.34 reproduces the time-difference expression of Equation 59.36, and the second term adds a correction to the time difference:

$$(\Delta T)_{ab}^a = (\Delta T)_{ab}^s + \varepsilon\Theta. \quad (59.38)$$

Since we model the measurement based on the assumption of a symmetric delay (Eq. 59.36), the time difference that we estimate is in error. The magnitude of the error is given by $\varepsilon\Theta$, the second term in Equation 59.38. This term is proportional

both to the magnitude of the asymmetry and to the round-trip delay. Thus a smaller round-trip delay guarantees a smaller error due to any asymmetry. The lesson is that NTP servers should be widely located so that the round-trip delay to any user is minimized.

The round-trip delay is often of order 100 ms (0.1 s), and typical asymmetries are on the order of a few percent of the delay. Therefore, we might expect that a typical NTP message exchange would have an error on the order of 5 or 10 ms due to the asymmetry of the path delay, and errors of this order should be considered as routine for a server and a client on a wide-area network.

In addition to possible asymmetries in the network delay, there may also be additional asymmetries in the client system. For example, if the process that manages the NTP message exchange runs in a standard user environment, then it must compete for processor cycles with all of the other processes that may be active on the system. In addition, it must issue a request to the system to retrieve the system time each time a message is sent or received in order to have the values to compute the time differences described in the preceding text. All of these effects add to the network delay measurement; depending on the details of the system and the processes that are active, it may also add to the asymmetry.

In order to minimize these effects, the NTP process can be moved into the system space where it runs at much higher priority as a system service. The ultimate version of this idea would be to move the NTP process into the network driver that receives and transmits the network packets, and some version of NTP and its cousin PTP (also called IEEE 1588) operate in this mode.

Although moving the NTP process into the system space or into the network driver itself will make the NTP process appear more stable and more accurate, the overall timing accuracy of an application that uses the system time may be degraded. This application normally runs as a standard user process, and it therefore experiences the same jitter as a user-level NTP process would experience when it issues a request to the system for the current time. This jitter is inside of the measurement loop when the NTP process runs as a user process, and the time-difference calculation therefore takes it into account (at least to some extent even if it also contributes to the asymmetry). However, this delay is completely outside of the measurement loop if the NTP process is pushed down to the system or driver levels so that the application process experiences the full impact of the delay jitter in requesting system services. Thus, while the NTP statistics improve, the accuracy realized by a user process may be degraded, and this problem is not reflected in any of the NTP statistics.

59.11.4.2 The ACTS Protocol The NIST (and a number of other timing laboratories) operate a time service that transmits the time in a digital format by means of standard dial-up telephone lines. The NIST system is called the Automated Computer Time Service (ACTS) [18], and it corrects for the transmission delay using a variant of the two-way protocol I have described.

The ACTS servers transmit a text string each second with the time derived from the NIST clock ensemble and an on-time marker (OTM) character. This character is initially transmitted using a default advance. If the user echoes the OTM back to the server, the server measures the round-trip delay, estimates the one-way delay as one-half of this value, and adjusts the advance of the next OTM transmission so that it will arrive at the user's system on time. The server changes the OTM character from "*" to "#" to indicate that it has entered this delay-calibrated mode. In every case, the server includes the estimate of the one-way delay in the message so that the client can determine the advance that was used. In the context of the previous discussion of the NTP protocol, the ACTS protocol assumes that $(T_{3b} - T_{2b})$ is essentially zero. That is, the client echoes the OTM character back to the server with only negligible delay.

This process continues on every transmission. The OTM character is advanced based on the one-way delays estimated from the average of the round-trip measurements of the previous seconds. The details of the averaging process are determined dynamically by the server based on the measured variation of the round-trip delay from second to second.

The original ACTS system was designed to minimize the complexity of the code in the receiving system. The receiver needs only to echo the OTM back to the server, and the next OTM will be transmitted so that it arrives on time. The receiver did not have to perform any calculations at all. The assumption of this design was that the delay variations were not accompanied by variations in the symmetry, and this assumption was largely confirmed by the original design, which could transmit time messages with an accuracy of order 0.5 ms RMS.

Placing the delay calculation in the server simplifies the design of the client system, but it has the unfortunate side effect that the server cannot detect a change in the symmetry of the delay whether or not this change in symmetry is accompanied by a change in the total round-trip delay. However, the client system is in a better position to determine what is really going on.

Since the true time difference between the client and the server changes by less than 1 ms from second to second, any significant change in the time difference measured from one second to the next one indicates that the advance algorithm in the server has been fooled by a change in round-trip delay that was accompanied by a change in symmetry. For example, it is possible that the change in the measured round-trip delay was really largely confined to either the inbound or outbound paths. The client can detect this possibility by noting the change in the measured time difference between the ACTS time and its system clock and the change in the measurement of the round-trip delay that the server has inserted into the transmission. As a simple example, if the change in the delay is confined to the outbound path between the server and the client, the server will see this as a change in the total round-trip delay and will advance the next OTM by one-half of this value. This is exactly one-half of the correct advance change so that the next OTM will not arrive on time by one-half of the change in the advance. The client will detect that the advance parameter has changed and that the

measured time difference has changed by one-half of that amount. This more sophisticated algorithm in the client system can almost completely compensate for the degraded stability of the dial-up telephone system, and the more sophisticated algorithm can transmit time over standard dial-up telephone lines with an uncertainty of order 0.5–0.8 ms RMS. This is about a factor of 10 better than the Internet time servers because the delay through the telephone system is more stable and more symmetric than the delay through a wide-area packet network.

59.11.4.3 Two-Way Satellite Time Transfer This method is used to compare the time scales of National Metrology Institutes and Timing Laboratories and to transmit time and frequency information to the International Bureau of Weights and Measures (the BIPM, in French) for the purpose of computing TAI and UTC.

The method uses the same assumption as in the previous discussions: the one-way delay can be estimated as one-half of the measured round-trip value. The configuration of the message exchange is similar to the NTP exchange discussed previously.

Each station encodes the 1 Hz tick of its local time scale using a pseudorandom code and a subcarrier whose frequency is about 70 MHz. The subcarrier is transmitted up to a communications satellite, which is in a geostationary orbit. (The satellite is located above the equator. The radius of its orbit is ~40,000 km, and its orbital period is 24 h. It therefore appears to be stationary with respect to an observer on the Earth.) The satellite retransmits the modulated signal down to the receiver, where the 1 Hz tick is recovered by a cross-correlation of the received pseudorandom code with a copy of the code generated in the receiver. The time difference between the recovered 1 Hz tick and the local clock is then stored. The message exchange is full duplex, and the time differences at each station are combined to estimate the time difference of the clocks at the two sites.

The uplink and downlink typically use different frequencies in the Ku band. The uplink frequency is nominally 14 GHz, and the downlink is nominally 11 GHz. Both frequencies are used on a portion of the path in each direction, but the paths are not the same so that the delays in the two directions may not be exactly equal. The dispersion of the refractivity of the ionosphere and the troposphere is small at these frequencies so that this asymmetry is generally not an important limitation. Balancing the transmit and receive delays in the ground station hardware is a more difficult problem, especially because these delays are often sensitive to fluctuations in the ambient temperature.

The transit time from one station up to the satellite and down to the other station is about 0.25 s so that the rotation of the Earth during this time must be taken into account. This is called the Sagnac [19] effect. The magnitude of this effect is $2\omega A/c^2$, where c is the speed of light, ω is the angular velocity of the Earth, and A is the area defined by the triangle formed by the satellite and the two stations projected onto the equatorial plane. The effect is positive for a message traveling eastward and negative for messages in the opposite direction.

The design of this method treats both stations as equal partners in the exchange. The transmissions at each end are generated by the local time scale and are not a response to a message received from the other end as with NTP. In principle, this is an important difference between this method and the NTP and ACTS methods described earlier, but the underlying assumptions of all of the methods are the same, and only the details of the analysis are somewhat different.

Based on the notation of the discussion of the NTP message exchange, a well-designed NTP system will have a very small time delay, $T_{3b} - T_{2b}$, between when a system receives a query and when it responds; the ACTS protocol assumes that this difference is negligible; the two-way system makes no assumptions about this difference except that it is accurately measured on each message exchange.

The time differences measured with the two-way satellite method are much more accurate than the measurements made with either of the systems discussed previously because the delays are much more stable and the assumption of the very small delay asymmetry is more accurate. The RMS uncertainty of the measurements is of order 0.1 ns. The spectrum of the noise is approximately white phase noise for short averaging times, but there are less favorable variations at longer periods; some links have a quasi-periodic approximately diurnal variation in the time-difference data. The source of this variation is not understood at present.

59.12 THE POLLING INTERVAL: HOW OFTEN SHOULD I CALIBRATE A CLOCK?

There are three different considerations that should be used to determine the interval between time-difference measurements. The first consideration is based on a statistical estimate of the noise processes, the second is derived from concerns about nonstatistical errors, and the third is based on a cost–benefit analysis. I will begin by considering choosing a polling interval based on a statistical analysis.

From the statistical perspective, the goal of the time-difference measurements is to improve the accuracy or the stability of the local clock once its deterministic properties have been determined and used to adjust the time, frequency, and aging (if appropriate) of the clock under test. The deterministic parameters can be applied directly to the physical device, or they can be used to adjust the readings of the clock administratively. (In general, timing laboratories usually do not adjust the physical parameters of a clock but rather apply the deterministic offsets administratively.)

Once the deterministic parameters have been included in the readings of the device under test, I assume in a statistical analysis that its time dispersion can be determined solely from its statistical characterization. At least at the beginning, I will assume that both the deterministic and stochastic parameters are constants that do not depend on time or on perturbations such as fluctuations in the ambient temperature.

In this model, the design of the calibration procedure is driven by the requirement that the accuracy or stability of the remote clock as seen through the channel should be better than the corresponding parameters of the local clock for the same query interval. (The channel includes the physical medium used to transmit the time signal and any measurement hardware at the end stations.) As a practical matter, it often turns out that the statistical characteristics of the channel are much poorer than the statistics of the remote clock itself. In this situation, which is quite common, improving the accuracy or the stability of the remote clock will have almost no effect on the performance of the synchronization process, which will be dominated by the statistics of the channel. The following discussion depends only on being able to characterize the combination of the remote clock and the channel by means of the two-sample Allan deviation. The analysis is not sensitive to whether the source of the fluctuations is in the clock or the channel connecting it to the device under test.

Suppose that the statistical characteristics of the remote clock seen through the channel can be described as white phase noise for all averaging times. This is the best that we can hope for—the measurement process using this channel is degraded by a noise process that has a mean of zero and a standard deviation that does not depend on the time the measurement was performed or on any external parameter such as the ambient temperature. The magnitude of the two-sample Allan deviation (the square root of the variance) that describes the statistics of the remote clock seen through this channel varies as the reciprocal of the averaging time. The time dispersion in this configuration is independent of the averaging time. (see Equations 59.10 and 59.11.) This is not a surprising result. If the measurement noise is characterized as white phase noise, then the fluctuations of every measurement are derived from the same distribution with a mean of zero and a fixed standard deviation, and there is no relationship between one measurement and any other one so that the time between measurements is irrelevant to the statistics of the time differences.

Now consider that the stability of the local clock is characterized as pure white frequency noise for moderate averaging times. Again, this is not a surprising result and is about the best we could ever hope to see; once the white phase noise of the measurement process has been accounted for, the next effect is the noise of the frequency control loop, which we also take to be a process with a mean of zero and a known standard deviation. (There are almost always longer-period effects that modify the assumption of pure white frequency noise, but I will assume for now that the averaging times that will be used will not be large enough to make this consideration important.) The two-sample Allan deviation for white frequency noise varies as the reciprocal of the square root of the averaging time so that time dispersion due to white frequency noise increases as the square root of the averaging time. (I will use the conservative estimate that I discussed in the preceding text for Eq. 59.11.)

We can now combine these two results to define two measurement strategies. The noise of the time-difference measurement process is characterized by a standard

deviation, M , which does not depend on averaging time. The time dispersion of the local clock due to its white frequency noise is characterized as a function of averaging time by $C\tau^{1/2}$. From the statistical perspective, the goal of the first synchronization procedure is to set the averaging time so that the remote clock seen through the channel is more stable than the local clock. In other words, the averaging time should be chosen so that the free-running time dispersion of the local clock due to its frequency noise is greater than the time dispersion of a time-difference measurement with respect to the remote clock seen through the channel:

$$\begin{aligned} C\tau^{1/2} &\geq M \\ \tau &\geq \left(\frac{M}{C}\right)^2. \end{aligned} \quad (59.39)$$

The result may be surprising but is easily explained. As the remote clock seen through the network becomes less stable (increasing M), the crossover between its stability and the stability of the local device, which increases as the square root of the averaging time, moves to longer and longer averaging times. Thus we would expect that the optimum polling interval for a time transfer that used the wide-area Internet to communicate between the local and remote devices would be longer than the optimum polling interval for the same devices that exchanged messages on a local network connection because the stability of the transmission delay in a wide-area network would be poorer.

Since the channel connection back to the remote clock is characterized by white phase noise, a second strategy would be to make measurements of the time difference as rapidly as possible and average the data. For example, suppose we could make measurements every second for a time interval of T seconds. If we averaged these T measurements, the standard deviation of the mean would be reduced from M to M/\sqrt{T} . The time dispersion due to the white frequency noise of the local device would be the same as before so that the comparison of Equation 59.39 becomes

$$\begin{aligned} CT^{1/2} &\geq \frac{M}{T^{1/2}} \\ T &\geq \frac{M}{C}, \end{aligned} \quad (59.40)$$

where the value of T in Equation 59.40 specifies the point at which the noise of the local clock is greater than or equal to the standard deviation of the average so that the averaging algorithm can improve the stability of the local clock starting at an averaging time of T . In this simple model, once this time is reached, additional averaging only makes things better because the standard deviation of the remote clock seen through the channel improves without bound, while the stability of the local clock degrades without bound. In the limit of very large T , we are not using the data from the local clock at all.

This simple model will break down at some point for one of two reasons. The first possibility is that channel noise is not purely white phase noise starting at some averaging time—longer-period fluctuations become important, and they are not zero-mean random processes. These fluctuations can be incorporated by modeling the time dispersion of the remote clock, M , as constant at short times but increasing as some power of the averaging time starting at some averaging time, T_m . The second possibility is that the requirements of the application limit the averaging time—we cannot average forever because we need to use the time difference for some application. The assumptions that I used in this discussion are somewhat artificial in that they are often better than real-world devices and channels. Therefore, these calculations are more illustrative of the method than rigorous derivations with very general applicability.

59.13 ERROR DETECTION

Any measurement protocol that receives data from a remote device over a noisy channel should be prepared to consider the possibility that the received data are in error, either because the remote clock has failed or the channel characteristics have changed suddenly. A purely statistical analysis cannot be the whole story here, since there is generally no objective way of distinguishing between an error and a very low-probability event that conforms to the statistical description. One method that is commonly used is to regard a measurement that differs from the mean (or from the predicted value) by more than three standard deviations as an error.

The machinery that I developed in the previous section can also be used to detect possible errors. For example, consider the averaging strategy presented in the discussion for Equation 59.40. Instead of waiting until all of the measurements have been completed to evaluate the average time difference, we could construct a running mean with an update each time a new time difference was acquired. The estimate of the mean at the k th step, \bar{X}_k , after receiving the time difference x_k can be calculated iteratively based on the estimate of the mean at the previous step,

$$\bar{X}_k = \frac{(k-1)\bar{X}_{k-1} + x_k}{k}, \quad (59.41)$$

where the estimate of the mean is initialized to zero.

One possibility is to ignore the k th estimate as having a one-time error if it differs from the running mean by more than three times the running estimate of the standard deviation computed from the current average or from a previous one:

$$\left| (x_k - \bar{X}_{k-1}) \right| > 3\sigma_{k-1} \quad (59.42)$$

The assumption that the difference is a one-time error would be confirmed if the next reading was consistent with the running mean value.

The situation becomes more complicated if the next measurement does not conform to the running mean either, and it may not be possible to distinguish between a problem with the remote clock, the local clock, or the channel. It is sometimes possible to decide this question if a second independent calibration source or an independent channel is available. This solution must be considered in the cost–benefit analysis that I will discuss in the next section.

59.14 COST–BENEFIT ANALYSIS

In this section I will consider the situation where a time-difference measurement has a finite cost in terms of computer cycles, network bandwidth, or some other finite resource. The trade-off between the accuracy of a time-difference measurement and the cost needed to realize it becomes important in this situation.

For example, consider again the simple case where the time-difference measurements are characterized as white phase noise with a mean of zero. The standard deviation of the mean of N measurements decreases as $1/\sqrt{N}$, and this improvement can continue without bound in principle. On the other hand, the cost of the measurement process increases linearly with N , assuming that each measurement has the same cost. In this situation, the cost–benefit analysis is always unfavorable—the cost of the measurements always increases faster than the standard deviation improves, and the best cost–benefit strategy is to make the minimum number of measurements consistent with the standard deviation that is required to meet the needs of the application.

More generally, I assume that the total cost of a measurement procedure, C , is given by the cost of a single measurement, c ; the interval between measurements, τ ; and the total measurement time, T :

$$C = c \frac{T}{\tau}. \quad (59.43)$$

I take the benefit of the procedure, B , as the time dispersion of the device for an averaging time, τ , where the time dispersion is calculated from the two-sample Allan deviation for that averaging time:

$$B = \sigma_y(\tau)\tau. \quad (59.44)$$

The goal is then to minimize the product BC , possibly with some additional constraint that the time dispersion must be less than some value required by the application.

Apart from the constants, the product BC is a function only of the two-sample Allan deviation so that it will always improve with increasing averaging time as long as the slope of the Allan deviation is negative, and the best strategy will be the longest averaging time that satisfies the time dispersion estimate in Equation 59.44. The slope

of the two-sample Allan deviation is negative in white phase noise and white frequency noise domains so that a pure cost–benefit analysis will always favor an averaging time that is at the onset of flicker processes where the slope of the Allan deviation approaches zero. The cost–benefit analysis product is a constant independent of averaging time in the flicker domain, but the time dispersion increases with averaging time (Eq. 59.44) so that the dispersion may not satisfy the requirements of the application in this domain. The cost–benefit analysis becomes unfavorable in the random walk of frequency domain where the slope of the two-sample Allan deviation is positive. The time dispersion of the local clock is increasing faster than the cost is decreasing in this region. This region might still be an acceptable choice if the accuracy requirement is very modest.

The method used for detecting errors also has a cost–benefit aspect. For example, if an Internet client queries N Internet servers on every measurement cycle in an attempt to detect an error, then the cost of the synchronization process has increased by a factor of N ; the benefit will depend on how often this procedure detects a problem. Shortening the polling interval to detect a problem with the local clock more quickly is subject to the same considerations. That is, do problems happen often enough to justify the increased cost of the algorithm? In general, comparing the measured time difference with a prediction based on the statistics of the local clock (e.g., Eq. 59.42) and querying multiple servers only when that test fails are a better strategy because it exploits the statistics of the local clock as a method for detecting a possible error with the remote clock or with the channel.

A cost–benefit analysis is very important from the perspective of the operators of the network and the public time servers—increasing the polling interval and reducing the number of servers queried on each measurement cycle translates directly into the number of users that can be supported with available, scarce resources.

59.15 THE NATIONAL TIME SCALE

The official time in the United States (and in most other countries) is UTC. The length of the UTC second is defined by the frequency of the hyperfine transition in the ground state of cesium. The frequency of this transition is defined to be 9,192,631,770 Hz, and counting this number of cycles defines the length of the second. The other time units (minutes, hours, ...) are multiples of this base unit.

The length of the day, computed as 86,400 Cs, is somewhat shorter than the length of the day in the UT1 time scale, which is a time scale based on the rotation of the Earth. The accumulated time difference is currently somewhat less than 1 s/year. In order to maintain a close connection between atomic time, defined by the cesium transition frequency, and the UT1 time scale, which characterizes the position of the Earth in space, additional seconds are added to UTC whenever the difference between UTC and UT1 approaches 0.9 s. The decision to add these “leap seconds” is made by

the International Earth Rotation and Reference System Service, and all national timing laboratories incorporate the leap second into their time services.

Leap seconds are normally added as the last UTC second of the last day of June or December. In the vicinity of a leap second, the time stamps are 23:59:58, 23:59:59, 23:59:60, and then 00:00:00 of the next day. Digital time services and most clocks cannot represent the leap second time of 23:59:60 and stop the clock for 1 s at 23:59:59. The time services operated by the National Institute of Standards and Technology implement the leap second by transmitting a time value equivalent to 23:59:59 twice, and most other time services do the same thing. Assigning the same time stamp to two consecutive seconds is ambiguous and has obvious difficulties with respect to causality, and the question of continuing the leap second procedure is currently (as of 2015) being discussed.

The details of the leap second procedure are important for users who must synchronize a clock to the official time scale in the vicinity of a leap second. Unfortunately, some time services implement the leap second in different ways. One method adds the leap second by duplicating the time 00:00:00 of the next day. This eventually results in the same time as the NIST method, but it adds the leap second in the wrong day and has time errors on the order of 1 s in the immediate vicinity of a leap second.

A more troubling method implements the leap second as a frequency adjustment during the last few minutes of the day. The clock is slowed down for some period of time until the additional second has been added. This method has both a time error and a frequency error during the time the leap second is being inserted. The clock is never stopped in this implementation, but both the time and the frequency are not correct with respect to national time standards during this interval. In addition, there is no generally accepted method for implementing the frequency adjustment so that different implementations of this method will also have errors with respect to the national standards of time and frequency in this vicinity of a leap second.

In addition to the two proposals, (i) not to make any changes or (ii) to stop adding future leap seconds to UTC but to continue the number of leap seconds that have already been added, a number of other alternatives have been suggested. One proposal would be to switch to TAI as the legal time scale, which would effectively reset the leap second count to zero in a single step. Another proposal would be to stop adding leap seconds to UTC and to change the name of the time scale to reflect this change in its implementation.

59.16 TRACEABILITY

There are many applications that require time stamps that are traceable to a national time scale, and realizing this requirement requires clocks that are synchronized to a national or international standard of time. A clock is *traceable* to a national time scale if there is an unbroken chain of time-difference calibration measurements between the

clock and the reference time scale by means of any of the methods that I have described before. Each one of these measurements must have an uncertainty estimate.

It can be difficult to establish the chain of measurements that is required for traceability. For example, the signal *in space* transmitted by a GPS satellite is traceable to UTC, the national and international reference time scale, through the US Naval Observatory, which monitors the time signals broadcast by the GPS constellation and computes the offset between GPS system time and UTC as maintained by the Naval Observatory. This offset is uploaded into the satellites and is transmitted as part of the navigation message.

However, the traceability of the signal in space does not necessarily extend to the timing signals produced by a GPS receiver unless the receiver, the antenna, and the connecting cable have been calibrated. The traceability almost certainly does not automatically extend to the application that uses the timing signals to apply time stamps as part of some application. This discussion does not suggest that these links in the chain are known to be inadequate or in error, but rather that they do not satisfy the strict definition of traceability without some sort of calibration procedure.

There are some situations where the requirements of strict traceability can be satisfied without a complex calibration procedure. For example, if an application requires that time stamps be traceable to a national time standard with an uncertainty of less than 1 s, then simply certifying that the satellite timing equipment is working properly is likely to be good enough. The uncertainty of the time signals produced by a receiver synchronized using signals from a GPS satellite is several orders of magnitude smaller than the requirements of the application so that the overall system is surely traceable at the level of 1 s if it is working at all. (Verifying that a GPS receiver is working properly may or may not be an easy job—it depends on the specific receiver that is used.)

A second aspect of traceability is *legal traceability*, by which I mean being able to establish in a legal proceeding that a time stamp was traceable to a national time scale. In this situation, “doing the right thing” might not be adequate if you can’t prove it to a judge and jury.

Given that the technical aspects of traceability that I discussed previously have been satisfied, establishing legal traceability is generally a matter of documentation—maintaining log files that show that the system was calibrated with an uncertainty consistent with the requirements of the application and that it was operating normally at the time in question. A log file that has entries only when there is a problem is unlikely to be adequate—it will have no entries when the system is working properly, and an empty log file is ambiguous and may not be of much help.

59.17 SUMMARY

I have discussed a number of methods for synchronizing a clock using a reference device that can be located either in the same facility or remotely and linked to the device under test by a communication channel. I have discussed a number of methods

of synchronizing a remote clock and the statistical considerations that characterize the accuracy of the procedure and how often to request a calibration. An important tool in these discussions is the two-sample Allan variance, and I have presented a simple introduction into how this estimator is calculated and used.

59.18 BIBLIOGRAPHY

The following list contains a few of the very large number of publications that contain additional information on time and frequency standards and distribution methods:

1. The publications of the Time and Frequency Division of the National Institute of Standards and Technology, generally available online at tf.nist.gov.
2. “Encyclopedia of Time.” Edited by Samuel L. Macey, New York, 1994, Garland Publishing, Inc.
3. Special issue on Time Scale Algorithms, *Metrologia*, Vol. 45, Number 6, December, 2008.
4. “Time and Frequency Measurement.” Edited by Christine Hackman and Donald B. Sullivan, College Park, MD, 1996, American Association of Physics Teachers. This was also published as resource letter TFM-1 in the *American Journal of Physics*, Vol. 63, Number 4, pages 306–317, April, 1995.
5. “Computer Network Time Synchronization,” 2nd edition. David L. Mills, New York, 2011, CRC Press.
6. Special Issue on Time and Frequency, *Proceedings of the IEEE*, Vol. 79, Number 7, July, 1991.
7. “Understanding GPS: Principles and Applications,” 2nd edition. Edited by Elliott D. Kaplan and Christopher J. Hegarty, Norwood, MA, 2006, Artech House, www.artechhouse.com.
8. “Global Positioning System: Signals, Measurements and Performance,” 2nd edition. Edited by Pratap Misra and Per Enge, Lincoln, MA, 2006, Ganga-Jamuna Press, GPStextbook@G-JPress.com.

REFERENCES

1. S. Bregni, “Measurement of the Maximum Time Interval Error for Telecommunications Clock Stability Characterization,” *IEEE Trans. Instrum. Meas.*, Vol. 45, pages 900–906, October 1996.
2. K. Biholar, T1.101-199X, Synchronization Interface Standard, Draft Standard of the American National Standards Institute, Inc.
3. S. R. Stein, “Frequency and Time—Their Measurement and Characterization,” *Precision Frequency Control*, Vol. 2, New York, Academic Press, 1985. This paper and a number of

- others on the same topic are reprinted in NIST Technical Note 1337, edited by D. B. Sullivan, D. W. Allan, D. A. Howe, and F. L. Walls, published by the US Department of Commerce, March 1990.
4. B. D. O. Anderson and J. B. Moore, *Optimal Filtering*, New York, Dover Publications, 1979, pages 54, ff.
 5. R. B. Blackman and J. W. Tukey, *The Measurement of Power Spectra*, New York, Dover Publications, 1959, pages 174, ff.
 6. Y. Lin, S. Lee, S. Li, Y. Xie, Z. Ren, and C. Nguyen, "Series-Resonant VHF Micromechanical Resonator Reference Oscillators," *IEEE J. Solid State Circuits*, Vol. 39, No. 12, pages 2477–2491, December 2004. See also Ville Kaajakari, *Practical MEMS*, Las Vegas, Small Gear Publishing, 2009. www.smallgearpublishing.com (accessed November 11, 2015).
 7. C. Audoin and B. Guinot, "Atomic Frequency Standards," *The Measurement of Time: Time, Frequency and the Atomic Clock*, Cambridge, Cambridge University Press, 2001.
 8. H. G. Andresen and E. Pannaci, "Servo controlled hydrogen maser cavity tuning," Proceedings 20th Annual Frequency Control Symposium, Atlantic City, NJ, Fort Monmouth, NJ, Electronic Components Laboratory, U.S. Army Electronics Command, 1966, pages 402–415. See also, C. Audoin, "Fast Cavity Auto-Tuning System for Hydrogen Maser," *Rev. Phys.*, Vol. A16, pages 125–130, 1981.
 9. G. E. P. Box and G. M. Jenkins, *Time Series Analysis: Forecasting and Control*, San Francisco, CA, Holden-Day, 1970.
 10. D. A. Howe, D. W. Allan, and J. A. Barnes, "Properties of signal sources and measurement methods," Proceedings 35th Annual Symposium on Frequency Control, Washington, DC, Electronic Industries Association, 1981, pages A1–A47. This is document AD-A110870 available from National Technical Information Service, 5285 Port Royal Road, Springfield, Virginia 22161.
 11. S. Stein, D. Glaze, J. Levine, J. Gray, D. Hilliard, D. Howe, and L. Erb, "Performance of an automated high accuracy phase measurement system," Proceedings 36th Annual Symposium on Frequency Control, Fort Monmouth, NJ, U.S. Army Electronics Research and Development Command, Electronics Technology and Devices Laboratory, 1982, pages 314–320. This is document AD-A130811 available from National Technical Information Service, 5285 Port Royal Road, Springfield, Virginia 22161.
 12. J. Levine, "Introduction to Time and Frequency Metrology," *Rev. Sci. Instrum.*, Vol. 70, No. 6, pages 2567–2596, 1999.
 13. G. Panfilo, A. Harmegnies, and L. Tisserand, "A new prediction algorithm for EAL," Proceedings of the 2011 Joint European Time and Frequency Forum and International Frequency Control Symposium, Piscataway, NJ, IEEE, 2011, pages 850–855.
 14. J. Levine, "The Statistical Modeling of Atomic Clocks and the Design of Time Scales," *Rev. Sci. Instrum.*, Vol. 83, page 021101, 2012.
 15. R. H. Jones and P. V. Tryon, "Estimating Time From Atomic Clocks," *J. Res. Natl. Bur. Stand.*, Vol. 88, pages 17–28, 1983. See also A. Gelb, *Applied Optimal Estimation*, Cambridge, MA, MIT Press, 1974, Chapter 4.
 16. Postprocessed orbits and clock solutions are computed by the International GPS Service and are available online at www.igs.org and <https://www.igsb.jpl.nasa.gov/components/prods.html> (accessed December 15, 2015).

17. Many references. See, for example, D. E. Comer, "Internetworking with TCP/IP," *Principles, Protocols, and Architecture*, Vol. 1, Englewood Cliffs, NJ, Prentice-Hall, 1991.
18. J. Levine, M. Weiss, D. D. Davis, D. W. Allan, and D. B. Sullivan, "The NIST Automated Computer Time Service," *J. Res. NIST*, Vol. 94, pages 311–322, 1989.
19. N. Ashby and M. Weiss, *Global Positioning Receivers and Relativity*, National Institute of Standards and Technology Technical Note 1385, Boulder, CO, U.S. Department of Commerce, Technology Administration, National Institute of Standards and Technology, 1999.

LABORATORY-BASED GRAVITY MEASUREMENT

CHARLES D. HOYLE, JR.

Department of Physics and Astronomy, Humboldt State University, Arcata, CA, USA

60.1 INTRODUCTION

Gravitational experiments invariably inspire images of planetary and galactic measurements, black hole mergers, large-scale gravitational wave detectors, and other massive undertakings. However, laboratory-scale tests of gravity involve relatively small masses and introduce many systematic effects not present in astronomical-scale investigations. In this regime, it is appropriate to discuss gravity as the Newtonian (weak-field) limit of General Relativity. Thus, we expect the gravitational force to obey the inverse-square law (ISL) if General Relativity is valid when small test masses are in close proximity. The higher order effects of General Relativity are not relevant at this scale because they are miniscule for any test masses that can be used in the laboratory. Therefore, tests of the ISL are effectively measurements of the validity of General Relativity at these scales. In addition, by selection of appropriate test masses and test mass geometries, the weak equivalence principle (WEP), a central feature of General Relativity, may also be effectively tested at short distances. Difficulty arises because the relative weakness of gravity compared to the other fundamental forces makes precision measurements of gravitational physics at laboratory scales precise and challenging work.

In addition to testing General Relativity itself, the motivation for improving tests of gravity at the laboratory scale spans many areas ranging from particle physics and cosmology to metrology and precision measurement. Tests of gravitational physics in this regime fall broadly in to five categories: tests of the Newtonian ISL, investigation of

the WEP, measurements of the gravitational constant (G), searches for new long-range forces that may couple to a variety of physical quantities or natural symmetries, and searches for effects that may yield insight into the nature of dark matter and/or dark energy. Many experiments are designed to test multiple areas at once by carefully choosing particular test masses and experimental configurations.

Due to the weakness of gravitational force between small test masses, it is necessary to eliminate any effects due to the Earth's gravitational field and other disturbances due to electromagnetic, thermal, and seismic effects in these experiments. Traditionally, torsion pendulums and other precision oscillators have been employed to decouple the experiments from environmental and systematic effects. However, novel atomic, molecular, and nuclear techniques are creating a new experimental arena for laboratory gravitational tests.

The question naturally arises, "What is meant by 'laboratory-scale'?" In the context of this chapter, "laboratory scale" means any experimental apparatus that is self-contained within the confines of a standard laboratory space and employs test masses small enough that the general relativistic corrections to the Newtonian ISL are smaller than the experimental resolution or uncertainty. This does not mean, however, that the effects under investigation are limited to operate only on this distance scale. For example, a long-range deviation from the ISL or WEP can be detected in certain laboratory-scale experimental configurations.

60.2 MOTIVATION FOR LABORATORY-SCALE TESTS OF GRAVITATIONAL PHYSICS

In recent years, tests of gravity have experienced a resurgence due in a large part to models of string or M-theory that predict modification of the gravitational force on short distance scales due to the influence of macroscopic extra spatial dimensions (for a review, see Ref. 1). However, many other theoretical scenarios predict possible violations of the ISL and WEP due to a variety of phenomena.

One of the most definitive string theory predictions in recent years suggests that gravity's strength increases at distances comparable to the size of proposed compactified spatial dimensions [2]. Other models predict weakening of gravity at small scales [3]. Such scenarios are proposed in an attempt to solve the gauge hierarchy problem (the discrepancy in energy scales between the Planck mass and the electroweak scale).

Attempts to explain the observed cosmic distance scale acceleration have also shown that the data would be consistent with a theory that predicts gravity to "turn off" at distances less than about 0.1 mm [4]. The observed value of the vacuum energy density, $\rho_{\text{vac}} \approx 3.8 \text{ keV/cm}^3$, corresponds to a length scale of $R_{\text{vac}} = \sqrt[4]{\hbar c / \rho_{\text{vac}}} \approx 85 \mu\text{m}$, which may also have fundamental significance [5].

Finally, unobserved particles predicted by string theories, such as the dilaton and moduli, may also produce new short-range forces operating through the "chameleon mechanism" [6] that could be observed in short-range tests of gravity.

Most alternative models of gravity predict a violation of the WEP at some level due to interactions coupled to quantities other than mass or modifications of gravity itself. Scalar or vector boson exchange produces forces that inherently violate the WEP over a range determined by the Compton wavelength of the exchange particle, $\lambda = h/m_b c$. The WEP has been tested with incredible precision over distance scales from 1 cm to ∞ [7–9] but has never been subjected to a dedicated test in the subcentimeter regime (corresponding to exchange boson masses $\gtrsim 0.1$ meV).

Measurements of the gravitational constant, G , have yielded widely varying results in recent decades, making it one of the least well-measured fundamental constants [10, 11]. Efforts to understand the discrepancy and determine an accurate value G are at the forefront of precision measurement research.

60.3 PARAMETERIZATION

A deviation from ISL behavior is generally modeled using a Yukawa addition to the classical Newtonian potential energy [1]. For point masses m_1 and m_2 separated by distance r , the modified potential energy becomes

$$V(r) = -\frac{Gm_1m_2}{r} \left(1 + \alpha e^{-r/\lambda}\right), \quad (60.1)$$

where G is the Newtonian gravitational constant, α is a dimensionless scaling factor corresponding to the strength of any deviation relative to Newtonian gravity, and λ is the characteristic length scale of the deviation.

It is generally assumed that a WEP violation would result in a coupling to some “charge” that is related to the seemingly conserved quantities of baryon number, B (atomic number, Z , plus neutron number, N), or lepton number, L ($L = Z$ for electrically neutral materials). A general scalar (–) or vector (+) Yukawa coupling would result a potential energy of the form [8]

$$V(r) = \mp \frac{1}{4\pi} \tilde{q}_1 \tilde{q}_2 \frac{e^{-r/\lambda}}{r}, \quad (60.2)$$

where \tilde{q}_i are the “charges” of the test masses and λ is the Compton wavelength of the exchange boson. A common parameterization assumes $\tilde{q} = \tilde{g} [Z \cos \tilde{\psi} + N \sin \tilde{\psi}]$ where \tilde{g} is a coupling constant and $\tilde{\psi}$ determines the type of charge. Recasting Equation 60.2 in a form similar to Equation 60.1 yields

$$V(r) = -\frac{Gm_1m_2}{r} \left(1 + \tilde{\alpha} \left[\frac{\tilde{q}}{\tilde{g}\mu} \right]_1 \left[\frac{\tilde{q}}{\tilde{g}\mu} \right]_2 e^{-r/\lambda} \right), \quad (60.3)$$

where μ is the mass of objects 1 or 2 in units of atomic mass units, u , and $\tilde{\alpha} = \pm \tilde{g}^2 / (4\pi G u^2)$.

60.4 CURRENT STATUS OF LABORATORY-SCALE GRAVITATIONAL MEASUREMENTS

As discussed in the following sections, there is an ever-increasing variety of experimental techniques for exploring gravity at laboratory distances. The workhorse of gravitational measurements is the torsion pendulum [8], although techniques such as high-frequency oscillators and atomic/molecular interferometry are promising new techniques in this field [12, 13]. The following sections summarize the current level of experimental results obtained from laboratory-scale experiments.

60.4.1 Tests of the ISL

Tests of the ISL can be considered as measurements of the parameters α and λ of Equation 60.1. Previous experiments utilizing various types of torsion pendulums have eliminated large portions of the Yukawa potential α - λ parameter space [8, 14–18]. The shaded region of Figure 60.1 shows the current constraints in the α - λ plane for λ between a few micrometers up to 1 cm. Laboratory tests at larger ranges have been performed [1, 12] and have also discovered no deviation from Newtonian behavior.

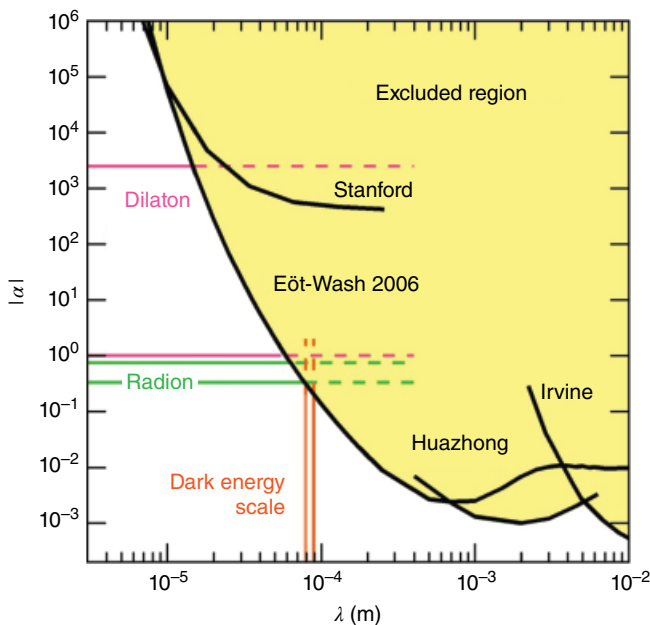


FIGURE 60.1 Current short-range experimental constraints in the $|\alpha|$ - λ Yukawa parameter space and theoretical predictions. The shaded region is excluded at the 95% confidence level. Results from previous experiments are shown by the curves labeled Stanford [18], Eöt-Wash [15], Huazhong [16], and Irvine [17]. For a more detailed discussion of theoretical predictions, see Ref. 1.

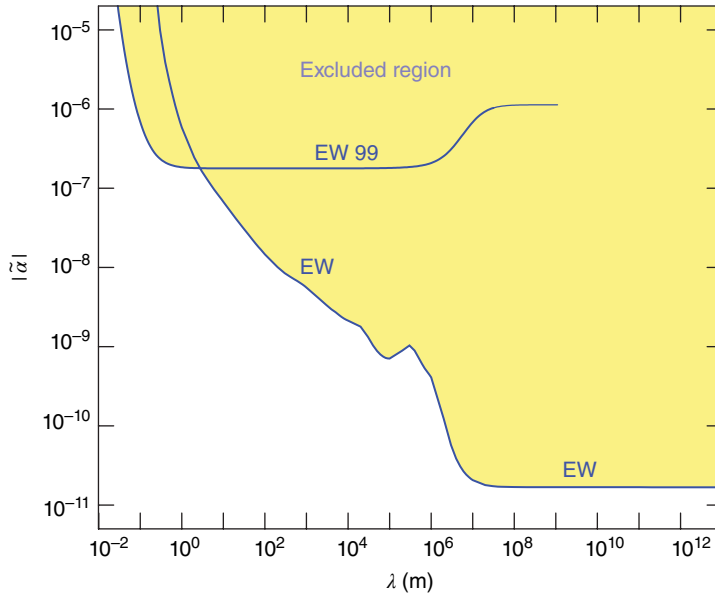


FIGURE 60.2 Best 95% confidence level constraints on violations of the WEP for interactions coupling to $\tilde{q} = B - L$. The curve labeled EW is from reference [7], while EW 99 is from reference [9]. Both limits were obtained with torsion pendulum experiments performed by the Eöt-Wash group at the University of Washington.

60.4.2 WEP Tests

The WEP is well characterized over distances from 1 cm to ∞ [7–9], but it is essentially untested below the centimeter scale. Various laboratory torsion pendulum experiments have set these constraints as summarized in Figure 60.2.

60.4.3 Measurements of G

The gravitational constant was one of the first to be measured with some precision by Cavendish (using a torsion balance) in 1798 [19]. It is surprising, therefore, that measurements performed over the last few decades involving a variety of techniques have yielded inconsistent results. In fact the published CODATA values of G have been essentially unchanged for decades even though there have been many precision measurements performed [11]. The inconsistency of these measurements has left the uncertainty in G unchanged. For a summary of recent measurements, see Figure 3 of Refs. 10 and 20.

60.5 TORSION PENDULUM EXPERIMENTS

Torsion pendulums (or torsion balances) have been used for gravitational measurements since at least the time of Cavendish [19]. A torsion pendulum is only sensitive to torque about the vertical axis and is therefore insensitive to the large vertical

component of the Earth's gravitational field or any uniform horizontal forces. Modern torsion pendulums have incredibly high torque sensitivity and are the most sensitive devices for measuring feeble forces between macroscopic objects. For a nice review of torsion pendulums used in fundamental physics experiments, see Ref. 8.

60.5.1 General Principles and Sensitivity

Torsion pendulums, when operated in a vacuum, behave as lightly damped harmonic oscillators in a single rotational dimension whose rotation axis is local vertical. Measurements of gravitational effects are performed by carefully choosing the mass distribution of the pendulum and applying a torque via a specially designed attractor mass. Measurement of the applied torque can be obtained through analysis of the pendulum's twist angle (time series analysis) or free oscillation period (torsion period analysis). For experiments where the applied attractor mass torque is time varying, the time series analysis is advantageous, whereas static attractor configurations sometimes measure the torsion period and the harmonic content of the motion to extract the gravitational influence on the pendulum [8]. Other tests have successfully employed feedback loops to maintain the pendulum stationary [21]. Using this technique, the torque is inferred from the applied feedback signal. An ideal torsion pendulum experiment is null; that is, one in which Newtonian gravity has no effect on the pendulum. Most modern experiments make an effort to achieve an experimental configuration that approximates a true null as much as is realistically possible.

Because torsion pendulums are resonant systems, the angular response to applied torques is frequency dependent. Torque applied on resonance yields the largest angular response; however, noise sources and false effects are also enhanced the most on resonance, and the best results to date have used off-resonance driving techniques [15]. Precision angular measurement is therefore paramount in torsion pendulum experiments. Optical autocollimators have been developed to measure small angular deflections of order 1 nrad or less. Recent developments have begun to push this sensitivity even further [22]. An example of one such system is the one currently in use by the author's group at the Humboldt State University (HSU) Gravitational Research Laboratory. A sample noise spectrum for this autocollimator is shown in Figure 60.3. The autocollimators typically employ position-sensitive photodiodes to measure the deflection of a laser diode beam used as an optical lever. To increase sensitivity, the beam is typically chopped and recorded using lock-in techniques.

60.5.2 Fundamental Limitations

60.5.2.1 Statistical Noise Sources Thermal and readout noise set the sensitivity limit for any torsion balance experiment. In the nonviscous vacuum regime (typically below 10^{-4} – 10^{-5} Torr), thermal noise within the torsion fiber itself is a limiting factor.

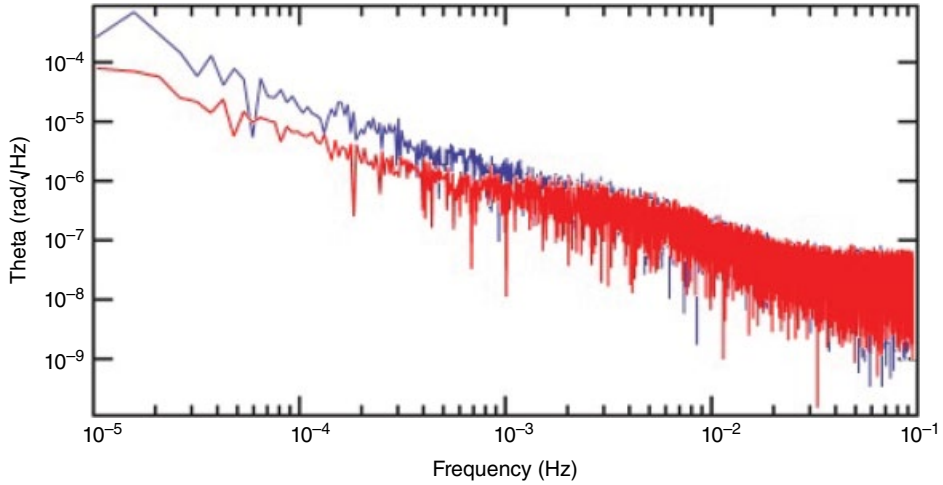


FIGURE 60.3 Sample noise spectrum of the Humboldt State autocollimator twist data taken off of a stationary mirror. The noise level at 5 mHz is $0.3 \mu\text{rad}/\sqrt{\text{Hz}}$, which yields an uncertainty in angle measurement of 1 nrad in roughly 1 day of integration time. The lower curve was taken with the laser off (electronic noise only), while the upper is with the laser pulsing and incident on the position-sensitive detector.

Internal damping in the fiber material produces frequency-dependent torque noise with amplitude spectral density

$$S_{\text{th}}^{1/2} = \sqrt{\frac{4k_{\text{B}}T\kappa}{2\pi fQ}}, \quad (60.4)$$

where k_{B} is Boltzmann's constant, T is the temperature, f is frequency, and κ and Q are the torsion constant and quality factor of the torsion pendulum, respectively. Typical values of Q for a tungsten fiber are approximately 4000, while higher values have been reported for silicon fibers; however, a challenge associated with the use of silicon fibers is the necessity for the fiber to be conducting so that the pendulum does not accumulate an excess electrical charge [8].

Another noise source arises in the optical readout and the associated electronics. This effective torque noise increases with angular frequency, ω , as the torsion pendulum's resonant motion tends to hide driving signals with frequencies much greater than its resonant frequency [14]:

$$S_{\text{ro}}^{1/2} = S_{\theta}^{1/2} \kappa \sqrt{\left(1 - \frac{\omega^2}{\omega_0^2}\right)^2 + \left(\frac{1}{Q}\right)^2}. \quad (60.5)$$

In Equation 60.5, $S_{\theta}^{1/2}$ is the angular noise spectrum of the autocollimator, ω is angular frequency, and ω_0 is the pendulum's resonant angular frequency.

Reduction of statistical noise can be attained by improving autocollimators and lowering the running temperature. Attempts to perform torsion pendulum experiments at cryogenic temperatures are underway; however it remains to be seen whether improved noise levels can be obtained given the extra complication of running a low temperature apparatus [12, 23].

Other broadband sources of torque noise include fluctuating electric and magnetic fields, temperature variations, gravity gradient fluctuations, and changes in apparatus tilt. Typically these environmental sources are only of concern at the specific signal frequency and are therefore classified as sources of “false effects.”

60.5.2.2 False Effects False effects are any coherent effects that could mimic a violation of the ISL or WEP or otherwise contaminate the signal of interest for a given experiment. Due to the relative weakness of gravity, all environmental disturbances must be characterized and/or suppressed to ensure any observed signal is in fact due to a modification of gravity or other new physics. To do so, sources of false effects are generally greatly exaggerated and the response of the pendulum is measured to establish a “feedthrough” for the given source. The source is then measured in a running configuration and, with the known feedthrough, its effect on the pendulum during data acquisition can be established. If a coherent effect is resolved, an effort can be made to suppress it or otherwise it may be subtracted from the data in some cases. If there is no resolved effect, an upper limit can be placed using the measured feedthrough and noise level of the source. A discussion of specific sources of false effects follows in the following text.

Gravity gradient fluctuations due to large- or small-scale mass distribution variations can directly cause torque on the pendulum that could mimic a detection of new physics. Typically efforts are undertaken to vary the attractor mass configuration to be able to subtract any spurious gravitational effects. Reversal or rotation of attractor masses are routinely employed as well as measurement of the local gravitational field using specially designed gradiometer pendulums. Multipole techniques may be used to assess long-range gravity gradient effects on a torsion pendulum [24, 25].

While electromagnetic interactions can contribute to false effects, their contribution to noise is small due to the fact that torsion pendulums are typically made from non-magnetic materials and reside in gold-coated conducting enclosures. Large fluctuating magnetic fields are typically applied to observe the response of the pendulum and determine its magnetic moment. Measurement of ambient field fluctuations then yields a measure of any magnetically induced false effects. Reversal of the attractor masses also reverses the phase of, or may otherwise change, any magnetic false effects. Characterization of patch charge effects and other electrostatic phenomena is difficult. Efforts are made to ensure that the pendulum is electrically grounded to its surroundings and is contained in a perfectly conducting enclosure, including a conducting shield that prohibits any electrostatic influence of the attractor mass (Casimir or otherwise).

Temperature fluctuations are reduced by shrouding torsion pendulum apparatuses in a thermal enclosure. An additional thermal mass also surrounds the torsion fiber inside the vacuum chamber to provide further isolation. Intentional thermal variations are applied to observe the response of the pendulum. The observed level of temperature fluctuations during normal operation multiplied by any observed feedthrough will yield a measure of any temperature-related false effect. Temperature-related effects are notoriously difficult to quantify given the variety of mechanisms by which thermal expansion can affect the twist of the pendulum, so a large effort is typically devoted to keep the apparatus temperature as constant as possible (typical temperature stability at the level of milliKelvin is desired).

Apparatus tilt variations can cause twist of the pendulum due to the well-documented tilt–twist effect [14]. Precision tilt sensors are employed to measure both the purposeful exaggeration of the apparatus tilt (to measure the tilt–twist feedthrough) and the changes in tilt during running configuration.

Radiation pressure noise from the optical angle sensor is not typically a limiting factor in torsion pendulum experiments due to the placement of the beam on the torsion fiber axis and low optical power used by the autocollimator [9].

The ultimate goal of most torsion pendulum experiments is to run at thermal limit, at which the torque resolution is only limited by statistical sources and not by environmental disturbances and false effects. Although not easy, experiments do reach this limit [14, 15].

60.5.3 ISL Experiments

Torsion pendulums have provided the best tests of the ISL at short distances to date. A plot summarizing recent short-range results is shown in Figure 60.1. Deviations from the ISL have been excluded at the 95% confidence level from distances of $55\mu\text{m}$ to beyond planetary scales. The following summarizes the specific experiments shown in Figure 60.1.

60.5.3.1 University of Washington Eöt-Wash Experiment At short distances, the torsion pendulum used by the University of Washington’s Eöt-Wash group has provided the most stringent constraints on gravitational strength ($\alpha = 1$) effects [15]. The pendulum used a 21-fold azimuthally symmetric mass distribution that was hung over a similar attractor mass that rotated at angular frequency ω . As the attractor rotated, the oscillation of the pendulum’s twist angle was recorded, and the distance between the pendulum and attractor was changed. Analysis of the harmonic content of the pendulum’s twist at 21ω , 42ω , and 63ω was compared to a Newtonian prediction to extract limits in the α – λ parameter space. No deviation from Newtonian behavior was found. The group is currently improving the design to incorporate the 120-fold “wedge” symmetry shown in Figure 60.4. A second experiment under development and spearheaded by C. Hagedorn of this group utilizes planar pendulum and attractor mass geometry.

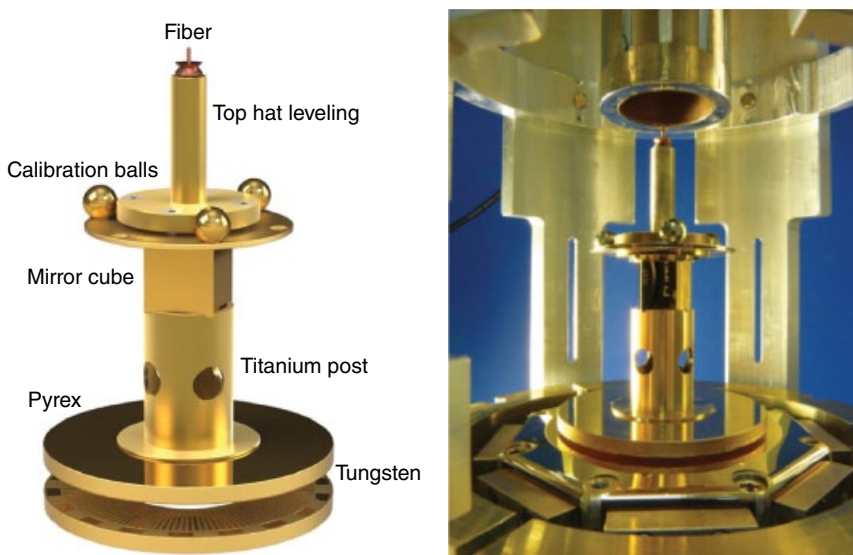


FIGURE 60.4 Left: diagram of next-generation Eöt-Wash torsion pendulum used for probing the ISL. The pendulum has a 120-fold symmetric “wedge” test mass design that will interact with a similarly shaped attractor mass that rotates beneath it. Right: picture of the pendulum suspended above the conducting membrane that electrostatically shields it from the attractor mass. Source: Reproduced with permission of Ted Cook.

60.5.3.2 Wuhan The Wuhan experiment employs a planar torsion pendulum design to achieve the best limits on the ISL in the 0.7–5 mm range [16]. The compensation masses and parallel-plate geometry used in this test provide an essentially null experiment.

60.5.3.3 Longer-Range Tests The UC Irvine/UW group has a well-established history of intermediate-range ISL tests [12]. The group is exploiting a torsion pendulum with cylindrical geometry that will have maximal sensitivity at $\lambda = 12$ cm. The specially designed pendulum and attractor mass will also yield an essentially null test.

60.5.4 Future ISL Tests

Several experimental efforts are underway to improve tests of the ISL at various distance scales. We describe here one approach in detail that is being developed by the author’s group at HSU.

60.5.4.1 HSU Experiment The HSU Gravitational Research Laboratory is developing a novel parallel-plate torsion pendulum and attractor configuration that proposes to obtain the best sensitivity down to $\lambda \approx 20 \mu\text{m}$ with a nearly perfect null experiment. The experiment is run primarily by undergraduates.

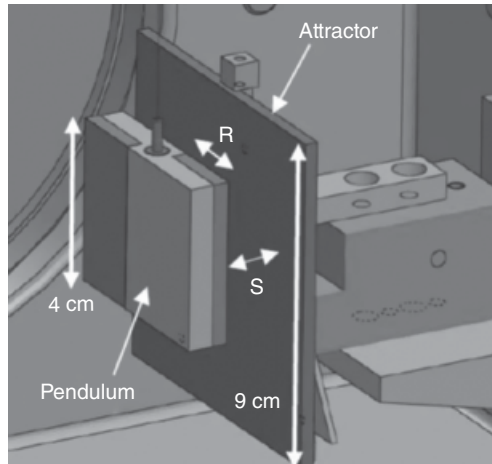


FIGURE 60.5 Basic geometry of the pendulum and attractor plate used in the Humboldt State experiment. When the pendulum/attractor separation, s , is modulated, the pendulum's stepped design results in potentially different short-range torques applied to each side, which in turn would result in twist motion at harmonics of the chosen attractor drive frequency. The pendulum is made from aluminum (light) and titanium (dark), and the attractor is copper. A conducting membrane that will separate the pendulum and attractor is not shown. The use of two different materials produces a composition dipole that provides sensitivity to violations of the WEP as well as the ISL.

The nearly null experiment is achieved by noting that the gravitational force does not depend on distance for a test mass interacting with an infinite plane of matter. The proposed tests exploit this fact by using a parallel-plate configuration with planar pendulums and a comparatively large attractor plate; a simplified geometry is shown in Figure 60.5, while Figure 60.6 shows an overview of the laboratory layout.

The proposed design is an aluminum/titanium pendulum with rectangular “steps” and the dimensions shown in Figure 60.5. The width of each step is $R = 2$ cm, while the thickness of each Al step is 6.25 and 3.75 mm for the Ti steps, ensuring equal mass in all steps. The total mass of the pendulum is 58 g. The attractor is a $9\text{ cm} \times 9\text{ cm} \times 0.3\text{ cm}$ -thick rectangular copper plate. The total mass of the plate is approximately 220 g.

The attractor–pendulum separation, s , is modulated by moving the attractor sinusoidally at angular drive frequency ω . In the ideal case of an infinite attractor plate, the Newtonian torque on the pendulum does not vary with attractor position, while any short-range interaction produces more torque on the closer, high-density “step” when s is smaller than the range of the interaction. This potential short-range torque modulation would cause a variation of the pendulum's twist at harmonics of the attractor modulation frequency (1ω , 2ω , 3ω , etc.). Note that Newtonian signal is only present due to the finite size of any *real* attractor mass and predominantly occurs at 1ω .

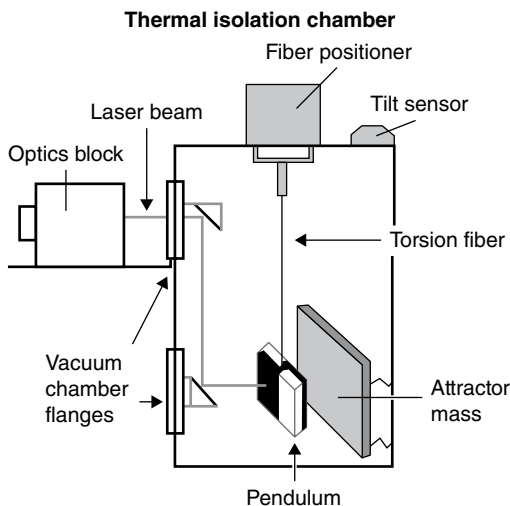


FIGURE 60.6 Overview of the existing laboratory setup at Humboldt State University.

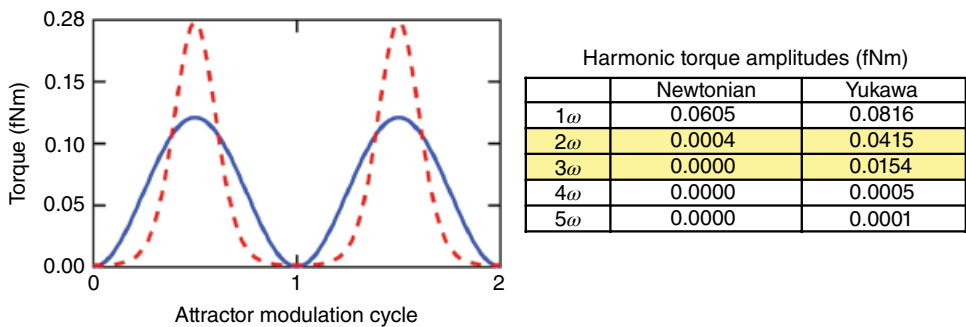


FIGURE 60.7 Left: calculated Newtonian and possible Yukawa torques on the Humboldt State pendulum as a function of time for two complete attractor modulation cycles. The peak-to-peak distance modulation amplitude is 0.5 mm and the minimum separation is 100 μm . Any Yukawa torque (dashed curve) with $\alpha=1$ and $\lambda=100\mu\text{m}$ would be clearly evident and larger than the Newtonian background (solid curve). Right: table of harmonic torque amplitudes for the times series shown on the left. Notice that for the chosen parameters, the 2ω and 3ω Yukawa signals are clearly different from the tiny Newtonian torque amplitudes. This difference in frequency dependence can be used to place constraints on Yukawa parameters, while systematic effects that will be largest at 1ω can be largely avoided.

Systematic false effects are also generally largest at 1ω . Thus the higher harmonics provide means to distinguish Newtonian and systematic false effects from short-range interactions. In fact, for an interaction with very small λ , the higher harmonics are similar in magnitude to one another because the applied torque is essentially a delta function at the closest pendulum/attractor separation (although the value of the torque diminishes rapidly as λ decreases). The harmonic content of a hypothetical interaction is shown in Figure 60.7.

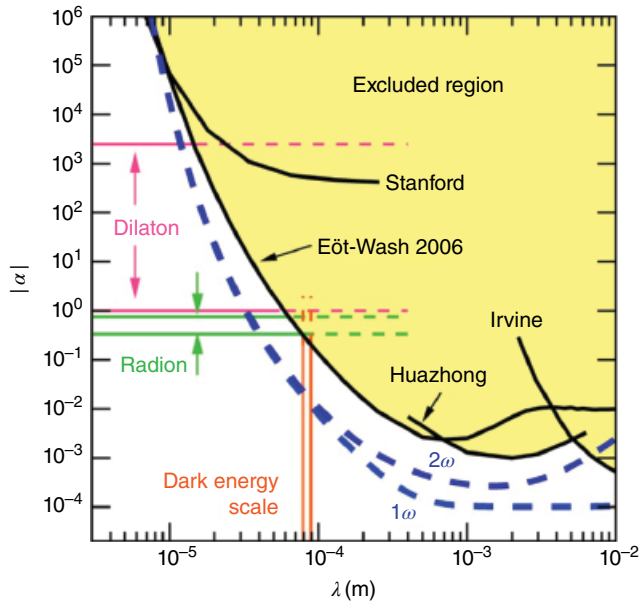


FIGURE 60.8 Reprint of Figure 60.1 including the predicted sensitivity for the HSU experiment. Each dashed line shows the predicted sensitivity of this apparatus for analysis of a single harmonic torque amplitude. Improved constraints may be obtained by analyzing multiple harmonics together. Note that for some values of λ , an improvement by a factor of approximately 50 is obtained over previous efforts.

The digitized angle signal is processed via Fourier techniques to determine the twist amplitude of the pendulum at harmonics of the attractor drive frequency. The amplitude of the twist oscillations will be compared to a detailed model of the expected pendulum/attractor Newtonian and possible Yukawa twist amplitudes. Limits in the α - λ parameter space will be obtained from this comparison, and any deviation from expected ISL (or WEP) values may be an indication of new physics. A prediction of the sensitivity for this design is presented in Figure 60.8

The surface of the pendulum will be polished to an optical finish and gold coated. The polish is necessary to reflect the autocollimator beam from the back side of the pendulum (side facing away from the attractor mass).

The two different materials naturally form a “composition dipole” and provide sensitivity to test the WEP. The materials will be joined first with an adhesive and subsequently machined and lapped to the appropriate dimensions. Fabrication of pendulums with different composition dipole pairs such as copper/titanium or molybdenum/aluminum will ensure that the WEP test is sensitive to all values of $\tilde{\psi}$.

60.5.4.2 Other Tests As previously noted, the Eöt-Wash and UC Irvine/UW group are developing several new ISL tests at short distances. Other research efforts utilizing torsion pendulums are also under way. Clive Speake’s group at the University of

Birmingham is implementing a superconducting torsion pendulum for gravitational tests [26]. The Cowsik group at Washington University is developing a pendulum supported by a torsion ribbon that uses a novel autocollimator [12].

60.5.5 WEP Tests

The best limits on violations of the WEP spanning the distance scale from 1 cm to ∞ have been obtained by composition dipole torsion pendulums in both rotating and nonrotating experimental configurations.

60.5.5.1 Eöt-Wash Rotating Torsion Balance Tests The best limits on a violation of the WEP over distance scales from roughly 3 m to ∞ (curve labeled EW in Fig. 60.2) have been determined with a rotating torsion balance used by the University of Washington Eöt-Wash group [7]. The rotating torsion balance used for WEP tests employs a composition dipole pendulum in a vacuum chamber that is rotated uniformly with frequency on the order of millihertz. Any differential acceleration resulting from the composition dipole would produce a twist of the pendulum at the rotation frequency. Thus, the experiment is sensitive to sources of WEP violations at any length scale greater than approximately the distance from the pendulum to the nearest stationary source mass.

The most recent results from this group were obtained using a torsion pendulum with two different composition dipole configurations: Be-Ti and Be-Al. A picture of this pendulum with its gold-coated test bodies is shown in Figure 60.9.

60.5.5.2 Eöt-Wash Short-Range WEP Test A separate experiment was conducted by the same University of Washington group that instead used a stationary Cu-Pb torsion pendulum interacting with a 3-ton rotating attractor mass made of ^{238}U [9]. The result from this experiment is labeled as EW 99 in Figure 60.2. The neutron-rich attractor allowed expanded exploration of the WEP parameter space; in particular, sensitivity was greatly enhanced for violations of the WEP that arise from neutron excess ($B - L$). The limits obtained represent the best from roughly 1 cm to 3 m. Constraints on violations of the WEP at shorter distances are inferred from ISL experiments. However, the previously mentioned experiment under development by the author's group at HSU is specifically designed to yield measurements of the WEP at subcentimeter distance scales with the possibility to probe diverse composition dipole mass pairings.

60.5.6 Measurements of G

Torsion pendulums have been widely used for measurement of the gravitational constant (see Ref. 10). The measurement of G with the smallest uncertainty was obtained by a torsion pendulum used in angular acceleration feedback mode [21]. Several torsion balance experiments are ongoing; however, atom interferometry and other new and promising techniques are on the horizon.

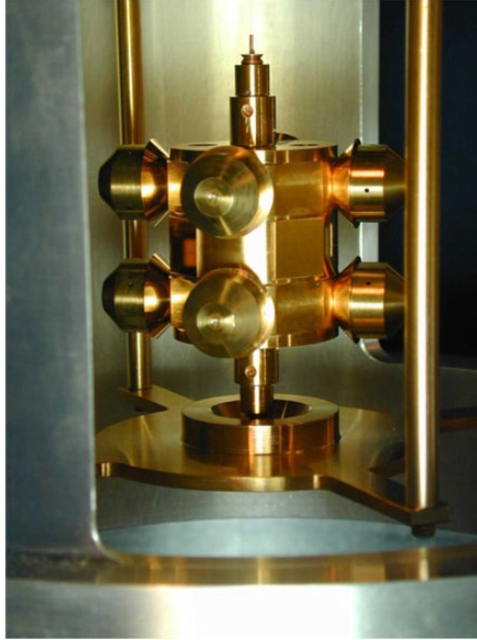


FIGURE 60.9 Eöt-Wash torsion pendulum used in the rotating torsion balance WEP tests of reference [7]. Source: Wagner et al. [7]. Reproduced with permission of IOP.

60.6 MICROOSCILLATORS AND SUBMICRON TESTS OF GRAVITY

For the $10\mu\text{m}$ scale and below, the practicality of suppressing electrostatic backgrounds in torsion pendulum experiments becomes insurmountable. At these scales, high-frequency oscillator techniques have been used to place constraints on the ISL; however, no experiment has yet achieved the sensitivity to measure gravitational strength effects.

60.6.1 Microcantilevers

Josh Long at Indiana University is continuing an experimental effort started by the John Price group at the University of Colorado that uses a resonant microcantilever system to search for violations of the ISL [27].

The Kapitlnik group at Stanford University has performed an extended series of measurements also employing a microcantilever system [18, 28].

60.6.2 Very Short-Range ISL Tests

To measure violations of the ISL at very short ranges ($<10\mu\text{m}$), experimenters must compete directly with the Casimir force. Most tests in this regime perform direct comparison of the measured force to the Casimir prediction to derive limits on

deviations from the ISL. Several recent Casimir measurements have yielded limits on Yukawa-type deviations from the ISL at short distances [29, 30].

60.7 ATOMIC AND NUCLEAR PHYSICS TECHNIQUES

The most recent measurement of the gravitational constant was performed using cold atom interferometry [10]. This novel technique along with other proposed atomic, nuclear, and molecular methods [13, 31–35] have the possibility to open up a wide range of new gravitational measurements in the near future.

ACKNOWLEDGEMENTS

The author's work and the HSU Gravitational Research Laboratory are supported by the National Science Foundation under grants PHY-1065697 and PHY-1306783. The author would like to thank Todd Wagner and Ted Cook for providing figures and images.

REFERENCES

1. E.G. Adelberger, B.R. Heckel, and A.E. Nelson, *Ann. Rev. Nucl. Part. Sci.* 53, 77 (2003).
2. N. Arkani-Hamed, S. Dimopoulos, and G.R. Dvali, *Phys. Lett. B* 436, 257 (1998).
3. G. Dvali, G. Gabadadze, M. Kolanovic, and F. Nitti, *Phys. Rev. D* 65, 024031 (2001).
4. R. Sundrum, *Phys. Rev. D* 69, 044014 (2004).
5. S.R. Beane, *Gen. Relativ. Gravit.* 29, 945 (1997).
6. J. Khoury and A. Weltman, *Phys. Rev. Lett.* 93, 171104 (2004).
7. T.A. Wagner, S. Schlamminger, J.H. Gundlach, and E.G. Adelberger, *Class. Quantum Gravit.* 29, 184002 (2012).
8. E.G. Adelberger, J.H. Gundlach, B.R. Heckel, S. Hoedl, and S. Schlamminger, *Prog. Part. Nucl. Phys.* 62, 102 (2009).
9. G.L. Smith, C.D. Hoyle, J.H. Gundlach, E.G. Adelberger, B.R. Heckel, and H.E. Swanson, *Phys. Rev. D* 61, 22001 (2000).
10. G. Rosi, F. Sorrentino, L. Cacciapuoti, M. Prevedelli, and G.M. Tino, *Nature* 510, 518–521 (2014).
11. P.J. Mohr, B.N. Taylor, and D.B. Newell, *Rev. Mod. Phys.* 84, 1527–1605 (2012).
12. R.D. Newman, E.C. Berg, and P.E. Boynton, *Space Sci. Rev.* 148, 175–190 (2009).
13. A.A. Geraci, S.B. Papp, and J. Kitching, *Phys. Rev. Lett.* 105, 101101 (2010).
14. C.D. Hoyle, D.J. Kapner, B.R. Heckel, E.G. Adelberger, J.H. Gundlach, U. Schmidt, and H.E. Swanson, *Phys. Rev. D* 70, 042004 (2004).

15. D.J. Kapner, T.S. Cook, E.G. Adelberger, J.H. Gundlach, B.R. Heckel, C.D. Hoyle, and H.E. Swanson, *Phys. Rev. Lett.* 98, 021101 (2007).
16. S.-Q. Yang, B.-F. Zhan, Q.-L. Wang, C.-G. Shao, L.-C. Tu, W.-H. Tan, and J. Luo, *Phys. Rev. Lett.* 108, 081101 (2012).
17. J.K. Hoskins, R.D. Newman, R. Spero, and J. Schultz, *Phys. Rev. D* 32, 3084 (1985).
18. A.A. Geraci, S.J. Smullin, D.M. Weld, J. Chiaverini, and A. Kapitulnik, *Phys. Rev. D* 78, 022002 (2008).
19. H. Cavendish, *Philos. Trans. R. Soc. Lond.* 88, 469–526 (1798).
20. C. Speake and T. Quinn, *Phys. Today* 67, 7, 27 (2014).
21. J.H. Gundlach and S. Merkowitz, *Phys. Rev. Lett.* 85, 2869 (2000).
22. M.D. Turner, C.A. Hagedorn, S. Schlamminger, and J.H. Gundlach, *Opt. Lett.* 36, 1479–1481 (2011).
23. F. Fleischer, E.G. Adelberger, M. Bassan, and B. Heckel, American Physical Society April Meeting/AAPT Meeting, February 13–17, Washington, DC (2010).
24. E.G. Adelberger, N.A. Collins, and C.D. Hoyle, *Class. Quantum Gravit.* 23, 125 (2006).
25. C. D’Urso and E.G. Adelberger, *Phys. Rev. D* 55, 7970 (1997).
26. G.D. Hammond, C.C. Speake, A.J. Matthews, E. Rocco, and F. Peña-Arellano, *Rev. Sci. Instrum.* 79, 025103 (2008).
27. J.C. Long, H.W. Chan, A.B. Churnside, E.A. Gulbis, M.C.M. Varney, and J.C. Price, *Nature* 421, 922–925 (2003).
28. S.J. Smullin, A.A. Geraci, D.M. Weld, J. Chiaverini, S. Holmes, and A. Kapitulnik, *Phys. Rev. D* 72, 122001 (2005).
29. V.M. Mostepanenko, R.S. Decca, E. Fischbach, G.L. Klimchitskaya, D.E. Krause, and D. López, *J. Phys. A: Math. Theor.* 41, 164054 (2008).
30. R.S. Decca, D. López, H.B. Chan, E. Fischbach, D.E. Krause, and C.R. Jamell, *Phys. Rev. Lett.* 94, 240401 (2005).
31. F. Sorrentino, M. de Angelis, A. Bertoldi, L. Cacciapuoti, A. Giorgini, M. Prevedelli, G. Rosi, and G.M. Tino, *J. Eur. Opt. Soc.* 4, 1990 (2009).
32. A. Peters, K.Y. Chung, and S. Chu, *Metrologia* 38, 25 (2001).
33. T. Jenke, G. Cronenberg, J. Burgdörfer, L.A. Chizhova, P. Geltenbort, A.N. Ivanov, T. Lauer, T. Lins, S. Rotter, H. Saul, U. Schmidt, and H. Abele, *Phys. Rev. Lett.* 112, 151105 (2014).
34. V.V. Nesvizhevsky, V.V. Nesvizhevsky, H.G. Börner, A.K. Petukhov, H. Abele, S. Baessler, F.J. Ruess, T. Stoferle, A. Westphal, A.M. Gagarski, G.A. Petrov, and A.V. Strelkov, *Nature* 415, 297–299 (2002).
35. S. Dimopoulos, P.W. Graham, J.M. Hogan, and M.A. Kasevich, *Phys. Rev. Lett.* 98, 111102 (2007).

61

CRYOGENIC MEASUREMENTS

RAY RADEBAUGH

*Applied Chemicals and Materials Division, National Institute of Standards and Technology¹,
Boulder, CO, USA*

61.1 INTRODUCTION

Cryogenics usually refers to temperatures less than about 120 K, but in this chapter we use such a definition rather loosely. Typically there is a gradual change in the type of sensors used or measurement methodology as temperature is lowered rather than an abrupt change as the temperature is lowered below 120 K. In some cases, a sensor appropriate for 100 K may also be the best for a temperature as high as 300 K. In this chapter we will specify the temperature range appropriate for a particular sensor or measurement methodology. Instrumentation for processes or experiments involving cryogenic temperatures often requires the use of sensors that must operate at these low temperatures. Certainly the measurement of temperature can only be done with a thermometer at the temperature of interest. However, certain other parameters, such as pressure, flow, liquid level, and magnetic field, have often been made with the active sensor located at room temperature but which could infer the property at cryogenic temperatures. This procedure usually resulted in a loss of accuracy, particularly under dynamic conditions. When cryogenic sensors were required in the early years of cryotechnology, they were usually constructed in the laboratory. The demand for cryogenic

¹Contribution of the United States Government; not subject to copyright in the United States.

sensors has grown sufficiently that commercial sensors are often available for use at cryogenic temperatures. In some cases, simple modifications of commercial sensors suffice to make them adaptable for use at cryogenic temperatures. In this chapter we review the availability and properties of commercial sensors and discuss the necessary modifications to make them useful for cryogenic temperatures.

The temperature range of primary interest here is between 4 and 300 K. Except for the case of some thermometers, a sensor that functions at 4 K will usually continue to function at lower temperatures as long as the power input is not too great. One of the main reasons commercially available sensors or transducers cannot be used at cryogenic temperatures is because of the choice of materials. In some cases, a material (e.g., rubber) undergoes a brittle transition at some low temperature that prevents its use at cryogenic temperatures. In other cases the differential contraction of different materials may be great enough at cryogenic temperatures to cause too high stresses or interference with moving components. Sensors with moving parts (such as flow sensors) are particularly difficult to operate at cryogenic temperatures because of the need for dry lubrication. Often electrical power inputs that are satisfactory for operation at room temperature can cause a sensor to self-heat or interfere with the overall experiment at cryogenic temperatures. Thus, commercial sensors can often be adapted for use at cryogenic temperatures by reducing the power input and/or changing a few key materials. Calibration of the sensor at the temperature it is to be used is nearly always necessary. Such calibrations often involve a comparison with a standard at room temperature.

61.2 TEMPERATURE

61.2.1 ITS-90 Temperature Scale and Primary Standards

The temperature scale in use today is known as the International Temperature Scale of 1990 (ITS-90), which extends from 0.65 to 1358 K [1]. It is a very close approximation to a true thermodynamic temperature scale. The scale is established by use of physical phenomena known so well that temperature can be calculated without any unknown quantities. Examples include equation of state of a gas, the velocity of sound in a gas, the thermal voltage or current noise in a resistor, blackbody radiation, and the angular anisotropy of gamma-ray emission from some radioactive nuclei in a magnetic field. A provisional extension covering the range from 0.9 mK to 1 K was established in 2000 and is known as PLTS-2000. It uses the melting curve of ^3He as the defining scale [2]. The ITS-90 is defined through a set of fixed points, interpolating primary thermometers, and interpolating equations [1, 3–5]. Fixed points are triple points and superconducting transition points. A standard platinum resistance thermometer (SPRT) is an example of an interpolating primary thermometer for temperatures between the triple point of equilibrium hydrogen at 13.8033 K and the freezing point of silver at 961.78 K. Use of fixed points and primary thermometers is a complex and

expensive undertaking, which limits their use mostly to national standards institutions. The primary standards are transferred to secondary standards, such as high-purity platinum or Rh—Fe alloy resistance thermometers. The secondary standards are then used to calibrate commercial (industrial) thermometers for customer use.

61.2.2 Commercial Thermometers

Thermometry at cryogenic temperatures has become very well developed with a wide range of commercially available thermometers for the measurement of temperatures from the millikelvin temperatures up to room temperature and above. Excellent reviews of cryogenic thermometry have been published [6–12]. A new class of thermometers not discussed in these previous reviews (except Ref. 7) is the zirconium oxynitride ceramic film thermometers, which have low magnetoresistive effects. Most commercial thermometers for cryogenic temperatures are resistors, diodes, thermocouples, or capacitors. The change of their electrical characteristic with temperature determines their suitability as a thermometer. A good thermometer should have high sensitivity and be stable over time. For dynamic measurements it should also have a fast response time.

61.2.2.1 Metallic Resistance Thermometers The resistance of most pure metals varies roughly linearly with temperature until at some low temperature the scattering of electrons by impurities dominates the resistance, which leads to a lower limit to the resistance. In most cases, this limit is reached by 4 K, which means it can no longer function as a thermometer. For standard-grade platinum the ratio of resistance at 4.2 K to that at 273.16 K is usually less than 4×10^{-4} . More impurities cause this ratio to increase and the lower limit to be reached at higher temperatures. Platinum is the most widely used metallic resistance thermometer because it is so reproducible over long periods of time. It is often used as a secondary standard to calibrate other commercial thermometers. The platinum wire used in standard thermometers is of very high purity and care is taken in the construction of the thermometer to eliminate strains in the wire, which can affect the resistance. Figure 61.1 shows the typical construction of a capsule type of SPRT. The capsule type is most commonly used for cryogenics as opposed to the long-stem type in which heat conduction in the stem from ambient temperature to low temperature can cause unacceptable heat leaks. A helix of platinum wire (about

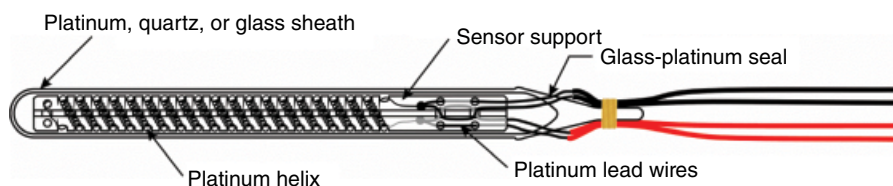


FIGURE 61.1 Cross section of a standard platinum resistance thermometer (SPRT).

75 μm diameter) is bifilarly wound on a notched mica or ceramic cross and placed inside the sheath, after which it is annealed at about 600°C to remove all strains. The capsule is filled with helium gas to enhance heat transfer between the platinum element and the sheath. Platinum, Inconel, glass, and quartz are common sheath materials. Dimensions of the capsule including the platinum-glass electrical feedthrough are about 5.8 mm in diameter by 56 mm long, with a resistance of 25.5 Ω at 0°C. The ITS-90 temperature range for these thermometers is from 13.8 K up to about 250°C. Special high-temperature versions are used for temperatures up to the silver triple point at 961.93°C. The reproducibility of these standard-grade thermometers is about 1 mK. Miniature capsule-type SPRTs have been developed recently that are about 3.2 mm in diameter and 9.7 mm long that can be used down to 13.8 K with some sacrifice in reproducibility [13].

Figure 61.2 shows how resistance varies with temperature for the most common metallic resistance thermometers. These thermometers have a positive temperature coefficient of resistance. A capacitance thermometer characteristic is also shown in Figure 61.1, but it will be discussed later. Two curves are shown for platinum thermometers, one for laboratory-grade SPRTs and one for PRTs made with lower-purity platinum, referred to as an industrial-grade platinum resistance thermometer (IPRT or just PRT). The resistance of the SPRT follows the ITS-90 definition down to 13.8 K, whereas the PRT meets the ITS-90 definition only down to about 70 K through the use of a slightly different resistance versus temperature curve. Their reproducibility is about 5 mK or higher. The distinction between the two grades is based on the purity of

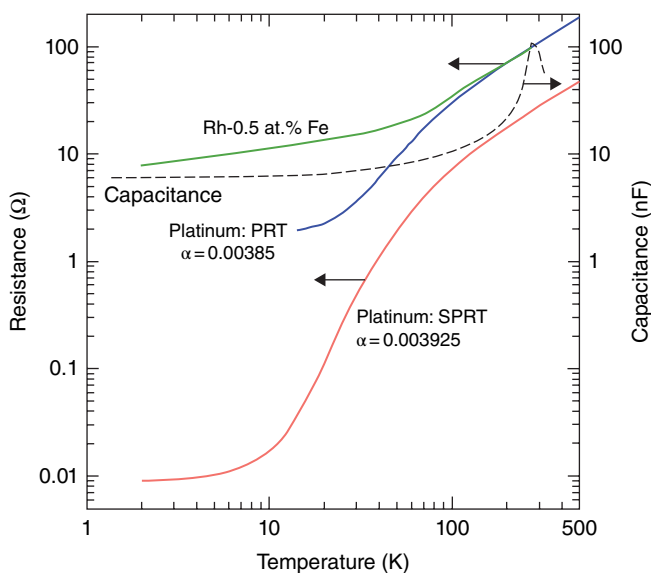


FIGURE 61.2 Characteristics of metallic resistance thermometers and capacitance thermometers.

the platinum and how the active element is supported to eliminate strain. The distinction is quantified by use of the resistance ratio given by

$$W(T) = \frac{R(T)}{R(273.16\text{K})}. \quad (61.1)$$

To be suitable for a SPRT, the platinum must be of sufficient purity to satisfy at least one of the two following relations [1]:

$$W(302.9146\text{K}) \geq 1.11807 \quad (61.2a)$$

$$W(234.3156\text{K}) \leq 0.844235. \quad (61.2b)$$

The defining equations that relate temperature to the ratio W are given by [1]. The constants in the equations are determined from calibrations. An alternate specification of platinum purity used in thermometers is the temperature coefficient of resistance α defined by

$$\alpha = \frac{R(100^\circ\text{C}) - R(0^\circ\text{C})}{100 R(0^\circ\text{C})}, \quad (61.3)$$

which technically has units of ohms/(ohm·°C) because the 100 in the denominator represents 100°C temperature change, but conventionally α is considered dimensionless. To meet the ITS-90 conditions given by Equations 61.2a and 61.2b, the temperature coefficient α should satisfy

$$\alpha \geq 0.003925. \quad (61.4)$$

This high value of alpha is achieved only with expensive reference-grade platinum (99.999% purity) wound in a strain-free manner and used in laboratory-grade SPRTs. The resistance at 0°C is normally 25.5 Ω. With reference-grade platinum in industrial thermometers, the temperature coefficient is 0.003920. Lower-purity platinum is used in most industrial-grade thermometers. Different standards organizations have adopted different temperature coefficients as their standard. The most widely used standard is the European standard (also widely used in the United States and elsewhere), designated by DIN IEC 60751 and ASTM E-1137 in which $\alpha=0.0038500$ and the resistance at 0°C is 100 Ω. The resistance of industrial-grade PRTs follows a standard curve down to about 70 K. They are usable for lower temperatures but require an individual calibration. There are three tolerance grades (A, B, and C) for the DIN standard and two for the ASTM standard. For the ASTM standard the grade A has a tolerance ranging from ±0.47 K at 73 K to ±0.13 K at 273 K, whereas the grade B tolerance is ±1.1 K at 73 K and ±0.25 K at 273 K. The tolerance indicates the level of interchangeability for the thermometer.

Resistance thermometers made with Rh-5 at.% Fe have a resistance that continues to change even below 1 K. They are useful for the temperature range of 0.65–500 K, with a linear response above 100 K, as shown in Figure 61.2. RhFe thermometers are not interchangeable like that of Pt thermometers, but their reproducibility of about 0.2 mK in models fabricated like that of the SPRT makes them a candidate for an interpolating standard below 25 K for the ITS-90 temperature scale [14]. Other metals and alloys are sometimes used in resistance thermometers for special reasons. Resistance thermometers that use pure metals are not very sensitive for temperatures of 4 K and below. For these lower temperatures, thermometers made with semiconductors or other negative temperature coefficient materials become a better choice.

61.2.2.2 Semiconductor-Like Resistance Thermometers Figure 61.3 shows the resistance response curves for several types of semiconductor-like resistance thermometers. As the figure indicates, their resistance is very sensitive to temperature below about 100 K, unlike that of platinum thermometers. The response curves for these semiconductor-like thermometers have a negative temperature coefficient. The disadvantage with these thermometers is that except for the RuO₂ thermometers, they do not follow a standard response curve as do platinum thermometers, so they are not interchangeable and must be individually calibrated. The RuO₂ thermometers are only interchangeable for the same manufacturer. The zirconium oxynitride thermometers, sold under the trade name as Cernox thermometers, are a commonly used type and are

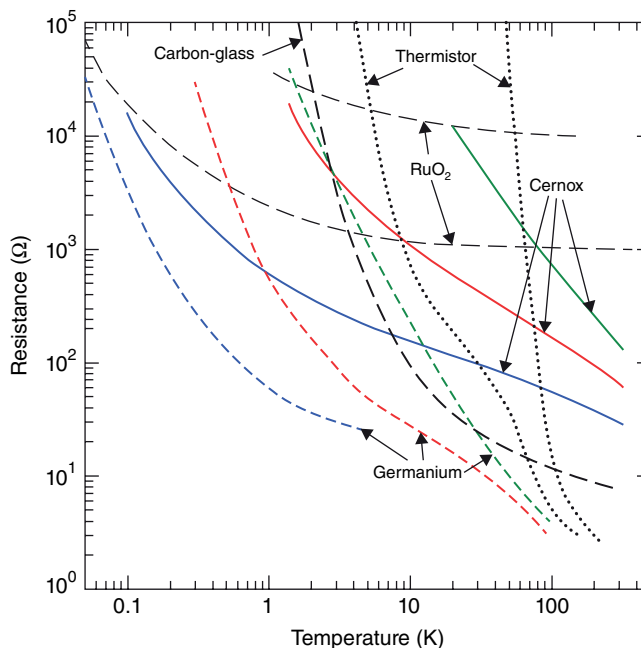


FIGURE 61.3 Characteristics of semiconductor and semiconductor-like thermometers.

available in several versions to cover a wide temperature range. They are a thin film resistor deposited on an alumina substrate. Their reproducibility of 3 mK at 4 K and 15 mK at 77 K allows them to be used in most laboratory settings. Germanium resistance thermometers are generally used for precision measurements below about 80 K. Their reproducibility is about ± 0.5 mK at 4 K, but they must be individually calibrated. They can yield accuracies of about 5–15 mK for temperatures between about 1 and 20 K with commercial calibrations. Selected germanium thermometers are useful as thermometers down to about 50 mK and are available with commercial calibrations down to that temperature. Carbon-glass thermometers have a very steep response curve, which gives them high sensitivity. They are made by impregnating porous glass with carbon [15]. They are not interchangeable, but their reproducibility is quite good. Their high sensitivity makes them useful for temperature control. Carbon resistors in the form of electrical circuit resistors have been used often for very low-cost thermometers, but only particular brands have been found to be useful. A useful brand may be discontinued or experience a composition change that greatly affects its low-temperature resistance behavior. Thermistors usually have the steepest response curves of all thermometers, which limits the temperature range for an individual thermistor. Their use is then limited to special applications, such as in precise temperature control.

61.2.2.3 Diode Thermometers The forward voltage of diodes with constant current excitation varies with temperature and makes a good thermometer. Figure 61.4 shows typical voltage curves for Si and GaAlAs diodes with a $10\mu\text{A}$ current excitation.

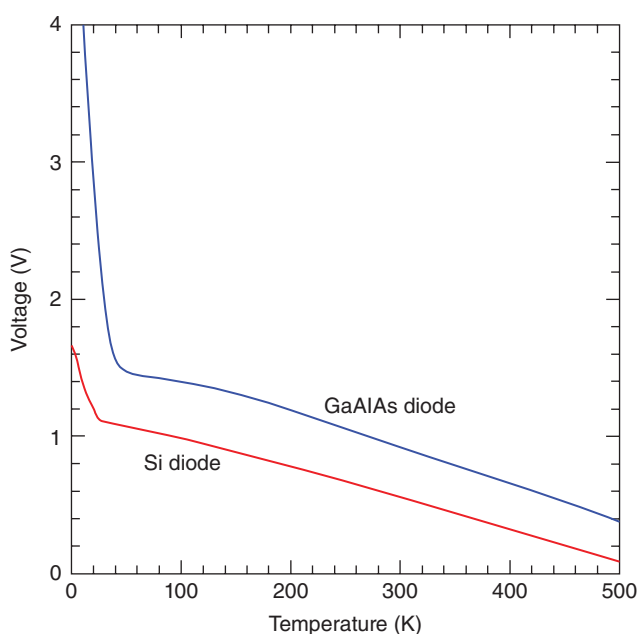


FIGURE 61.4 Characteristics of two types of diode thermometers with $10\mu\text{A}$ current.

Special Si diodes make excellent thermometers because of their interchangeability for the same manufacturer and their relatively large voltage signal. They follow a standard curve to within 0.25 K for the “A” grade and 0.5 K for the “B” grade. Magnetic fields have a strong effect on Si diodes but much less effect on GaAlAs diodes. Unfortunately, GaAlAs diodes do not follow a standard curve, so they must be individually calibrated.

61.2.2.4 Thermocouples Thermocouples make use of the thermopower or the Seebeck coefficient in metals. If an electrical conductor is placed in a temperature gradient, current carriers (electrons or holes) will diffuse from the hot to cold end to build up a voltage that prevents further diffusion. The voltage gradient with zero current flow is known as the absolute Seebeck coefficient, which is given by

$$S = -\frac{dV}{dT}, \quad (61.5)$$

where S has units of V/K. Note that it is not possible to measure S directly, since any attempt to measure the voltage with a voltmeter will introduce another conductor in the temperature gradient with its own Seebeck coefficient. What is measured is the relative Seebeck coefficient, which is the difference between the absolute values in the two conductors. The measured Seebeck coefficient of the conductor pair is given as

$$S_{AB} = S_A - S_B = \frac{dV_B}{dT} - \frac{dV_A}{dT} = -\frac{dV_{AB}}{dT}. \quad (61.6)$$

The sign convention is quite complicated and will not be discussed here. Because the entropy of charge carriers in a superconductor is zero, S of a superconductor is zero. Thus, by choosing one leg of a pair to be a superconductor, S of the other leg can be determined from the measurement of the pair. Such a technique is useful only up to about 120 K, above which no superconductors exist. For higher temperatures $S(T)$ is found from the difficult measurements of the Thomson coefficient μ and use of the relation

$$S(T) = \int_0^T \frac{\mu(T')}{T'} dT'. \quad (61.7)$$

Figure 61.5 shows the temperature dependence for the absolute Seebeck coefficient for materials commonly used in thermocouples [16]. For a finite temperature difference, the voltage across the conductor is the integral of S over the given temperature range. Though the absolute Seebeck coefficient has no practical use, it aids in understanding voltages developed in thermocouple measurement systems. It shows that voltages are developed along the length of conductors in temperature gradients and not at the junctions. The junction simply ensures that the electrical potential of both legs is the

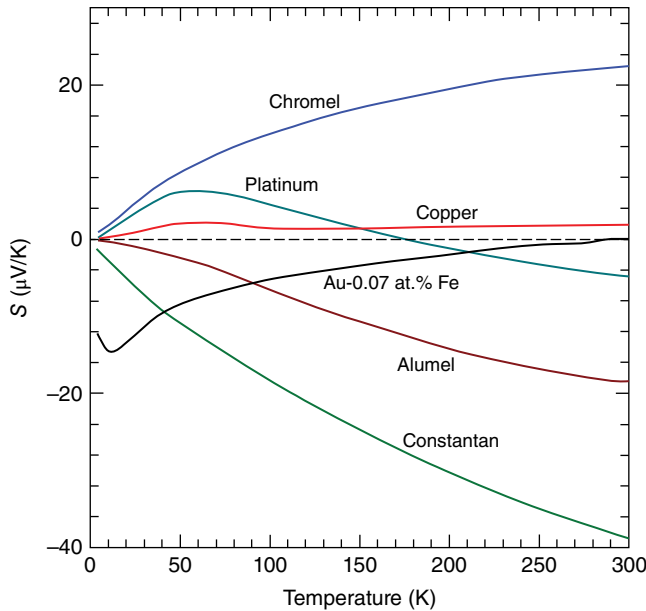


FIGURE 61.5 Absolute Seebeck coefficient of several metals used in thermocouples.

TABLE 61.1 Thermocouples Useful for Cryogenic Temperatures

Type	Positive Wire	Negative Wire	US Color Code	T Range (K)	Std. Error
E	Chromel	Constantan	+ purple; - red	3–1173	1.5 K
J	Iron	Constantan	+ white; - red	63–1073	1.5 K
K	Chromel	Alumel	+ yellow; - red	3–1573	1.5 K
T	Copper	Constantan	+ blue; - red	3–673	1% of T
Au–Fe	Chromel	Au-0.07 at.% Fe	—	1–573	0.2% of V

same at that point. For example, if both legs of a pair are of the identical material and the open ends are at the same temperature, then according to Equation 61.6 there will be no measured voltage difference between the pair. Also, if a third material is introduced in an isothermal part of the thermocouple circuit, then it has no effect on the output because no additional voltage is introduced in the isothermal section. Lastly, Figure 61.5 is useful in understanding the polarity of each leg in a thermocouple.

Several metal combinations are used for thermocouple thermometry, depending on the temperature and other environmental conditions. The most common combinations for use at cryogenic temperatures are given in Table 61.1. Designations for single-leg materials are:

Constantan: EN or TN, nominally 55 wt.% Cu and 45 wt.% Ni

Chromel (Trademark of Concept Alloys, Inc.): EP or KP, nominally 90 wt.% Ni and 10 wt.% Cr

Alumel (Trademark of Concept Alloys, Inc.): KN, nominally 95 wt.% Ni, 2 wt.% Al, 2 wt.% Mn, and 1 wt.% Si

Figure 61.6 shows the Seebeck coefficient S_{AB} (sensitivity) of common cryogenic thermocouple pairs. The integral of S_{AB} gives the emf of the pair with the junction held at 0 K, as shown in Figure 61.7. In practice, the reference junction is usually held at 0°C, so thermocouple tables are usually given with respect to a 0°C reference. The curves in Figure 61.7 with a 0 K reference are converted to a 0°C reference simply by subtracting the voltage at 0°C from the curves. Figure 61.8a shows a schematic of the theoretical wiring arrangement with the reference junction at 0 K and the voltmeter at a temperature T . Figure 61.8b shows the wiring arrangement for the typical scheme of a 0°C reference temperature.

Thermocouples are inherently a sensor for measuring temperature differences, so they are often used to measure small temperature differences. Figure 61.9 shows a schematic for such measurements. In this arrangement most of the temperature difference from some low temperature to ambient is spanned by identical materials on both legs, so no additional voltage is developed over most of the temperature gradient, and both legs can come from the same spool to ensure nearly identical Seebeck coefficients. The reference junctions can be at any temperature, but they must be at the same temperature. Often they are anchored at the low temperature T by potting both of them in a copper piece. With such an arrangement the leads extending to room temperature can be copper or phosphor bronze, which have low values of absolute Seebeck coefficients $S(T)$, so any material variation between the legs has only a minimal spurious

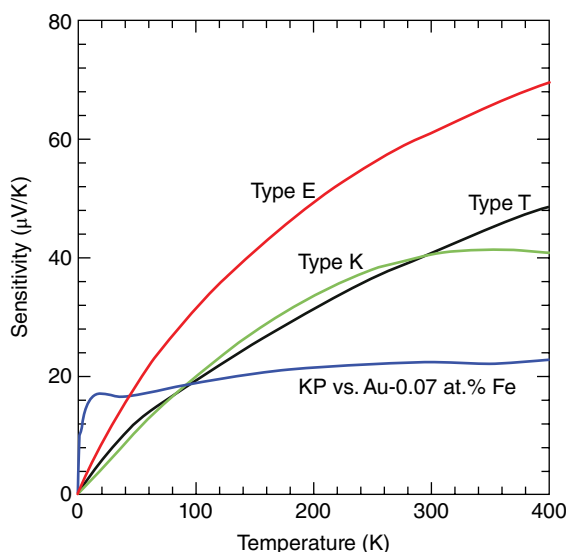


FIGURE 61.6 Relative Seebeck coefficient or sensitivity of common thermocouple types useful for cryogenic temperatures.

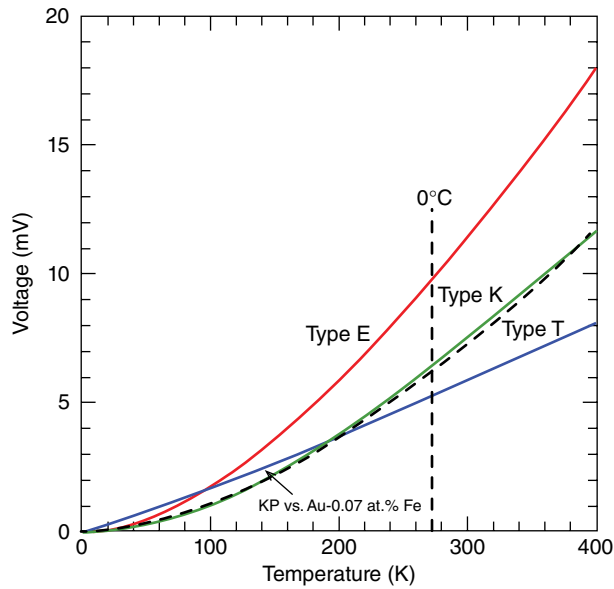


FIGURE 61.7 Voltage output of common cryogenic thermocouples with a 0 K reference temperature. The voltage at 0°C should be subtracted from these readings to convert to a 0°C reference temperature.

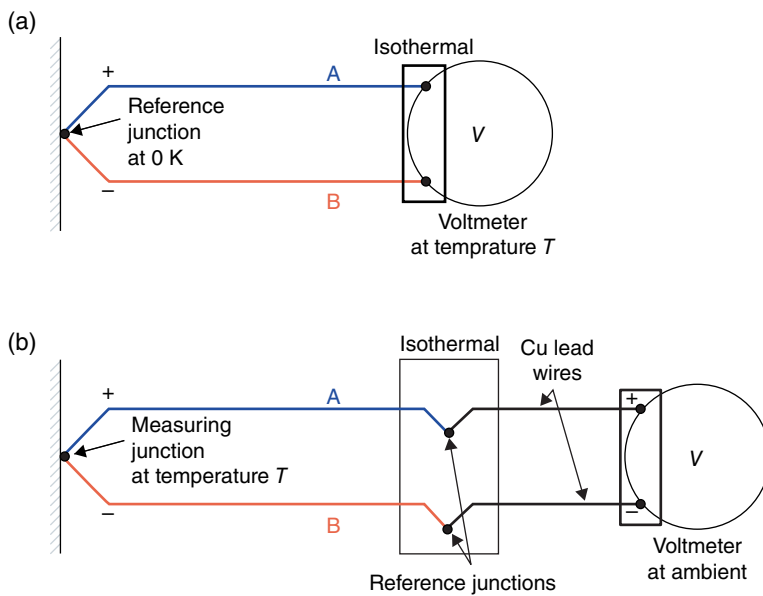


FIGURE 61.8 Schematic of a thermocouple measurement system with (a) 0 K reference temperature and (b) some other reference temperature.

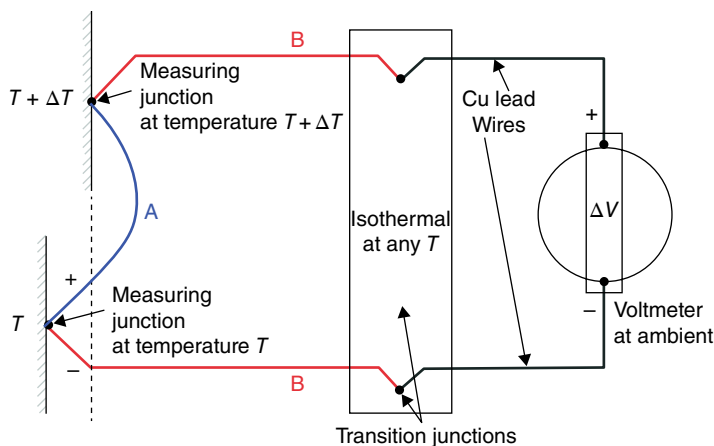


FIGURE 61.9 Schematic of a thermocouple measurement system for measuring small temperature differences.

voltage generation. The reference junctions could also be at the voltmeter connections, but ensuring isothermal conditions there is more difficult and the two leads extending from low T to ambient are thermocouple materials that may have high $S(T)$ and susceptible to larger variations due to material variations. The emf measured at the voltmeter due to a small temperature difference ΔT at the temperature T is given by

$$\Delta V = S_{AB}(T) \Delta T, \quad (61.8)$$

where the relative Seebeck coefficient $S_{AB}(T)$ is evaluated at the temperature T . The output voltage can be increased in such a measurement by the use of a thermopile, in which the thermocouple legs are crisscrossed between the two temperatures to form multiple junctions in series. For extremely high resolution, the thermocouple emf can be measured with a superconducting quantum interference device (SQUID) operating at 4.2 K [17]. Voltages of 10^{-13} V are easily resolved with a SQUID system, and their low-temperature operation eliminates most of the spurious thermal emfs in the system. The Au—Fe thermocouples should not be used in magnetic fields, since they are orientation dependent. The type E thermocouple has a small magnetic field dependence and can be used in moderate fields for temperatures above about 40 K.

61.2.2.5 Capacitance Thermometers The variation of capacitance for typical capacitance thermometers is shown in Figure 61.2 with capacitance shown on the right axis. The capacitance can shift equivalent to a kelvin or more upon thermal cycling, but after about an hour at temperature the drift is only a few tenths of a mK at 4.2 K but a few mK at 300 K. Their strong point is that they are very insensitive to magnetic fields. Their change in a magnetic field is less than 0.05% of the reading over the entire temperature range from 2 to 300 K. Thus, capacitance thermometers are best suited for

temperature control during the application of magnetic fields, but a different thermometer must be used to measure the temperature in zero field if that temperature must be known with an uncertainty less than about 1 K.

61.2.3 Thermometer Use and Comparisons

The selection of an appropriate thermometer for cryogenic applications makes use of a comparison of such factors as accuracy, reproducibility, temperature range, temperature resolution, size, cost, magnetic field effect, measuring instrument, and other factors. Accuracy involves how close the true thermodynamic temperature is measured, of which the ITS-90 temperature scale is the latest approximation to such a temperature. How close any thermometer follows the ITS-90 temperature scale depends on the uncertainty of calibration, the temperature resolution of the thermometer and temperature measurement system, and the reproducibility of the thermometer. Typical calibration uncertainties for commercial thermometers range from about ± 4 mK at 4 K, ± 10 mK at 80 K, and ± 25 mK at 300 K. Total thermometer uncertainties must also take into account reproducibility, of which typical values for the various thermometer types are given in Table 61.2. Individual thermometer calibrations over a wide temperature range ensure the lowest uncertainty, but the cost can be quite high. Much lower cost can be achieved by using interchangeable thermometers, such as platinum resistance thermometers, Si diodes, and thermocouples. Their interchangeability is indicated in Table 61.2. Figure 61.1 showed the construction details and size of typical SPRTs. The diameter and length of about 5.8 mm by 56 mm makes them too large for most experimental work. Most industrial thermometers are much smaller and are available in a variety of configurations, as shown in Figure 61.10. For example, wire wound PRTs conforming to IEC 60751 down to 70 K are available in capsules as small as 1.8 mm in diameter and 5 mm long. Thin film platinum RTDs are available in sizes as small as 2 mm \times 2 mm \times 1 mm thick that conform to IEC 60751 class B down to -50°C and cost about US\$1 each in quantities of 100. Quantitative results on the magnetic field effects of many types of thermometers are given by Sample and Rubin [18] and by Rubin et al. [19].

61.2.3.1 Temperature Resolution and Sensitivity The relative temperature resolution is given by

$$\frac{\Delta T}{T} = \frac{(\Delta V/V)}{S_d}, \quad (61.9)$$

where ΔV is the voltage resolution of the measurement system, V is the voltage, and S_d is the dimensionless sensitivity of the thermometer. The expression also applies to a resistor or a capacitor by replacing V with either R or C . For sensors with high-voltage outputs, such as volt-level signals with diode thermometers, a $5\frac{1}{2}$ digit voltmeter can

TABLE 61.2 Characteristics of Various Commercial Thermometers

Sensor	Excitation	Useful Range	Interchangeability	Reproducibility	Long-Term Drift	B Field $\Delta T/T$ in 10 ⁷
Pt (PRT) Pt100	1 mA	70–800 K <70 K cal.	0.5–0.1 K (A) 1–0.25 K (B)	5 mK	10 mK	1% at 77 K 0.1% at 0°C
Rh–Fe	1 mA	0.65–500 K	Poor	0.2 mK	0.2 mK	Poor
Cernox™	10 mV	0.1–420 K	Poor	0.02% of <i>T</i>	25 mK	<0.5%
Ge	1–3 mV	0.05–100 K	Poor	0.5 mK	1–10 mK	Poor
Carbon-glass	1–3 mV	1–325 K	Poor	1 mK@4K	4 mK	<5%
Carbon	1–3 mV	0.1–300 K	5% of <i>T</i>	0.1% of <i>T</i>	8 mK	<5%
Ru–O	10 mV	0.01–40 K	5–10% of <i>T</i>	15 mK	40 mK	<1%
Thermistor			Poor			Good
Si diode	10 μA	1.4–500 K	0.25 K (A) 0.5 K (B)	5–20 mK	10–40 mK	Poor
Thermocouple	—	1.2–1500 K	1 K	20 mK	50 mK	<5%
Capacitance	5 V at 5 kHz	1.4–290 K	Poor	0.3 K	1 K	<0.05%

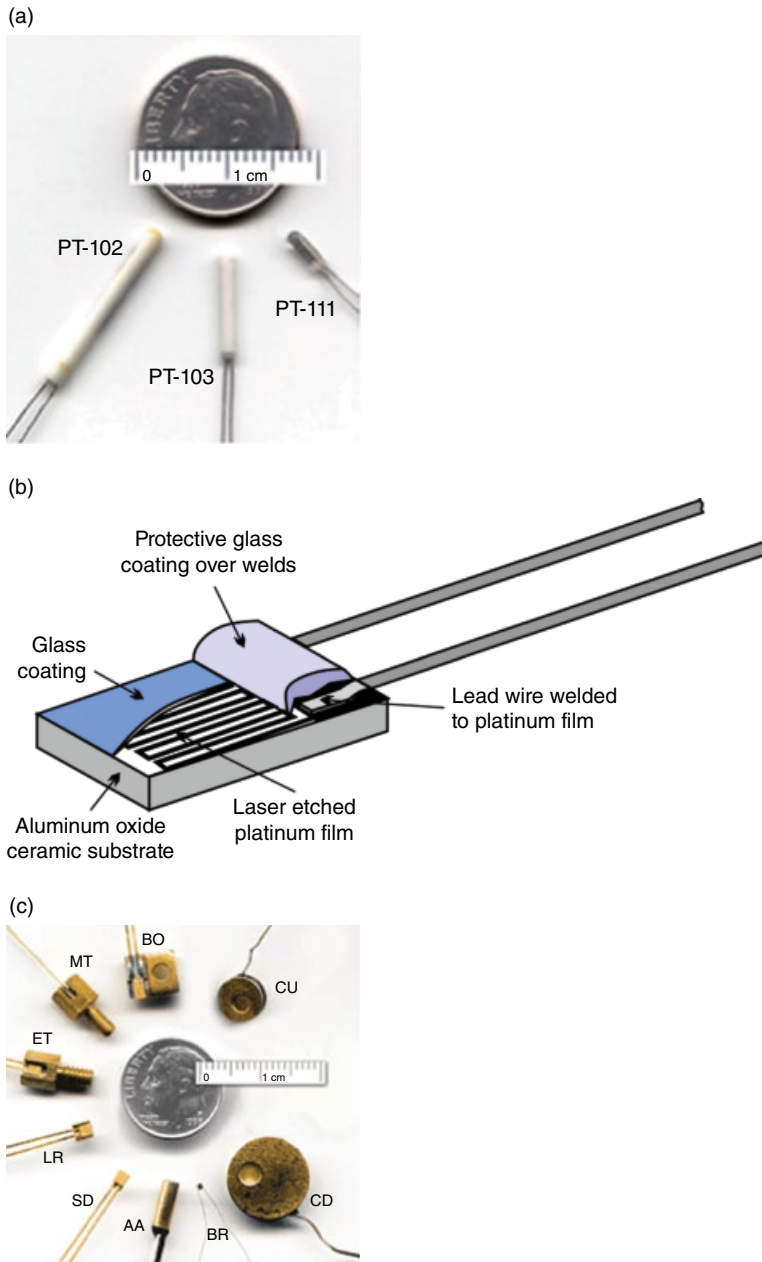


FIGURE 61.10 (a) Photo of several typical industrial wire wound platinum resistance thermometers. Source: Reproduced with permission of Lake Shore Cryotronics. (b) Drawing of a platinum film resistance thermometer. Typical dimensions are 2 mm×2 mm. (c) Photo of various types of packages available for many resistance and diode thermometers.

provide a voltage resolution of about 2×10^{-6} . For thermocouples the maximum output voltage may only be about 10 mV, so a voltmeter with 1 μ V resolution yields only $\Delta V/V = 1 \times 10^{-4}$. The dimensionless sensitivity of the thermometer is given by

$$S_d = \left| \frac{T}{V} \frac{dV}{dT} \right|, \quad (61.10)$$

where the voltage V can be replaced with R or C . A comparison of the dimensionless sensitivity for the various thermometer types discussed earlier is shown in Figure 61.11. The parameter O in this figure represents V , R , or C . The dimensionless sensitivity for thermocouples is calculated using the 0 K reference temperature, which can be misleading. For thermocouples it is best to determine the temperature resolution by the equation

$$\Delta T = \frac{\Delta V}{S_{AB}}, \quad (61.11)$$

where S_{AB} is determined from the sensitivity shown in Figure 61.6. For example, at 100 K the 30 μ V/K sensitivity for type E thermocouples results in a temperature

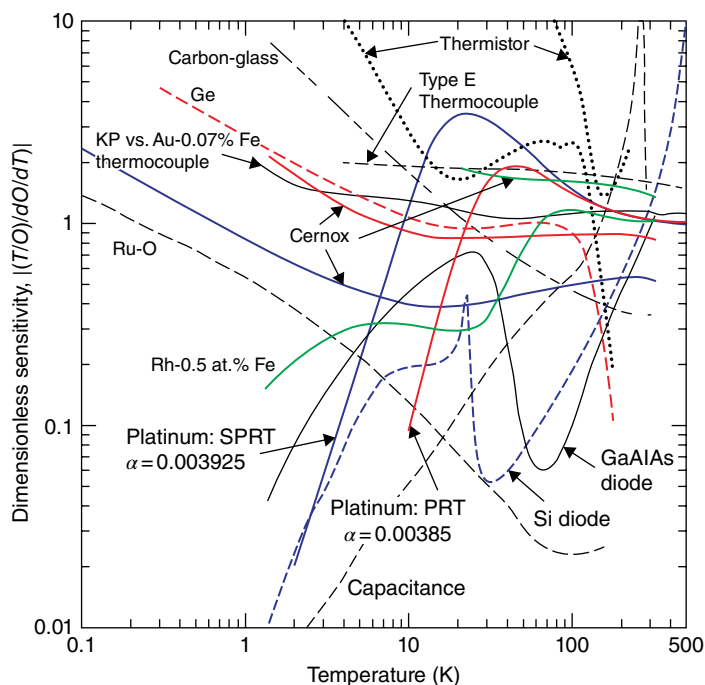


FIGURE 61.11 Dimensionless sensitivity of many types of thermometers. Sensitivity for thermocouples is shown with voltage for 0 K reference temperature. The parameter O can be R , V , or C .

resolution of 0.03 K for a 1 μ V voltage resolution. However, the uncertainty of the absolute temperature is best given by the standard error of 1.5 K shown in Table 61.1 that takes into account interchangeability, reproducibility, and long-term drift.

61.2.3.2 Thermometer Electrical Excitation Except for the case of thermocouples, all other thermometers discussed here require some type of electrical excitation. For most accurate measurements, a four-lead measurement should be used right to the thermometer element: two leads for current and two leads for voltage. Such an arrangement eliminates the error caused by voltage drop in a current-carrying lead. The standard calibration curves given for diodes are for an excitation of 10 μ A, as listed in Table 61.2. For resistance thermometers, any current or voltage can be used, provided it does not cause self-heating of the thermometer. Thermometers with positive temperature coefficients, such as the metallic resistance thermometers, are best excited with a constant current over a wide temperature range to allow for reduced power at lower temperatures. However, thermometers with negative temperature coefficients, such as semiconductor or semiconductor-like thermometers, are best excited with constant voltage. These typical currents and voltages are listed in Table 61.2. In practice, the maximum excitation can be determined by increasing the current or voltage until a change in resistance is detected. To prevent self-heating errors, the power dissipation should usually be less than about 1–10 μ W at 300 K and decrease to about 0.01–0.1 μ W at 4.2 K. The excitations listed in Table 61.2 usually yield these power levels. The power dissipation in the diode thermometers is in the range of 20–50 μ W at 4.2 K, which will result in self-heating if special care is not taken to thermally anchor them very well.

Resistance thermometers can be excited with either alternating current (AC) or direct current (DC). DC is most commonly used because of the availability of lower-cost instrumentation. However for the most precision work, AC is commonly used to allow for noise reduction with lock-in amplifiers. AC bridge networks with null detectors provide the ultimate in resistance resolution and have the added advantage of eliminating thermal EMFs caused by temperature gradients. A resistance resolution of 1 $\mu\Omega$ is possible, which gives a temperature resolution of 1 μ K in a 25 Ω SPRT [20]. A low frequency of about 30 Hz is commonly used to avoid any problems with reactance in the circuit. Precision work using DC excitation must reverse the current and take the average of the resistance from the two current directions to eliminate thermal EMFs.

61.2.3.3 Thermal Anchoring of Thermometers and Leads Any of the thermometers discussed here measure the temperature of the thermometer, so the uncertainties apply only to the thermometer itself and not to the sample to be measured. Because of heat leak through the electrical leads and the self-heating due to thermometer excitation, the thermometer temperature can be higher than the sample unless special care is taken to ensure good thermal contact between thermometer and sample and to minimize heat

conduction through the leads. Thermometer manufacturers are careful to provide good thermal contact between the thermometer element and any package. Examples of packages are shown in Figure 61.10. Canister packages must be inserted into a close fitting hole in a high thermal conductivity block or spool. Use of a thermal grease or epoxy ensures good thermal contact between the canister and the block, but the hole must not be blind to allow for air escape and easy thermometer disassembly by pushing on the end rather than pulling on the electrical leads. Figure 61.10 also shows many packages in which the thermometer canister has already been mounted by the manufacturer in a gold-plated copper spool. The gold plating minimizes radiation heating of the spool and provides a high thermal contact conductance when bolted to the sample. For contact areas greater than about 1 cm^2 , the use of thermal grease can provide improved thermal contact.

The second consideration in thermal anchoring thermometers is the thermal anchoring of the electrical leads separately from the thermometer. The Wiedemann–Franz law can be used to provide a good rule of thumb to relate heat leak in electrical leads to the electrical resistance of the lead between room and low temperature. For a low temperature of 77 K and a high temperature of 300 K, the Wiedemann–Franz law gives

$$\dot{Q}R \approx 1\text{ mW} \cdot \Omega, \quad (61.12)$$

where \dot{Q} is the heat flow and R is the electrical resistance. For a two-lead measurement circuit, the lead resistance must be much smaller than that of the thermometer. For a $100\text{ }\Omega$ Pt thermometer, the resistance is only about $10\text{ }\Omega$ at 80 K, so even a $0.1\text{ }\Omega$ resistance in each lead gives a 2% error in resistance and a heat leak of 20 mW in both leads according to Equation 61.12. A typical thermal conductance between a thermometer spool and the sample may be about 1 W/K , which then leads to a temperature difference of 20 mK between the spool and the sample. Another temperature difference will occur between the thermometer and the spool. The problem is not so serious with higher-resistance thermometers or with diodes that have a resistance of about $10^5\text{ }\Omega$. To eliminate the lead conduction problem, electrical leads should be thermally anchored to the sample independently of the thermometer. Often the leads are wrapped around a separate spool that is then bolted to the sample near the thermometer. Typical thermal tempering lengths for various wire sizes and materials are shown in Table 61.3.

Thermocouples are particularly difficult to thermally anchor to a cryogenic sample because of the small tip and the fact that most cryogenic samples will be in a vacuum. A portion of the thermocouple wire can be thermally anchored to the sample by thermal grease, epoxy, or a bolted spool as discussed for resistance thermometers. The tip can also be greased to the surface but care must be taken to ensure electrical isolation unless the measuring instrument has inputs isolated from ground. If that is the case, soldering the tip to the sample provides excellent thermal contact.

TABLE 61.3 Typical Wire Tempering Lengths for Thermometer Leads of Various Sizes and Materials^a

Material	T_h (K)	T_c (K)	Tempering Length for Various Wire Gages (cm)			
			0.080 mm (#40 AWG)	0.125 mm (#36 AWG)	0.200 mm (#32 AWG)	0.500 mm (#24 AWG)
Copper	300	80	1.9	3.3	5.7	16
	300	4	8.0	13.8	23.3	68.8
Phosphor bronze	300	80	0.4	0.6	1.1	3.2
	300	4	0.4	0.7	1.3	3.8
Manganin	300	80	0.2	0.4	0.4	2.1
	300	4	0.2	0.4	0.7	2.0
Stainless steel 304	300	80	0.2	0.3	0.6	1.7
	300	4	0.2	0.3	0.5	1.4

^aData from Ekin [7].

61.2.4 Dynamic Temperature Measurements

Most of the thermometers discussed previously have thermal time constants of several seconds. With some exceptions they are not designed for dynamic temperature measurements. Fast response times are achieved with thermometers that have low heat capacity (low mass or low specific heat) and good thermal contact between the sensing element and the object to be measured. The object to be measured can either be a solid object, in which case it can provide support for the thermometer, or a fluid, in which case the thermometer must be supported by some nearby solid and thermally insulated from it. The second case occurs, for example, in the measurement of the instantaneous temperature of the helium working fluid in regenerative cryocoolers. Typical operating frequencies may range from 1 to 60 Hz. To make such measurements at 60 Hz with negligible phase shifts requires a thermal time constant of less than about 300 μ s. In the measurement of instantaneous gas temperature, the dominant thermal resistance is often between the sensing element and the gas. As a result, the measurement of dynamic gas temperatures is usually more difficult to measure than that of liquids or solids.

By their nature, thermocouples have a small mass and a potential for fast response times. The use of small diameter thermocouple wire at cryogenic temperatures can often yield response times of a few tenths of seconds. Faster response times are obtained by using commercially available thin foil thermocouples. Foil thicknesses down to 5 μ m are available, but considerable care is required in handling the unsupported foil. For measurements of the surface temperatures of solids, a thermocouple film of 3–6 μ m thickness can be sputtered on the surface [21]. The internal response time of such a thin film will be about 1 μ s or less, but with the thermal resistance at the interface, the response to temperature changes at the surface may be considerably longer. The response time of unsupported 5 μ m thick type E thermocouple foil to

oscillating helium gas temperatures was measured at NIST and found to be about 10 ms at 80 K.

Thin film platinum or carbon thermometers can also be used for dynamic temperature measurements and have response times comparable to those of the thin film thermocouples. Louie and Steward [22] used an unsupported $4\mu\text{m}$ thick Pt foil in the measurement of transient heat transfer to liquid hydrogen for response times down to $10\mu\text{s}$. Carbon films on a quartz substrate have also been used for transient heat transfer experiments to liquid helium [23] and to liquid hydrogen [22]. Giarratano et al. [24] measured a response time of about $50\mu\text{s}$ at 77 K for an 18 nm thick platinum film on a quartz substrate.

For high-speed temperature measurements in the range of 1–20 K, a silicon-on-sapphire (SOS) thermometer is the fastest ever reported. The response time of these thermometers in both liquid and gaseous helium was found to be about 300 ns. [25]. These thermometers are made with a $1\mu\text{m}$ thick silicon layer on a 0.13 mm thick sapphire substrate. The silicon was ion implanted with phosphorus to give a resistance versus temperature curve similar to germanium resistance thermometers. These thermometers were used to study the temperature oscillations that occur in thermoacoustic oscillations inside small tubes closed at the room temperature end and open to a dewar of liquid helium at the other end [25]. The same thermometers were used for the measurement of instantaneous temperature of the helium gas inside a Stirling cryocooler next to the regenerator. Figure 61.12 shows how these thermometers were suspended from a fiberglass-epoxy support to measure the gas temperatures in a

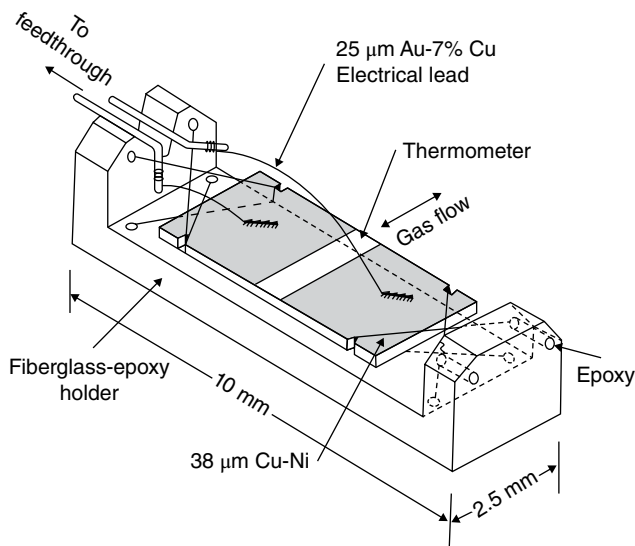


FIGURE 61.12 Drawing of a silicon-on-sapphire (SOS) thermometer for dynamic temperature measurements of flowing fluids at low temperature.

Stirling refrigerator at temperatures of about 10 K. The 38 μm diameter Cu—Ni support wires minimize the thermal contact between the thermometer and the support.

The thermal response times of several carbon, germanium, and diode thermometers at cryogenic temperatures were measured at NIST [26]. The response times reported there for the SOS thermometer were superseded by the measurements of Louie et al. [25]. At 4 K, 1/8 W carbon resistors showed response times as fast as 6 ms. Commercial Si diode thermometers in their basic sensor package have a response time at 4 K of about 10 ms as reported by their manufacturer. A response time of about 6 μs at 4 K has been reported for a miniature silicon diode thermometer [27]. The thin film metal oxynitride resistance thermometers have a reported time constant at 4 K of 1.5 ms as an unpackaged chip. When packaged inside a copper canister, the response time increases to about 0.4 s at 4 K.

For high-speed measurements of oscillating gas temperatures, platinum wire of about 5 μm gives a response time in still helium gas at 300 K of about 300 μs . Unfortunately, platinum wire of this small diameter is not strong enough to withstand oscillating mass flows that usually accompany the oscillating temperatures in gas. Instead, 3.8 μm diameter tungsten wire can be used, which is much stronger and has a measured response time of about 260 μs at 300 K in still helium gas [28, 29]. Thermometers made with 2 mm long segments of the 3.8 μm diameter tungsten wire have been used in oscillating gas flows at cryogenic temperatures for hours with no breakage, providing the gas is very clean. Such wire is commercially available from manufacturers of hot-wire anemometers. A description of a small demountable probe using this wire is given in the section on flow. The resistance of the tungsten wire has a linear temperature dependence down to about 77 K. Above that temperature its dimensionless sensitivity is 1.032, which is comparable to that of platinum.

61.3 STRAIN

The measurement of strain has many applications beyond its direct measurement. For example, strain gages are commonly used to measure force and pressure. In turn, pressure transducers are often used in the measurement of flow. Most measurements of strain, including those at cryogenic temperatures, are performed with bonded resistance strain gages. We restrict our discussion to these devices. An excellent review of resistance strain gages is given by Hannah and Reed [30], although their book emphasizes temperatures of 300 K and above. The principles are no different at cryogenic temperatures, but a proper materials selection is important. The resistance strain gage is an element whose resistance is a function of the applied strain. The relative resistance change can be expressed as

$$\frac{\Delta R}{R} = F_s \left(\frac{\Delta L}{L} \right), \quad (61.13)$$

where F_s is the gage factor or strain sensitivity factor and $\Delta L/L$ is the strain. Typical gage factors are about 2 for most of the commonly used metallic alloys. Their gage factors are nearly independent of strain for strain levels up to about ± 2000 microstrain (2000×10^{-6}). The metal alloy gages are usable at cryogenic temperatures for strain levels up to about 1.5%. Semiconductors can have gage factors of about 100 or more, but they are very temperature sensitive.

61.3.1 Metal Alloy Strain Gages

Most measurements of strain at temperatures from 4 to 300 K are made with a nickel-chromium alloy or a modified nickel-chromium alloy (73%Ni + 20%Cr + Al + Fe) either in the form of a wire or, more recently, in the form of photoetched foil. The copper-nickel alloy most often used at ambient temperatures has larger temperature and magnetic field effects and is seldom used for cryogenic temperatures. Typical resistance values for these gages are in the range of 60–1000 Ω , although 120 and 350 Ω gages are most commonly used at cryogenic temperatures. The alloy grid is bonded to a carrier matrix (backing) and usually has a geometry like that shown in Figure 61.13. Gage lengths vary from about 0.20 to 100 mm. The large areas in the region of the bends reduce the effects of transverse strain. Typical ratios of gage factors between transverse and longitudinal strains are only a few percent and have negligible effect on most measurements, unless high accuracy is desired. Many other gage geometries are available from gage manufacturers for use in various applications. The most common backing material for use at cryogenic temperatures is glass fiber-reinforced epoxy-phenolic. The polyimide backing commonly used for large strains at room temperature or above is seldom used for cryogenic temperatures. Most gages have a top layer of insulation bonded over the grid and backing. This top layer is known as the overlay or encapsulating layer. It is particularly important for use in liquid cryogenics to prevent the formation of bubbles at the surface of the metal caused by self-heating. These bubbles can lead to rapid localized temperature rises, which cause considerable

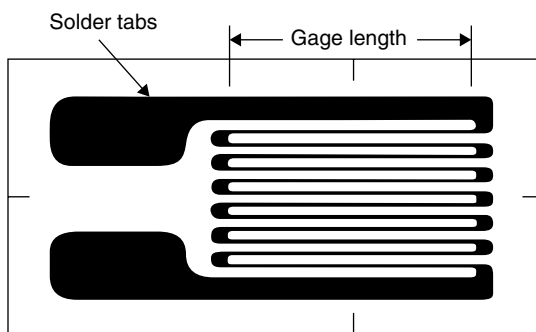


FIGURE 61.13 Geometry of a metal foil strain gage.

noise in the signal. For cryogenic use, the gage is usually bonded to the test specimen with an epoxy recommended by the manufacturer of the gage.

61.3.2 Temperature Effects

Various temperature effects can have a significant impact on the measurement of strain at cryogenic temperatures. There are three different temperature effects that need to be considered. The first is the effect of temperature on the gage factor. The gage factor for the modified nickel-chromium alloy varies linearly with temperature in such a way that the gage factor at 4 K is about 4–5% higher than the gage factor at 297 K [31]. For the copper-nickel alloy, the gage factor decreases by about 3% when cooled to 4 K from 297 K.

The second temperature effect is caused by the change in resistivity with temperature, or the temperature coefficient of resistivity (TCR). With the cryogenic alloys discussed here, the change in resistance of a strain gage when cooled from ambient temperature to 4 K is usually less than 5%. A 5% resistance change is the same change that would occur with a strain of 2.5% in a gage alloy with a gage factor of 2. The resistance change caused only by a temperature change is referred to as apparent strain or thermal output.

The third temperature effect is caused by the strain induced in the gage due to the difference in the thermal expansion between the specimen and the gage. This difference is a function of the specimen material and the gage material. With most metals, the difference in thermal contraction from ambient to 4 K is only a few tenths of a percent. In practice, the second and third temperature effects are combined into one apparent strain (A.S.) or thermal output that is a function only of temperature and the difference in thermal expansion between the gage and the specimen. Strain gage manufacturers can minimize this thermal output for a particular specimen thermal expansion by a proper heat treatment of the gage alloy. That technique is known as self-temperature compensation (STC). Figure 61.14 shows how this apparent strain varies with temperature for the modified nickel-chromium gage bonded to various materials and with the curves normalized at 280 K [32]. Shown for comparison is the dashed curve for a copper-nickel alloy gage bonded to a 304L stainless steel specimen. When the modified nickel-chromium curves are normalized at 4.2 K, all the curves agree with each other up to 20 K and reach a minimum of -700×10^{-6} at a temperature of 15 K.

In order to correct for the thermal output or apparent strain, the temperature of the specimen must be measured. Often it is submersed in liquid cryogenics, in which case the barometric pressure defines the bath temperature. When the specimen is not in a liquid bath, its temperature must be measured with a thermometer in thermal contact with a strain-free area of the specimen that is also in good thermal equilibrium with the portion subjected to the strain. Alternatively, the thermal output can be reduced to zero by using a temperature compensating circuit. This circuit is a Wheatstone bridge with two identical resistance strain gages used for the active and

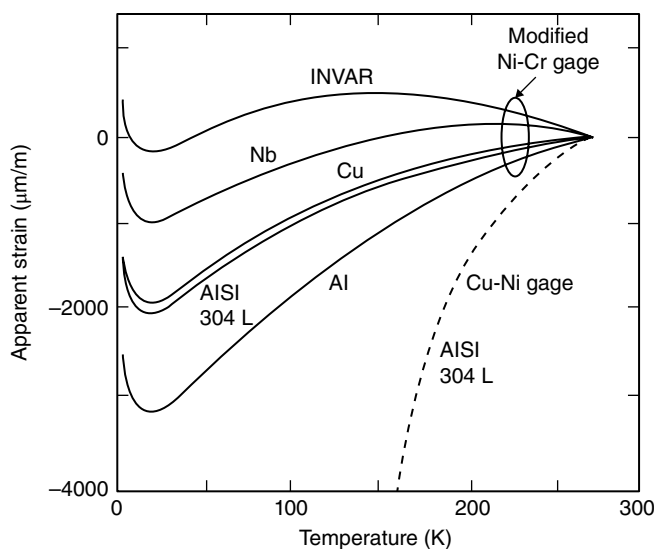


FIGURE 61.14 Apparent strain caused by temperature change from 280 K for Ni—Cr gages and Cu—Ni gages on various test materials.

the compensating (reference) arm. The compensating gage is mounted on a strain-free region of the sample that is at the same temperature as the strained region. If the specimen can remain strain-free until the test temperature is reached, then the thermal output can simply be canceled out by adjusting the reference resistor in the Wheatstone bridge circuit.

61.3.3 Magnetic Field Effects

A magnetic field can cause a change in resistance, which is known as the magnetoresistance effect. This resistance change leads to a strain error. The magnetic field can also affect the gage factor. Walstrom [33] measured the effect of magnetic fields up to 6 T on several strain gage alloys. For the nickel-chromium alloy, a strain error of $+160 \times 10^{-6}$ was found at 4.2 K in a magnetic field of 6 T. The error varied approximately quadratically with magnetic field. There was no detectable magnetoresistance strain error at 296 or 77 K with this alloy. The copper-nickel alloy gage showed much larger magnetoresistance effects. It showed a strain error of -250×10^{-6} at 296 K in a field of 6 T. Presumably, the error would be much larger at cryogenic temperatures, but it was not measured. These results indicate that copper-nickel gages should not be used in magnetic fields. The results of Walstrom also showed that the magnetoresistance effect was independent of field direction and independent of strain for strains up to 10^{-3} . There was about a 1% effect on the gage factor for fields above 3 T with the field perpendicular to the gage surface. No effect on the gage factor was seen with other field orientations.

Freynik et al. [34] extended the magnetoresistance measurements on nickel-chromium alloy gages up to magnetic fields of 12 T at 4.2 K. The strain error was found to be $+400 \times 10^{-6}$ at 4.2 K in a magnetic field of 12 T. Their data were in good agreement with those of Walstrom [33] for fields of 6 T and below.

61.3.4 Measurement System

The measurement of strain with resistance strain gages entails the detection of resistance changes that are often less than 1% of the resistance. Such small changes are best measured with a Wheatstone bridge adjusted for zero output at some known reference condition. An amplifier is used on the output voltage to significantly increase the resolution. Either DC or AC bridge excitation can be used, although most commercial strain gage systems use DC voltage. The use of AC excitation eliminates any thermal EMFs generated in the circuit and allows the use of lock-in amplifiers for enhanced signal-to-noise ratio. Any change in resistance in the electrical leads to the gage can be compensated by a three-wire bridge. As a result, most static measurements of strain should be made with three-wire bridges. The bridge excitation voltage must be kept sufficiently low to prevent self-heating of the gage. Usually, the excitation voltage is determined experimentally. For cryogenic applications, a bridge excitation of 2 V with a 350Ω gage is typical for use in a liquid cryogen. In vacuum, voltage levels down to 0.5 V may be necessary to prevent self-heating [35].

61.3.5 Dynamic Measurements

The intrinsic frequency response of the resistance strain gage should be in the tens or hundreds of kilohertz range. The adhesive joint will lower this frequency response some, but the resulting frequency response should be well above the maximum frequency used in most measurements. Few measurements are made at frequencies above 100 Hz because of the limitations in the equipment used to apply the dynamic strain [36]. In measurements of dynamic strain, the output of the Wheatstone bridge circuit is coupled to the amplifier through a capacitor (AC coupled) to eliminate any DC component. The AC coupling eliminates all effects associated with slow temperature changes of the specimen. Dynamic strain measurements are typically associated with fatigue measurements. The fatigue life of a properly selected gage can be as high as 10^8 cycles at a strain level of $\pm 2000 \times 10^{-6}$.

61.4 PRESSURE

The easiest and most common method for measuring pressure at cryogenic temperatures is to connect a capillary line between the desired pressure location and a pressure transducer at ambient temperature. In this case, a conventional pressure transducer can

be used. This method is limited to measurements of static pressure because of the low-frequency response of the capillary line. This method also has limitations in the low-pressure range because of the thermomolecular pressure correction that occurs in a temperature gradient [37]. The correction becomes particularly large for pressures below about 130 Pa. In this region, the pressure at the warm end of the capillary is higher than that at the cold end. Most commercial pressure transducers are designed for use at ambient temperature and cannot be used at cryogenic temperatures. We discuss some of the exceptions here. There are four types of pressure sensors or transducers that are commonly used at cryogenic temperatures, (i) capacitance, (ii) variable reluctance, (iii) strain gage or piezoresistive, and (iv) piezoelectric.

61.4.1 Capacitance Pressure Sensors

Variable capacitance pressure sensors are one of the most common types of pressure sensors used for precision work at cryogenic temperatures. However, we are not aware of any commercial units that have been used at these temperatures. An excellent review of capacitance pressure sensors used at cryogenic temperatures is given by Jacobs [38]. The sensor consists of a thin, stretched membrane, or for high pressures, a machined diaphragm, which deflects with pressure. It forms one electrode of the capacitor. The other electrode is formed by a stationary disk. The two electrodes must be electrically insulated from each other. The diaphragm and other parts of the sensor are usually of BeCu. A well-constructed capacitance sensor has less than a 5% change in sensitivity when cooled from 300 to 4 K. The capacitance of these sensors is typically in the range of 20–50 pF and can be measured with a capacitance bridge. For better accuracy, a three-lead bridge should be used to eliminate the effects of temperature-dependent lead capacitance. A frequency-to-voltage converter can also be used to measure the capacitance (pressure). Sensitivities of one part in 10^8 have been achieved with some capacitance pressure sensors, although a resolution of one part in 10^5 would be more common with inexpensive electronics. A disadvantage of the capacitance sensors is the requirement for coaxial cables between the sensor and the electronics.

61.4.2 Variable Reluctance Pressure Sensors

The variable reluctance pressure sensor utilizes a magnetically permeable stainless steel diaphragm. Deflection of the diaphragm is sensed by a pair of inductance coils on each side of the diaphragm, as shown in Figure 61.15. The magnetic reluctance of each of the circuits is a function of the gap between the diaphragm and the “E” core. A change in the reluctance on each side of the diaphragm changes the inductance of each of the coils. These two coils are connected in a bridge circuit, with the coils forming one-half of the bridge and a center tapped transformer forming the other half of the four-arm bridge. An AC signal of 3–5 kHz is commonly used in the bridge circuit. A carrier demodulator amplifies the output signal and converts it to a DC voltage

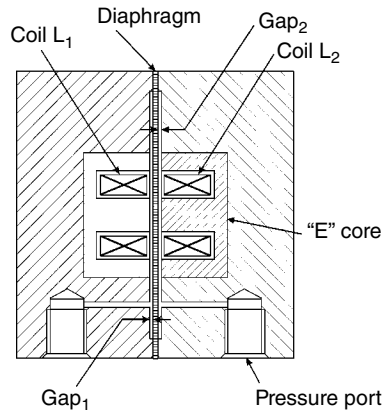


FIGURE 61.15 Cross section of a variable reluctance pressure transducer.

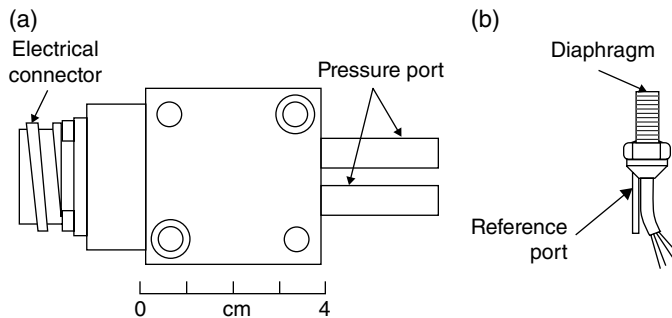


FIGURE 61.16 Two types of pressure transducers adaptable to cryogenic temperatures. (a) Variable reluctance and (b) piezoresistive.

proportional to the pressure. Because the coils are made of copper wire with very low resistance, the power dissipation in the transducer is quite small and acceptable for most cryogenic use. For temperatures of 77 K and below, the power dissipation is about 1 mW with conventional electronics using a 5 V excitation.

Commercial variable reluctance transducers are made with either all-welded construction or with rubber O-ring seals between the diaphragm and the transducer case. For cryogenic applications, the all-welded models must be chosen. Such transducers perform satisfactorily at temperatures down to 2 K. The differential pressure models with relatively low-pressure ranges (below about 100 kPa) remain linear at these low temperatures. Their sensitivity decreases by 12–13% from 300 to 77 K but remains unchanged from 77 to 4 K. They have a natural frequency of about 20 kHz and an internal volume of about 0.07 cm³ on each side of the diaphragm. Their external dimensions are relatively large, as shown in Figure 61.16a. Unfortunately, we have experienced some large deviations from linearity at 4 K with the high-pressure (5 MPa) models.

The calibration and vibration testing of many of these variable reluctance pressure transducers was reported by Kashani et al. [39] for use at temperatures down to 2.1 K.

The full-scale range of these transducers varied from 0.86 to 138 kPa. They found that the sensitivity of a 0.86 kPa transducer increased by about 4% from 300 to 4.2 K and then decreased by about 2% at 2.1 K. A 138 kPa transducer showed a 6% decrease in sensitivity when cooled from 300 to 4.2 K. There was no significant change in their responses after they were overpressurized. After temperature cycling the transducers two or three times, they exhibited linearity and repeatability to within $\pm 1\%$ of full scale in liquid helium. Daney [40] reported that a 5.5 kPa transducer showed a sensitivity change of only 0.2% between 300 and 4 K. Because of the different values of sensitivity changes reported by different authors on different transducers, it is important that each transducer be calibrated at cryogenic temperatures. Once calibrated, it should be repeatable to within $\pm 1\%$. The zero reading also shifts with temperature, but the transducer is normally rezeroed electronically after it has reached the desired temperature.

61.4.3 Piezoresistive Pressure Sensors

The piezoresistive pressure sensors use a strain gage to measure the deflection of a diaphragm subjected to a pressure on one side. They are the most common commercial pressure sensor. Some of them can be used at cryogenic temperatures if the materials have been properly selected. One type of sensor uses a metal diaphragm, typically stainless steel, with a thin film strain gage deposited on the diaphragm, or a metal foil strain gage bonded to the diaphragm. Various strain gage materials have been used, but the nickel-chromium alloy discussed in the previous section on strain is particularly good for use at cryogenic temperatures. Cerutti et al. [41] reported on tests with several commercial strain gage pressure sensors at temperatures down to 4.2 K and in magnetic fields up to 6 T. As is expected, their behavior is similar to that discussed for strain gages. For example, the calibration factor changes by about 5% for the nickel-chromium gages when cooled from 300 to 4 K. Thermal zero shifts of these same sensors were less than 3% of full scale (F.S.) for a temperature change from 293 to 4.2 K. The apparent pressure error at full scale due to a temperature change from 293 to 4 K was about 4% F.S. The combined nonlinearity and hysteresis was about $\pm 0.2\%$ F.S. at a fixed temperature of 4.2 K. When the sensor was cycled between 293 and 4.2 K, the nonlinearity and hysteresis increased to $\pm 2.3\%$ F.S. A magnetic field caused a maximum signal variation at 5 K of about 0.5% F.S., which occurred at low pressures and a field of 1.6 T.

Another type of piezoresistive pressure sensor uses a semiconductor strain gage to measure the deflection of a diaphragm. Doped silicon is nearly always used for these gages because the processing of silicon is a very well-established technology. In some cases, the diaphragm is also made of silicon with the strain gage grid diffused into the diaphragm. The silicon diaphragm is usually bonded to a stainless steel case by epoxy. The silicon diaphragm is often etched (micromachined) into a particular geometry to concentrate the stress at the region where the strain gage is located. This construction has the advantage of nearly eliminating all mechanical hysteresis in the sensor. The epoxy bond has the disadvantage of limiting the maximum negative pressure differential

to about 1 MPa before the epoxy bond will crack. The epoxy bond can occasionally develop leaks after rapid cooldown. The sensors that employ a stainless steel diaphragm welded to the case do not suffer from the limited negative pressure differential, and they are less likely to develop small leaks past the diaphragm. The silicon strain gage elements are bonded directly to the back side of the diaphragm to maintain a high-frequency response. The bond does have the disadvantage of slightly higher nonlinearity and hysteresis (0.5–1%) compared with that of the all-silicon construction (about 0.5%). These silicon pressure sensors, with stainless steel or silicon diaphragms, can be made as small as 1.3 mm diameter. A common configuration available in a wide range of full-scale pressure ratings has a diameter of 3.9 mm at the tip where the diaphragm is located. A 10–32 UNF-2A thread (4.8 mm diameter) allows the sensor to be screwed into a small pressure port. Figure 61.16b shows the geometry of this type of pressure sensor or transducer. A rubber O-ring fits in a groove under the head of these transducers to seal against the pressure port. For use at cryogenic temperatures, the O-ring must be replaced with a Teflon gasket about 0.13 mm thick. The gasket is compressed by a circular tongue on the mating assembly that fits closely within the O-ring groove to prevent extrusion of the Teflon. The geometry of this type of transducer makes it easy to place the diaphragm very close to the location where the pressure is to be measured. Their natural frequency is about 500 kHz, which permits pressure measurements at very high frequencies. These transducers can be used in the differential mode by utilizing the reference port, but the permissible pressure on the backside of the diaphragm is limited to about 1 MPa. We have not used these transducers in the differential mode.

The advantage of the silicon strain gage over the metallic strain gage in these pressure transducers is the greatly enhanced sensitivity. The 5 MPa transducers with silicon strain gages have sensitivities in the range of 3600–8700 mV/(V MPa) compared with 0.65 mV/(V MPa) for a transducer with a metal strain gage. The silicon transducers can be used with much cheaper readout electronics. The disadvantage of the silicon device is that it is much more temperature sensitive. Manufacturers of these transducers build in a temperature compensation circuit as part of the Wheatstone bridge that is good for the region of about 260–360 K. The zero shift and the sensitivity shift within this temperature range are less than 4%.

The sensitivity (V/Pa) and the zero reading of these silicon-based pressure sensors will change considerably when used at cryogenic temperatures. Boyd et al. [42] reported on precision calibration measurements of this type of sensor over the temperature range of 78–300 K. A total of 37 sensors were measured. Their sensitivities increased by a factor of 1.7–1.8 as the temperature was decreased from 278 to 78 K. The zero offset changed by about 1% F.S. over this temperature range. The curves for both the sensitivity and the zero offset indicate that they will continue to change as the temperature is reduced below 78 K, but at a slower rate. A thermal hysteresis of about $\pm 0.1\%$ F.S. was reported after many thermal cycles. An excitation of 1 mA was used with these sensors, which had a bridge resistance of about 5 k Ω . The power dissipation was

about 5 mW. Hershberg and Lyngdal [43] have shown that for pressure measurements at about 10 K or below, the power dissipation in the sensor should be less than about 5 mW. The normal power dissipation at 300 K in these transducers using the commercial electronics is about 50 mW.

Measurements made in our laboratory with a 5 MPa sensor containing a silicon diaphragm with a diffused silicon strain gage showed a sensitivity increase of 1.9 between 300 and 76 K with almost no change between 76 and 4 K. Such behavior is consistent with results reported by Clark [44] in which the sensitivity increased by about a factor of 1.8 between 300 and 77 K with little change below that temperature. The large change in sensitivity between 300 and 77 K means that the temperature of the pressure transducer must be measured accurately for use in this temperature range. For lower temperatures, the transducer temperature need not be measured accurately. Walstrom and Maddocks [45] tested a series of rather inexpensive semiconductor pressure sensors in the temperature range of 1.6–4.2 K. These sensors had an initial failure rate of about 20% upon cooldown, but those that survived could be cooled repeatedly. They found that the sensitivity at 4.2 K was about a factor of 2.4–2.6 higher than the room temperature value. They also found that the sensitivity at 1.6 K was about 5% higher than the value at 4.2 K and that the sensitivity at high pressure ($P \approx 100$ kPa) was about 5% less than the sensitivity at low pressure ($P < 10$ kPa).

For dynamic pressure measurements in a gas, the temperature of the gas and the silicon strain gage can vary with time. Some corrections may be necessary when using these transducers for dynamic pressure measurements between 77 and 270 K when the dynamic temperature is large. The piezoresistive pressure sensors are commonly used for pressure measurements in cryocoolers because of their small size and fast response.

61.4.4 Piezoelectric Pressure Sensors

Piezoelectric pressure sensors convert the stress applied to the sensing element (typically quartz crystal) to an electrical charge of the order of picocoulombs. A high-impedance charge amplifier converts the charge to a voltage output that will decay with time when the stress remains constant due to charge leakage through resistance in the output leads. These sensors therefore can only measure pressure changes or dynamic pressures. Some commercial piezoelectric sensors have the charge amplifier built into the sensor package to eliminate the need for special low-noise coaxial cables between the sensor and the charge amplifier at room temperature. The charge amplifier converts the high-impedance charge output from the sensor to a low-impedance voltage output. Any low-noise cable between the sensor and the charge amplifier must have insulation resistances as high as $10^{13} \Omega$. Sensors with a built-in charge amplifier are powered with a low-cost 24–27 VDC, 2–20 mA constant current supply. The voltage output is usually ± 5 V at full scale. A long coaxial or two-conductor ribbon

cable can be used to connect the sensor to the room temperature electronics without signal degradation.

Materials and mounting techniques in many of these sensors are compatible with cryogenic use. Some special models designed specifically for cryogenic operation and calibrated at 77 K are commercially available. A low-frequency limit in the range of 0.5 Hz for a 5% error is typical for these piezoelectric pressure sensors due to charge dissipation in the circuit resistance and capacitance. Resonant frequencies are generally greater than 250 kHz. Resolutions of 2×10^{-5} of full scale are possible. Models with full-scale dynamic pressure from 0.3 to 35 MPa are available. They can be used with static pressures much higher than the full-scale dynamic pressure, which makes them very useful for measurements of small dynamic pressure amplitudes superimposed upon a large static pressure. Typical applications are in the measurement of dynamic pressures in regenerative cryocoolers, such as pulse tube cryocoolers. A typical sensor package is about 38 mm long with a 3/8-24 UNF-2A thread or M6 thread for mounting. A brass or copper gasket is available for a leak-tight seal at cryogenic temperatures, but other models have rubber O-ring seals that cannot be used at low temperatures. Miniature models are also available with a 10–32 (4.8 mm) mounting thread and a total length of about 15 mm.

Piezoelectric pressure sensors are usually calibrated by the manufacturer at room temperature. Such a calibration is more complex than that with other pressure sensors because of the need to use a rapid pressure change or dynamic pressure. Calibrations at 77 K are available for some models. The temperature coefficient of sensitivity is about 0.07%/°C. In-house calibrations are often performed by comparison with piezoresistive pressure sensors, which can be calibrated with static pressure from a room temperature sensor, but are also capable of measuring high-frequency dynamic pressure.

61.5 FLOW

Many types of flowmeters have been used successfully at cryogenic temperatures for measurements of gas or liquid flows. Some are mass flowmeters while others are volumetric flowmeters. The determination of mass flow rate from a volumetric flow measurement requires a measurement of the fluid density. Some volumetric flowmeters contain a densitometer to yield an inferential mass flowmeter. The measurement of mass flow of cryogenic fluids is particularly important for custody transfer of cryogenics from tank trucks to the customer. Pipe sizes commonly used for this application vary from 3 to 9 cm with volumetric flow rates up to about 20 L/s. For liquefied natural gas (LNG), pipe sizes up to 20 cm are commonly used with flow rates up to about 100 L/s [46]. Due to space limitations, we cannot discuss all the types of flowmeters used for cryogenic service. The flowmeters considered here and their type (M=mass, V=volumetric, N=neither) are positive displacement (V), angular momentum (M), turbine

(V), differential pressure (N), thermal or calorimetric (M), and hot-wire anemometer (M). The differential pressure element can be in the form of an orifice, venturi, packed screens, or laminar flow channels. Other flowmeters not discussed here but used for cryogenic service are ultrasonic (V), vortex shedding (V), dual turbine (M), and Coriolis or gyroscopic (M). Detailed descriptions of many types of flowmeters used in cryogenic service are given by Alspach et al. [47], Brennan et al. [48, 49], and Brennan and Takano [50]. A discussion of the NIST cryogenic flowmeter calibration facility for use with liquid nitrogen and argon is given by Brennan et al. [46]. Most cryogenic flowmeters can be calibrated to an uncertainty of $\pm 0.5\%$ for volume flow and $\pm 0.2\%$ for mass flow in this facility [51]. The repeatability of individual flowmeters may be larger than these uncertainty values.

61.5.1 Positive Displacement Flowmeter (Volume Flow)

The positive displacement flowmeter works on the principle that the flowing liquid must displace some mechanical element. The movement of the mechanical element is then sensed electronically. The various mechanical elements used are screw impeller, rotating vane, and oscillating piston. A detailed description of these various types of positive displacement flowmeters and an evaluation of them for cryogenic service is given by Brennan et al. [48]. They are generally used with moderate flow rates (1–10 L/s) and are capable of being operated over a 5 to 1 flow range, which means the minimum flow is 1/5 of the maximum flow. A pressure drop of about 30 kPa is typical with these meters at maximum flow. With care it is possible to achieve an uncertainty of $\pm 1\%$ with these meters. It is important to subcool the liquid below the saturation curve to prevent the formation of vapor in the flowmeter. A disadvantage of these meters is that they are subject to wear and need to be recalibrated periodically.

61.5.2 Angular Momentum Flowmeter (Mass Flow)

The angular momentum flowmeter has a rotating member with vanes oriented parallel to the axis. The rotating member is driven by an electric motor through a constant torque clutch (hysteresis drive) as shown in Figure 61.17. The liquid enters the meter through a flow straightener and passes by the rotating vanes. The liquid tends to retard the rotational speed of the rotor in a manner that is inversely proportional to the mass flow rate. Rotor speed is sensed by a magnetic pickup, and the resulting signal treated electronically to indicate mass flow rate. A flow range of 8 to 1 is typical with these meters. Maximum flow rates for these meters may vary from about 2 to 15 kg/s. Pressure drops of 20–50 kPa are typical with these flowmeters at maximum flow. They have been tested with liquid hydrogen [47] and liquid oxygen, nitrogen, and argon [49]. Flowmeters of this type are often used as custody transfer flowmeters on delivery vehicles. Uncertainties of $\pm 2\%$ or less are typical for these flowmeters.

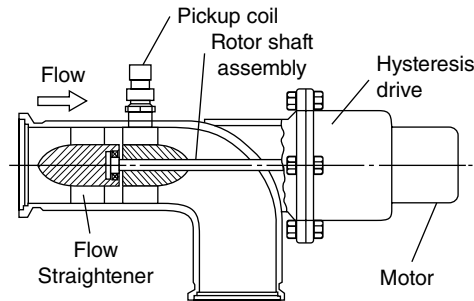


FIGURE 61.17 Angular momentum flowmeter. From Brennan et al. [49].

61.5.3 Turbine Flowmeter (Volume Flow)

The turbine flowmeter consists of a freely rotating bladed rotor, supported by bearings, inside a housing, and an electrical transducer that senses rotor speed. Rotor speed is a direct function of flow velocity. They are used mostly for liquid flows and have a useful range of at least 10 to 1. Calibrations with liquid cryogenics may differ from water calibrations by up to $\pm 2\%$ [52]. They are susceptible to errors caused by upstream swirl, so some means of flow straightening is usually required for accurate measurements. An evaluation of several cryogenic turbine flowmeters was reported by Brennan et al. [51]. They ranged in size from 3.2 to 5.1 cm with maximum flow rates between 5 and 14 L/s. Maximum pressure drops ranged from 20 to 100 kPa. Uncertainties were generally less than $\pm 1\%$.

A commercial turbine flowmeter with ball bearings has even been used for measuring flow in normal and superfluid helium with flow rates between 0.01 and 0.3 L/s [40]. It had a bore diameter of 9.35 mm. The meter output was about 0.5% higher for superfluid helium compared with normal helium. However, care must be taken with superfluid helium to prevent cavitation. A custom-made turbine flowmeter with magnetic bearings has also been used for liquid helium flow [53]. The success of these meters depends very much on maintaining a low drag from the bearings. To insure this, the liquid must be free of solid particles such as frozen air or water. An upstream filter is often used with these flowmeters. Short-term repeatability of $\pm 0.15\%$ was reported for the magnetic-bearing flowmeter. Like other flowmeters with moving parts, they do have a limited lifetime.

Turbine flowmeters have been reported to have response times in the range of 1–10 ms, depending upon blade angle, flowmeter size, and flow rate [52]. These authors also report on the successful use of these meters in reverse flow. As a result, they can be used to a limited extent in transient flow or oscillating flow for frequencies less than about 1–10 Hz.

61.5.4 Differential Pressure Flowmeter

The differential pressure flowmeter can be used with either gas or liquid flows. They operate on the principle that the pressure drop across some flow element is proportional to the flow rate. These meters can be used at cryogenic temperatures if the

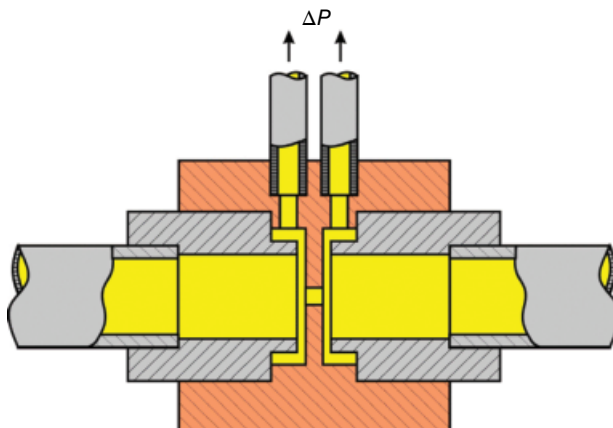


FIGURE 61.18 Orifice flowmeter for oscillating flow.

pressure transducer is located at ambient temperature or if a compatible pressure transducer is used at the cryogenic temperature. These flowmeters have no moving parts; thus, they are desirable for applications that require high reliability. The most common type of flow element is the sharp-edge orifice plate. Usually, the orifice plate is designed for flow in one direction with the sharp edge of the orifice on the entrance side. A symmetric design for the orifice plate as shown in Figure 61.18 has been used in our laboratory for use with oscillating flow [54]. When used for oscillating flows, it is important that the connecting lines are of the same length and that the differential pressure transducer is symmetrical. We have experienced some problems with shifts in the zero reading of differential pressure transducers when the differential pressure changes sign.

The relation between the mass flow rate and the pressure drop ΔP across the orifice is given by

$$\dot{m} = C_o A_o \left[\frac{2\rho\Delta P}{(1-\beta^4)} \right]^{1/2}, \quad (61.14)$$

where C_o is the orifice or discharge coefficient (≈ 0.6), A_o is the cross-sectional area of the orifice, ρ is the fluid density, and β is the ratio of the orifice diameter to the tube inside diameter. Because of the square root dependence of \dot{m} on ΔP , a range of 10 for ΔP yields a range of 3 in \dot{m} . This low flow range is a disadvantage of the orifice meter. The orifice coefficient determined from a water calibration can be used for most liquid cryogenics with $\pm 2\%$ uncertainty [49]. The uncertainty for use with gas can be somewhat higher. To obtain high accuracy with these meters, it is necessary to have a straight length of tube upstream of the orifice that is at least 20 times the tube diameter, and a length at least five times the tube diameter should be placed downstream of the orifice. Alternatively, flow straighteners in the form of tube bundles can be used if length is

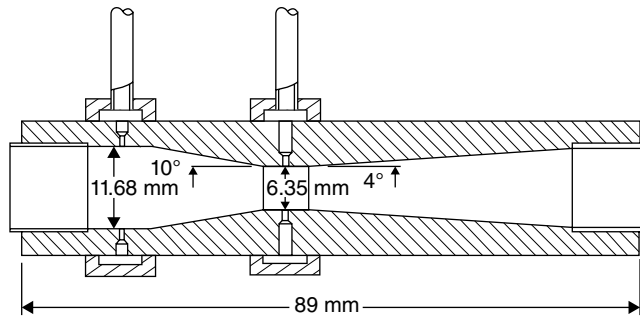


FIGURE 61.19 Cross section of a venturi flowmeter.

restricted. Equation 61.14 shows that this flowmeter is neither an intrinsic mass flowmeter nor an intrinsic volumetric flowmeter because of the square root dependence on the density. The simple design of these meters means that they can be scaled over a very wide range of flow rates. Pressure drops can be made quite small, although for ΔP less than about 1% of the mean pressure, the signal-to-noise ratio may begin to decrease.

A disadvantage of the orifice meter is the large amount of turbulence created by the flow through the orifice. This problem is reduced by using a venturi as the flow element, as shown in Figure 61.19. The throat diameter is usually about one-half the tube diameter. The flow rate through the venturi meter is governed by the same relationship as for the orifice meter (Equation 61.4); however, the discharge coefficient is near unity. Venturi meters have been used in many applications for measuring flow rates of normal, supercritical, and superfluid helium [40, 55]. Short-term repeatability of $\pm 0.5\%$ was reported. Discharge coefficients varied by about 3% over a flow range of 10 to 1 and for temperatures between 1.7 and 4.2 K. The design requirements of the venturi prevent it from being used for reverse flow, such as in fully reversing oscillating flow.

The third type of flow element that can be used in differential pressure flowmeters is the laminar flow element. The laminar flow element gives rise to a linear relationship between mass flow rate and pressure drop. As a result, it can be used over a wider range of flow rates than can the orifice meter and the venturi meter. The governing equation is given by

$$\dot{V} = \frac{\dot{m}}{\rho} = \frac{\Delta P}{Z_f}, \quad (61.15)$$

where \dot{V} is the volumetric flow rate and Z_f is the flow impedance of the laminar element. For laminar flow in a gap, the flow impedance is given by

$$Z_f = 12\mu L / wt^3, \quad (61.16)$$

where μ is the viscosity, L is the length of the gap, w is the width of the gap, and t is the thickness of the gap. Equation 61.15 shows that the laminar flow element is intrinsically a volumetric flowmeter. In order to achieve laminar flow conditions, the gap thickness must be sufficiently small. As an example, for helium gas with $\dot{m} = 1$ g/s, $P = 2$ MPa, $L = 1$ cm, and $\Delta P = 20$ kPa, the gap thickness must be less than $34\text{ }\mu\text{m}$ at 80 K and less than $9.8\text{ }\mu\text{m}$ at 10 K. The gap width must be 116 mm at 80 K and 240 mm at 10 K. Even though the overall gap width can be achieved with many parallel gaps, the outside dimensions of the laminar flow element will be relatively large. The small gap thickness at cryogenic temperatures makes the laminar flow element difficult to fabricate. No commercial laminar flow elements are available for cryogenic use. A custom-made device has been used with some success to measure oscillating mass flow rate at temperatures of about 10 K in high-pressure helium gas [54]. One caution to note on the use of a gap flow element for measuring oscillating flow is that the high velocity gas flow in the gap can lead to a significant inertance term, which will cause the phase of the pressure drop to lead that of the flow [56]. The inertance I (fluid equivalent of electrical inductance) of a gap is given by

$$I = \frac{\rho L}{wt}, \quad (61.17)$$

and the complex impedance is given by

$$Z_I = j\omega I, \quad (61.18)$$

where $j^2 = -1$ and ω is the angular frequency of sinusoidal oscillation. The presence of j in Equation 61.18 indicates that the component of the dynamic pressure drop due to the inertance leads the flow by 90° . The phase shifting becomes more important at higher frequencies. For oscillating flow the compliance (fluid analog of electrical capacitance) must also be considered to calculate the phase between the pressure drop and the flow.

The fourth type of flow element that can be used in differential flowmeters is that of packed screen or packed spheres. Correlations for the friction factor in such a packing must be used to find the relation between the pressure drop and the flow. With such geometries the relation between the pressure drop and flow is nonlinear. The effect of inertance is generally less in packed screen or packed spheres compared to that in gaps because of slower fluid velocities.

61.5.5 Thermal or Calorimetric (Mass Flow)

In the thermal flowmeter, the flowing fluid is heated with a constant power \dot{Q} , which causes its temperature to rise by an amount ΔT . A thermocouple or thermopile measures this temperature difference between the outgoing and incoming fluid flow. The mass flow rate is given by

$$\dot{m} = \frac{\dot{Q}}{C_p \Delta T}, \quad (61.19)$$

where C_p is the specific heat of the fluid. This flowmeter is a true mass flowmeter. A commercial thermal flowmeter was used by Bugeat et al. [57] to measure mass flow rates around 10 mg/s of hydrogen and helium gas at temperatures between 100 and 300 K. They reported a nonlinearity of less than 2% and a response time of about 0.22 s at 158 K in hydrogen gas for a flow of 9 mg/s. This type of flowmeter should work with flow in either direction.

61.5.6 Hot-Wire Anemometer (Mass Flow)

The hot-wire or constant temperature anemometer (CTA) infers mass flow rates from the changing heat transfer rates associated with a heated element [58]. The resistively heated element, often a fine wire, has a large temperature coefficient of resistance and a large length-to-diameter ratio. With feedback electronics, the electrical power to the element is varied automatically to maintain the element at a constant resistance (temperature) as the flow rate varies. The power, or voltage squared, is correlated to the mass flow rate from a calibration of the device against a standard flowmeter at ambient temperature. For an ideal gas, the CTA is a true mass flowmeter [59].

We have used commercial hot-wire anemometer probes successfully for the measurement of helium gas flows at temperatures down to 77 K [54, 59]. The calibration changes in a linear manner with the gas temperature, but it is not a strong function of this temperature. The probes were fabricated with a 3.8 μm diameter tungsten wire about 2 mm long attached to the wire supports. The wire was heated to about 297 K to give high sensitivity and reproducible results. Since the power input with this high wire temperature was about 1 W, the CTA was turned on only briefly in order to make the needed measurements. The response time of the CTA was measured to be less than 15 μs in zero flow. The response time is even faster at finite flow rates. The fast response time of the CTA makes it ideal for measuring turbulence, transient flow, or oscillating flow. Because of the fine wire, it can only be used in very clean gas flows.

The CTA has been used with modified commercial vacuum fittings as shown in Figure 61.20 to measure oscillating mass flow rates in compressed helium gas at temperature down to 77 K within a pulse tube refrigerator [59]. Oscillating frequencies up to 30 Hz could be measured. To provide the necessary temperature correction, an identical tungsten wire probe to serve as an RTD was inserted in the assembly from the opposite side (see Fig. 61.20). The diameter of the tubes extending from the device shown in Figure 61.20 was 3.2 mm. Consequently, this type of flowmeter can be made with very small gas volumes, which is necessary for measurements of the oscillating gas flow within small Stirling or pulse tube refrigerators. Several layers of stainless steel screen on each side of the probes ensured that the flow was uniform in both

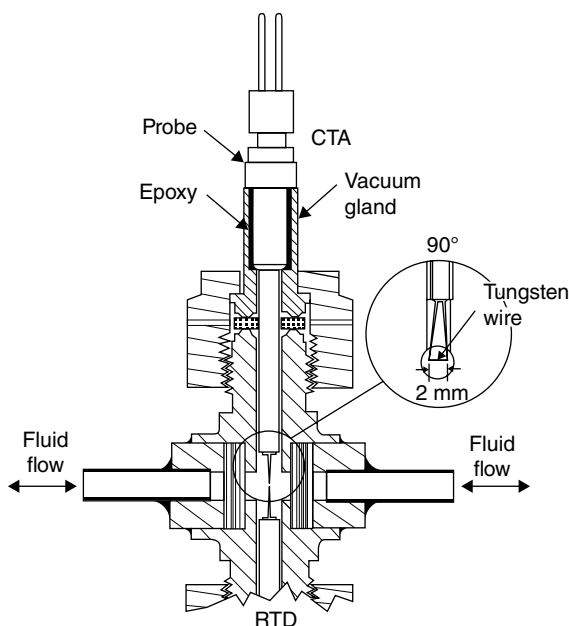


FIGURE 61.20 Modified vacuum assembly with CTA and RTD probes in place.

directions. The integrated mass flows measured for each direction of flow agreed to within 1.6% for measurements near 80 K.

61.6 LIQUID LEVEL

A common measurement problem in cryogenics is that of determining the level of cryogenics in storage dewars or within an experimental apparatus. For the heavier cryogenics (hydrogen and helium excluded), a measurement of the hydrostatic head with a differential pressure gage can be correlated to the liquid level. This technique is commonly used on storage dewars of liquid nitrogen for approximate ($\pm 10\%$) indications of liquid level. More precise measurements can be made with capacitance liquid-level gages that are available for use with liquid nitrogen from a few different manufacturers. The detection of liquid level at discrete locations (for level control) is often performed in many commercial devices with a self-heated resistance or diode thermometer. The higher heat transfer rate in the liquid phase causes the temperature of the thermometer to decrease when it is immersed in the liquid. When several of these thermometers are located at varying heights, a semicontinuous reading of liquid level is available. For continuous readings, a vertical wire or foil can be resistively heated [60, 61]. The resistance of the wire will be a function of the liquid level. The heat input to the liquid cryogen with this device can be minimized by using a current pulsed periodically to update the previous reading. Modern electronics makes this a simple task.

Greater sensitivity and less heating occurs when the wire is made with a superconducting material that has a critical temperature slightly above the normal boiling temperature of the liquid cryogen of interest. A current very near the critical current is passed through the wire so only the portion in the liquid remains in the superconducting state. Again, intermittent use reduces the total heat dissipation. For liquid nitrogen, the wire can be a high-temperature superconductor. For liquid helium, the wire must be a low-temperature superconductor such as tantalum ($T_c = 4.4$ K). The current through the wire must be varied if the temperature of the helium bath is lowered. These superconducting level sensors are commercially available from several manufacturers.

61.7 MAGNETIC FIELD

Instruments that measure magnetic field are usually called gaussmeters, teslameters, or magnetometers. For measurements of weak magnetic fields, the Superconducting Quantum Interference Device (SQUID) magnetometer is unsurpassed. It can detect changes in magnetic field as small as 10^{-15} T [17]. They can be used for fields up to about 1 mT.

For measurement of high magnetic fields at cryogenic temperatures, magnetoresistive sensors or Hall effect sensors can be used. The principle of magnetoresistive sensors is that the resistance of a metal or semimetal changes with the applied magnetic field. This change can vary from 2 to 3% change in a nickel-iron alloy to as much as 10^6 in bismuth. For small values of magnetic field, the change in resistance is proportional to the square of the magnetic field, whereas for larger fields it may have several higher order terms. Magnetoresistance is also a function of temperature. This complex behavior makes it difficult to use magnetoresistive sensors for measuring magnetic fields over a wide range of temperatures and fields.

The Hall effect sensor works as follows: A current is passed through the sensor in the x direction. With a magnetic field applied in the z -direction, a voltage in the y -direction is generated that is related to the magnetic flux density B . With proper cancellation of offset voltages, the output of the Hall effect element is given by

$$V = \left(2\mu_H \frac{w}{L} R \right) IB, \quad (61.20)$$

where μ_H is the electron Hall mobility, w/L is the effective width-to-length ratio for the Hall element, I is the excitation current, R is the resistance of the Hall element, and B is the magnetic flux density. Any temperature dependence comes about as a result of μ_H and R . The expression within the parentheses makes up the Hall sensitivity γ . It is a weak function of B . Glowacki and Ignatowicz [62] showed that $\text{Cd}_x\text{Hg}_{1-x}\text{Te}$ ($x=0.175$) films produced good Hall effect sensors for use between 4.2 and 20 K. There was no temperature dependence in this range and the Hall sensitivity varied by about 20%

between a magnetic flux density of 0.1 and 2 T. The output was very reproducible after many thermal cycles.

Commercial cryogenic Hall effect sensors are available along with the control electronics from a variety of manufacturers. The usual materials are InAs, InSb, and GaAs. Sample and Rubin [18] have measured the temperature dependence and the linearity of several of these Hall sensors. They recommend the Hall sensor over other magnetic field sensors for use at cryogenic temperatures. These sensors are available in axial or transverse field models. Typical diameters are about 6 mm and a length of 5 mm for the axial sensor. The transverse sensor is flat with a width of 5 mm and a length of 16 mm. They can be used in magnetic fields up to ± 15 T with deviations from linearity less than 1.5%. Outputs at 4.2 and 77 K are within $\pm 1.5\%$ of the calibration at 300 K. Repeatability after many thermal cycles is within 1%. However, they are susceptible to damage from thermal shock after repeated cycling.

61.8 CONCLUSIONS

We have discussed and compared the various sensors and instrumentation that are commonly used for measurements at cryogenic temperatures. In most cases, commercial products are available for this need. We have reviewed available instrumentation for measurements of temperature, strain, pressure, flow, liquid level, and magnetic field at cryogenic temperatures. The comparisons of various sensors should allow the reader to quickly determine which sensor is best suited for the task at hand. The references cited should be useful if more details are needed.

REFERENCES

1. Preston-Thomas, H., 1990, "The International Temperature Scale of 1990 (ITS-90)," *Metrologia*, Vol. 27, pp. 3–10, and 107.
2. BIPM, 2011, "Supplementary information for the realization of the PLTS-2000", Adopted by the Consultative Committee for Thermometry, International Committee for Weights and Measures, pp. 1–25; slightly modified version of, Rusby, R. L., Fellmuth, B., Engert, J., Fogle, W. E., Adams, E. D., Pitre, L., and Durieux, M., 2007, "Realization of the 3He melting pressure scale, PLTS-2000," *J. Low Temp. Phys.*, Vol. 149, pp. 156–175.
3. Mangum, B. W., Furukawa, G. T., Kreider, K. G., Meyer, C. W., Ripple, D. C., Strouse, G. F., Tew, W. L., Moldover, M. R., Carol Johnson, B., Yoon, H. W., Gibson, C. E., and Saunders, R. D., 2001, "The kelvin and temperature measurements," *J. Res. Natl. Inst. Stand. Technol.*, Vol. 106, pp. 105–149.
4. Strouse, G. F., 2008, *Standard Platinum Resistance Thermometer Calibrations From the AR TP to the Ag FP*, NIST Special Publication SP 250-81, U.S. Department of Commerce, Technology Administration, National Institute of Standards and Technology, Gaithersburg, MD.

5. Tew, W. L. and Meyer, C. W., 2003, "Recent Results of NIST Realization of the ITS-90 Below 84 K," *Temperature: Its Measurement and Control in Science and Technology*, Vol. 7, D. C. Ripple, ed., American Institute of Physics, New York, pp. 143–148.
6. Courts, S. S., Holmes, D. S., Swinehart, P. R., and Dodrill, B. C., 1991, "Cryogenic thermometry," *Applications of Cryogenic Technology*, Vol. 10, J. P. Kelley, ed., Plenum Press, New York, pp. 55–69.
7. Ekin, J. W., 2006, *Experimental Techniques for Low-Temperature Measurements*, Oxford University Press, Oxford, pp. 185–225.
8. Holmes, D. S. and Courts, S. S., 1992, "Resolution and accuracy of cryogenic temperature measurements," *Temperature: Its Measurement and Control in Science and Industry*, Vol. 6, J. F. Schooley, ed., American Institute of Physics, New York, pp. 1225–1230.
9. Rubin, L. G., Brandt, B. L., and Sample, H. H., 1982, "Cryogenic thermometry: a review of recent progress, II," *Cryogenics*, Vol. 22, pp. 491–503.
10. Rubin, L. G., 1997, "Cryogenic thermometry: a review of progress since 1982," *Cryogenics*, Vol. 37, pp. 341–356.
11. Sparks, L. L., 1983, "Temperature, strain, and magnetic field measurements," *Materials at Low Temperatures*, R. P. Reed and A. F. Clark, eds., American Society for Metals, Metals Park, OH, pp. 515–571.
12. Yeager, C. J. and Courts, S. S., 2001, "A review of cryogenic thermometry and common temperature sensors," *IEEE Sensors J.*, Vol. 1, pp. 352–360.
13. Courts, S. S. and Krouse, J. K.; *Temperature: Its Measurement and Control in Science and Industry*. 2013, "A new capsule platinum resistance thermometer for cryogenic use," *AIP Conf. Proc.*, Vol. 1552, 8, pp. 168–173.
14. Peng, L., 2012, "A proposal for RhFe thermometer used as the interpolating instrument of temperature scale in the range of 1 K to 25 K," 26th International Conference on Low Temperature Physics, *J. Phys. Conf. Ser.*, Vol. 400, p. 052016.
15. Lawless, W. N., 1972, "Thermometric properties of carbon impregnated porous glass at low temperatures," *Rev. Sci. Instrum.*, Vol. 43, pp. 1743–1747.
16. Bentley, R. E., 1998, *Handbook of Temperature Measurement, Vol. 3: The Theory and Practice of Thermoelectric Thermometry*, Springer, Singapore.
17. Fagaly, R. L., 1987, "Superconducting magnetometers and instrumentation," *Sci. Prog.*, Vol. 71, pp. 181–201.
18. Sample, H. H. and Rubin, L. G., 1977, "Instrumentation and methods for low temperature measurements in high magnetic fields," *Cryogenics*, Vol. 17, pp. 597–606.
19. Rubin, L. G., Brandt, B. L., and Sample, H. H., 1986, "Some practical solutions to measurement problems encountered at low temperatures and high magnetic fields," *Advances in Cryogenic Engineering*, Vol. 31, R. W. Fast, ed., Plenum Press, New York, pp. 1221–1230.
20. Cutkowsky, R. D., 1970, "An A-C resistance thermometer bridge," *J. Res. Natl. Bur. Stand.*, Vol. 74C, pp. 15–18.
21. Kreider, K. G., 1992, "Thin-film thermocouples," *Temperature: Its Measurement and Control in Science and Industry*, Vol. 6, J. F. Schooley, ed., American Institute of Physics, New York, pp. 643–648.

22. Louie, B. and Steward, W. G., 1990, "Onset of nucleate and film boiling resulting from transient heat transfer to liquid hydrogen," *Advances in Cryogenic Engineering*, Vol. 35A, R. W. Fast, ed., Plenum Press, New York, pp. 403–412.
23. Giarratano, P. J. and Steward, W. G., 1983, "Transient forced convection heat transfer to helium during a step in heat flux," *J. Heat Transf.*, Vol. 105, pp. 350–357.
24. Giarratano, P. J., Lloyd, F. L., Mullen, L. O., and Chen, G. B., 1982, "A thin platinum film for transient heat transfer studies," *Temperature: Its Measurement and Control in Science and Industry*, Vol. 5, J. F. Schooley, ed., American Institute of Physics, New York, pp. 859–863.
25. Louie, B., Radebaugh, R., and Early, S. R., 1986, "A thermometer for fast response in cryogenic flow," *Advances in Cryogenic Engineering*, Vol. 35A, R. W. Fast, ed., Plenum Press, New York, pp. 1235–1246.
26. Linenberger, D., Spellicy, E., and Radebaugh, R., 1982, "Thermal response times of some cryogenic thermometers," *Temperature: Its Measurement and Control in Science and Industry*, Vol. 5, J. F. Schooley, ed., American Institute of Physics, New York, pp. 1367–1372.
27. Rao, M. G., Scurlock, R. G., and Wu, Y. Y., 1983, "Miniature silicon diode thermometers for cryogenics," *Cryogenics*, Vol. 23, pp. 635–638.
28. Rawlins, W., Radebaugh, R., and Timmerhaus, K. D., 1991, "Monitoring rapidly changing temperatures of the oscillating working fluid in a regenerative refrigerator," *Applications of Cryogenic Technology*, Vol. 10, J. P. Kelley, ed., Plenum Press, New York, pp. 71–83.
29. Rawlins, W., Timmerhaus, K. D., and Radebaugh, R., 1992, "Resistance thermometers with fast response for use in rapidly oscillating gas flows," *Temperature: Its Measurement and Control in Science and Industry*, Vol. 6, J. F. Schooley, ed., American Institute of Physics, New York, pp. 471–474.
30. Hannah, R. L. and Reed, S. E., 1992, *Strain Gage Users' Handbook*, Elsevier Applied Science, New York.
31. Starr, J. E., 1992, "Basic strain gage characteristics," *Strain Gage Users' Handbook*, R. L. Hannah and S. E. Reed, eds., Elsevier Applied Science, New York, pp. 1–77.
32. Pavese, F., 1984, "Investigation of transducers for large-scale cryogenic systems in Italy," *Advances in Cryogenic Engineering*, Vol. 29, R. W. Fast, ed., Plenum Press, New York, pp. 869–877.
33. Walstrom, P. L., 1975, "The effect of high magnetic fields on metal foil strain gauges at 4.2 K," *Cryogenics*, Vol. 15, pp. 270–272.
34. Freynik, H. S., Roach, D. R., Deis, D. W., and Hirzel, D. G., 1978, "Evaluation of metal-foil strain gauges for cryogenic application in magnetic fields," *Advances in Cryogenic Engineering*, Vol. 24, K. D. Timmerhaus, R. P. Reed, and A. F. Clark, eds., Plenum Press, New York, pp. 473–479.
35. Ferrero, C. and Marinari, C., 1990, "Strain analysis at cryogenic temperatures: self-heating effect and linearization of the apparent-strain curve," *Advances in Cryogenic Engineering*, Vol. 35B, R. W. Fast, ed., Plenum Press, New York, pp. 1609–1616.
36. Hartwig, G. and Wüchner, F., 1975, "Low temperature mechanical testing machine," *Rev. Sci. Instrum.*, Vol. 46, pp. 481–485.

37. McConnville, G. T., 1969, "Thermomolecular pressure corrections in helium vapour pressure thermometry: the effect of the tube surface," *Cryogenics*, Vol. 9, pp. 122–127.
38. Jacobs, R., 1986, "Cryogenic applications of capacitance-type pressure sensors," *Advances in Cryogenic Engineering*, Vol. 31, R. W. Fast, ed., Plenum Press, New York, pp. 1277–1284.
39. Kashani, A., Wilcox, R. A., Spivak, A. L., Daney, D. E., and Woodhouse, C. E., 1990, "SHOOT flowmeter and pressure transducers," *Cryogenics*, Vol. 30, pp. 286–291.
40. Daney, D. E., 1988, "Behavior of turbine and venturi flowmeters in superfluid helium," *Advances in Cryogenic Engineering*, Vol. 33, R. W. Fast, ed., Plenum Press, New York, pp. 1071–1079.
41. Cerutti, G., Maghenzani, R., and Molinar, G. F., 1983, "Testing of strain-gauge pressure transducers up to 3.5 MPa at cryogenic temperatures and in magnetic fields up to 6T," *Cryogenics*, Vol. 23, pp. 539–545.
42. Boyd, C., Juanarena, D., and Rao, M. G., 1990, "Cryogenic pressure sensor calibration facility," *Advances in Cryogenic Engineering*, Vol. 35B, R. W. Fast, ed., Plenum Press, New York, pp. 1573–1581.
43. Hershberg, E. L. and Lyngdal, J. W., 1994, "Self heating in piezoresistive pressure sensors at cryogenic temperatures," *Advances in Cryogenic Engineering*, Vol. 39, P. Kittel, ed., Plenum Press, New York, pp. 1123–1130.
44. Clark, D. L., 1992, "Temperature compensation for piezoresistive pressure transducers at cryogenic temperatures," *Advances in Cryogenic Engineering*, Vol. 37B, R. W. Fast, ed., Plenum Press, New York, pp. 1447–1452.
45. Walstrom, P. L. and Maddocks, J. R., 1987, "Use of Siemens KPY pressure sensors at liquid helium temperatures," *Cryogenics*, Vol. 27, pp. 439–441.
46. Brennan, J. A., LaBrecque, J. F., and Kneebone, C. H., 1976, "Progress report on cryogenic flowmetering at the National Bureau of Standards," *Instrumentation in the Cryogenic Industry, Proceedings of the First Biennial Symposium*, Vol. 1, Houston, TX, October 11–14, 1976, Instrument Society of America, Pittsburgh, PA, pp. 621–636.
47. Alspach, W. J., Miller, C. E., and Flynn, T. M., 1966, "Mass flowmeters in cryogenic service," *Flow Measurement Symposium, ASME Flow Measurement Conference*, Pittsburgh, PA, September 26–28, 1966, American Society of Mechanical Engineers, New York, pp. 34–56.
48. Brennan, J. A., Dean, J. W., Mann, D. B., and Kneebone, C. H., 1971, *An Evaluation of Positive Displacement Cryogenic Volumetric Flowmeters*, National Bureau of Standards Technical Note 605, U.S. Department of Commerce, Washington, DC.
49. Brennan, J. A., Stokes, R. W., Kneebone, C. H., and Mann, D. B., 1974, *An Evaluation of Selected Angular Momentum, Vortex Shedding, and Orifice Cryogenic Flowmeters*, National Bureau of Standards Technical Note 650, U.S. National Bureau of Standards, Washington, DC.
50. Brennan, J. A. and Takano, A., 1982, "A preliminary report on the evaluation of selected ultrasonic and gyroscopic flowmeters at cryogenic temperatures," *Proceedings of the Ninth International Cryogenic Engineering Conference*, K. Yasukochi and H. Nagano, eds., Butterworth, Guildford, Surrey, pp. 655–658.
51. Brennan, J. A., Mann, D. B., Dean, J. W., and Kneebone, C. H., 1972, "Performance of NBS cryogenic flow research facility," *Advances in Cryogenic Engineering*, Vol. 17, K. D. Timmerhaus, ed., Plenum Press, New York, pp. 199–205.

52. Alspach, W. J. and Flynn, T. M., 1965, "Considerations when using turbine-type flowmeters in cryogenic service," *Advances in Cryogenic Engineering*, Vol. 10, K. D. Timmerhaus, ed., Plenum Press, New York, pp. 246–252.
53. Rivetti, A., Martini, G., Gorla, R., and Lorefice, S., 1987, "Turbine flowmeter for liquid helium with the rotor magnetically levitated," *Cryogenics*, Vol. 27, pp. 8–11.
54. Radebaugh, R. and Rawlins, W., 1993, "Measurement of oscillating mass flows at low temperatures," *Devices for Flow Measurement and Control—1993*, C. J. Blechinger and S. A. Sherif, eds., American Society of Mechanical Engineers, New York, pp. 25–32.
55. Rivetti, A., Martini, G., and Birello, G., 1994, "Metrological performances of Venturi flowmeters in normal, supercritical, and superfluid helium," *Advances in Cryogenic Engineering*, Vol. 39, P. Kittel, ed., Plenum Press, New York, pp. 1051–1058.
56. Yuan, S. W. K., Curran, D. G. T., and Cha, J. S., 2010, "A non-tube inertance device for pulse tube cryocoolers," *Advances in Cryogenic Engineering*, Vol. 55, J. G. Weisend II, ed., American Institute of Physics, New York, pp. 143–148.
57. Bugeat, J. P., Petit, R., and Valentian, D., 1987, "Thermal helium mass flowmeter for space cryostat," *Cryogenics*, Vol. 27, pp. 4–7.
58. Perry, A. E., 1982, *Hot Wire Anemometry*, Clarendon Press, Oxford.
59. Rawlins, W., Radebaugh, R., and Timmerhaus, K. D., 1993, "Thermal anemometry for mass flow measurement in oscillating cryogenic gas flows," *Rev. Sci. Instrum.*, Vol. 64, pp. 3229–3235.
60. Maimoni, A., 1956, "Hot wire liquid-level indicator," *Rev. Sci. Instrum.*, Vol. 27, pp. 1024–1027.
61. Wexler, A. and Corak, W. S., 1951, "Measurement and control of the level of low boiling liquids," *Rev. Sci. Instrum.*, Vol. 22, pp. 941–945.
62. Glowacki, B. A. and Ignatowicz, S. A., 1987, "Hall probe $\text{Cd}_x\text{Hg}_{1-x}\text{Te}$ for use in magnetic investigations of Nb_3Sn superconducting layers," *Cryogenics*, Vol. 27, pp. 162–164.

TEMPERATURE-DEPENDENT FLUORESCENCE MEASUREMENTS

JAMES E. PARKS¹, MICHAEL R. CATES², STEPHEN W. ALLISON²,
DAVID L. BESHEARS², M. AL AKERMAN², AND MATTHEW B. SCUDIERE²

¹*Department of Physics, University of Tennessee, Knoxville, TN, USA*

²*Emco-Williams Inc, Knoxville, TN, USA*

62.1 INTRODUCTION

Temperature is one of the most important attributes of physical systems, and its measurement is critical to many aspects of scientific research and development. There are a significant number of circumstances, however, in which temperature is difficult or impossible to measure by standard means such as thermometers, thermocouples, and infrared surface emissions. Most of these circumstances are associated with challenges such as very high temperature, vibrating or moving surfaces, difficulty of access, hazardous locations, and the like. Typical examples include centrifuges, turbine engine components, high-speed motors, and vibrating or moving production machinery. In many situations, it is also necessary to measure temperature remotely, without direct contact, because of difficulties of access, intervening heated air or other gases, or movement of the component to be measured [1–6].

Fluorescent materials (phosphors), bonded to surfaces of interest, provide a very important approach to temperature measurement in many of these difficult circumstances. Most phosphors have characteristic emissions that are affected by temperature, since the phosphor molecular structures are directly correlated to vibrations and rotations associated with temperature. Some phosphors, depending on their molecular structure, tend to have emission bands in the visible and infrared which can

be sensed by standard photodetectors of many types. Certain temperature ranges for particular phosphors will show very strong temperature dependence; consequently, monitoring the fluorescence in these ranges can produce a very sensitive measurement of that phosphor's temperature. When the temperature of the phosphor layer is determined, the temperature of the component to which it is bonded can be inferred. The layers required are often no greater than about 50 μm , so they interfere minimally in most situations.

Phosphors that exhibit significant temperature sensitivity in various temperature ranges, and which are reasonably stable chemically, are often called thermographic phosphors (TPs). These TPs will have emissions that can be measured in various ways to determine temperature. The most common method has been to measure the fluorescent intensity as a function of time and extract the characteristic lifetime of the emission. That lifetime, for TPs in particular, is often a strong function of temperature. Another approach has been to select two or more emission wavelength bands, where the intensity of one or more is strongly dependent on temperature and at least one other for which it is not, and to use the ratio of their intensities as a normalized function of temperature. In other specialized situations it is also possible to simply use fluorescence intensity to determine temperature. In yet other circumstances the emission wavelength shift, which can be temperature dependent, can be monitored. Certain emission bands, too, can vary in width as a function of temperature; those variations in width can be used for temperature measurement.

A number of energy sources can be used to stimulate the fluorescence of a TP. The most common source used to date is ultraviolet radiation from a laser, laser diode, or light-emitting diode (LED). The stimulated fluorescence, then, is less energetic, therefore of longer wavelength, typically in the visible region of the electromagnetic spectrum. Other sources, such as X-rays or other high-energy electromagnetic emissions, electrons, protons, and the like will also activate TPs, as indeed, for certain materials, can acoustic impact or related mechanical processes.

In all cases, it is important to recognize that the material measured is the phosphor itself not the surface beneath it. That limitation, however, can be used to advantage when two or more TP layers are used, perhaps with interstitial material layers. In such arrangements heat flux, for example, can be determined by measuring the two or more temperature differences and applying the heat transfer parameters of the interstitial layers. Relatedly, the wear of a surface can be monitored by measuring the emissions from that thinning surface, or the efficiency of a thermal barrier coating can be determined by its fluorescent properties or the emissions of TPs mixed with it or on its surface. TPs in powder form, with particle diameters down to a fraction of micrometer, can also be injected into moving fluids, liquid, or gas, and, when monitored, indicate velocity, temperature, or fluid density. In short, TPs can be used in a variety of ways to make measurements of temperature and related physical properties, often in situations where other methods to make such measurements would be futile or severely limited.

62.2 ADVANTAGES OF PHOSPHOR THERMOMETRY

The use of TPs for temperature measurement has several major advantages over more standard temperature measurement methods. Some of these advantages will be made clear in the following sections, but it is useful to list a number of them here. (i) TP response has no dependence on surface properties such as emissivity or reflectance. The fluorescent characteristic measured is only associated with the molecular conditions of the phosphor layer. (ii) There are TPs covering a vast temperature range, from cryogenic systems up to systems near 2000 K. (iii) Calibrated TP systems do not drift over time or require any kind of reference measurement. (iv) Most TPs are chemically stable and have low electrical and heat conductivity, so their presence on a surface is minimally perturbing. (v) TP emission characteristics have fast time responses, typically on the order of microseconds, so measurement systems with rapid time dependence are straightforward to produce. (vi) TPs can be distributed in very thin layers over significant areas; consequently, they have the promise of effective use in two-dimension temperature measurement.

62.3 THEORY AND BACKGROUND

Many compounds have fluorescent properties, but one important class of these is rare-earth oxides and similar structures that have a small addition of a different rare-earth ion distributed through the molecular lattice. This small addition is called an activator or dopant. Its purpose is to make the fluorescence more likely upon absorption of a stimulating energy source. These are particularly true with rare-earth compounds because the activator ion is relatively isolated from others of its type in the molecular lattice, leaving it with fewer quantum decay routes upon excitation. Many of those routes are photon emitting, hence the fluorescent emission. Rare-earth metals are chemically unique in that the chemical valence shell of the atom is lower in electronic energy than the closed electronic shell above it. Consequently, all rare-earth compounds have very similar chemical behaviors and tend to be stable up to high temperatures. Some refractory materials, such as Y_2O_3 and other similar compounds, can be made into effective TPs by appropriate small additions of other metal ions, such as Eu, Tb, Dy, and others. The use of high-temperature materials that can be applied as powders in binders or sprayed directly on surfaces by other means has made the effective use of TPs for many applications possible.

In this section we will restrict our mathematical analysis to rare-earth phosphors activated by Eu ions, both because they are common and effective and because they are relatively simple in their molecular behavior. For example, there are many europium-doped phosphor compounds in which the lifetime of the fluorescence from certain emission lines is dependent on temperature. These include $\text{Y}_2\text{O}_3:\text{Eu}$, $\text{La}_2\text{O}_2\text{S}:\text{Eu}$, $\text{LaPO}_4:\text{Eu}$, and $\text{LuPO}_4:\text{Eu}$.

There are a number of mechanisms responsible for phosphor temperature responses in general. Given here is an explanation that often applies to Eu-activated phosphors. The Eu ion replaces a small fraction of the dominant rare-earth ions in the lattice. That relative isolation allows the Eu ion to excite to higher energy levels without significant competition from vibrational states. When those levels deexcite the probability of photon emission is high, thereby making the phosphor more efficient.

Now, we will select one of the very common TPs, especially useful near room temperature, $\text{La}_2\text{O}_2\text{S}:\text{Eu}$, to illustrate the theory of temperature-dependent behavior among TPs. Further, we will consider only the fluorescent lifetime method of extracting temperature. This method is not only the most commonly used but also helps clarify the atomic behavior within the molecular structures involved.

An energy level diagram for $\text{La}_2\text{O}_2\text{S}:\text{Eu}$ is shown in Figure 62.1. In Figure 62.1a, for clarity, the potential energy diagrams for the lowest ground and excited electronic states are shown. The abscissa corresponds to the position of the activator in the lattice. It will undergo various allowed vibrations. One might picture the europium atom connected by springs to its neighboring oxygen and sulfur atoms and oscillating through the equilibrium center with increasing amplitude as temperature is increased. The horizontal lines within the potential wells illustrate that the vibrational energy is quantized. The higher the temperature, the more vibration states are occupied. The vertical line, arrow #1, corresponds to the energy of an ultraviolet photon, ν_{uv} . Upon

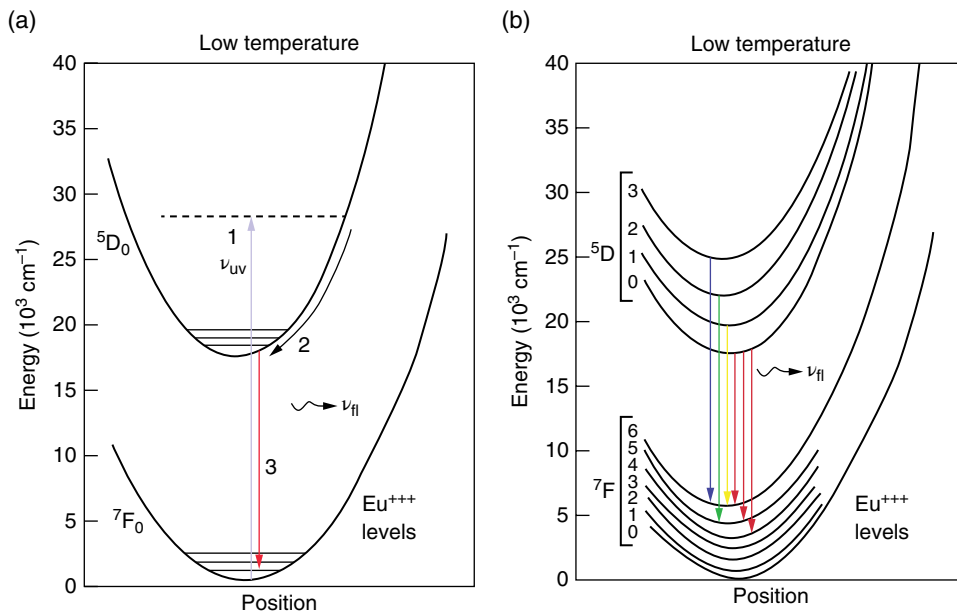


FIGURE 62.1 Energy level diagram of europium phosphor (a) lowest ground and excited electronic states and (b) multiple levels of the $5D$ and $7F$ states with fluorescence transitions indicated.

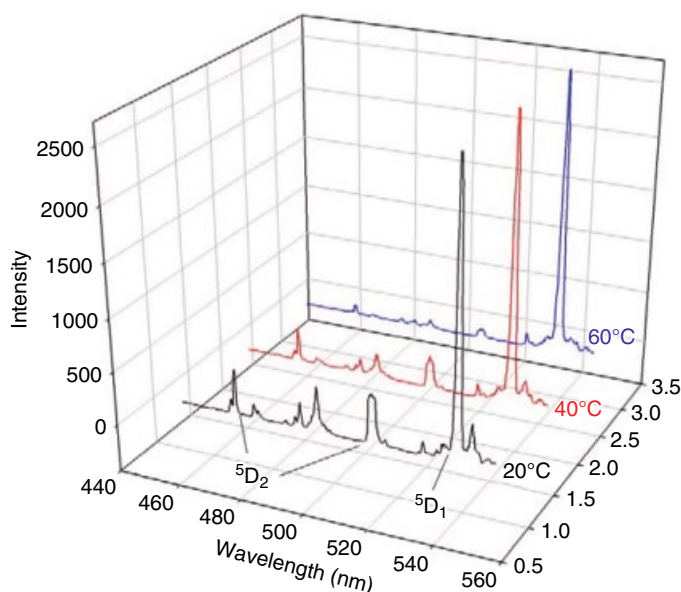


FIGURE 62.2 Fluorescent spectrum of $\text{La}_2\text{O}_2\text{S}:\text{Eu}$.

absorption by the host or dopant, almost immediately (i.e., in $< 10^{-9}$ s), excess energy is given to the lattice vibrations following path 2 as seen. The energy is redistributed among the vibration levels as governed by a Boltzmann temperature. At low temperatures, the potential energy in the excited electronic state has only one path for returning to the ground state: to emit the energy in the form of a fluorescence photon, ν_{fl} . The picture is a little more complicated in that, as seen in Figure 62.1b, there are several ground electronic states, denoted by $^7\text{F}_j$, as well as several $^5\text{D}_i$ excited electronic states that lie sequentially higher in energy. Various fluorescence transitions are possible, as denoted by the arrows, the corresponding transitions emit from red to blue with increasing energy. They occur with differing probabilities and at different energies (or wavelengths). A scan of the fluorescence intensity with wavelength is called a fluorescence spectrum. A portion is shown in Figure 62.2 at three different temperatures. The $^5\text{D}_2$ states change in going from 20 to 60°C. The $^5\text{D}_1$ state is unchanged.

Another level of complexity has been added to the energy level diagram for $\text{La}_2\text{O}_2\text{S}:\text{Eu}$ in Figure 62.3, which depicts a state termed the “charge transfer (CT) state.” The CT state explanation, first put forward by Fonger and Struck [7], involves the transfer of an electron from a neighboring atom to the europium. It is indicated qualitatively in the right-hand figure such that only the lower vibrational levels of any excited electronic state may be occupied at low temperatures. At low temperatures, following excitation, the $^5\text{D}_i$ state is populated and fluorescence is essentially the only deexcitation pathway. However, at sufficiently high temperatures, the

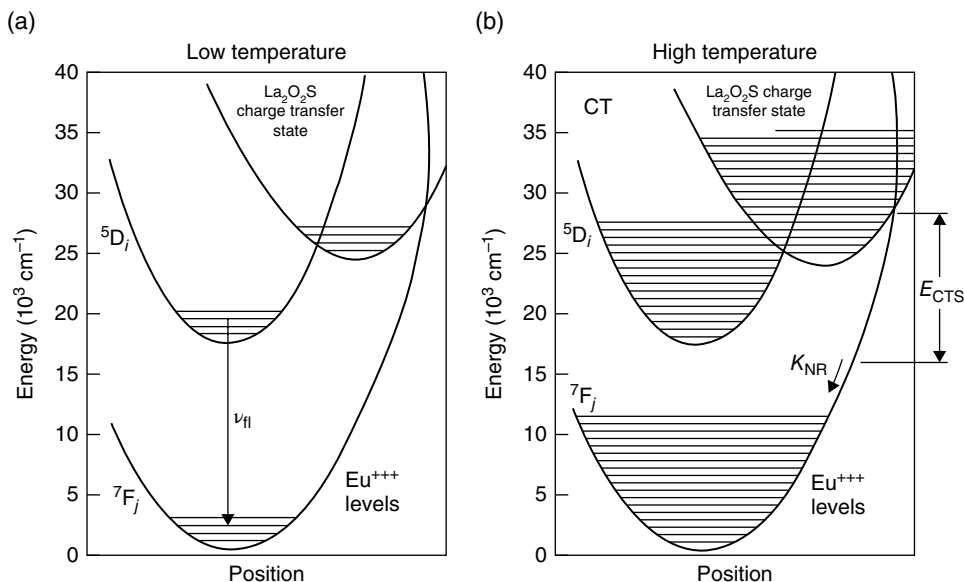


FIGURE 62.3 Energy level diagram for (a) low temperature and (b) high temperature with the charge transfer state.

excited state vibrational levels overlap with those of the CT state. The CT state provides another path for deexciting the phosphor. In this case, cross over to the CT state can occur. The CT state crosses the ground state potential energy curve at E_{CTS} . When this happens, energy is transferred to the host lattice through vibrations (or some energy can leak into a lower $^5\text{D}_i$ state) and not via fluorescence. The higher the temperature, the more likely and the faster this occurs. This depopulation of the electronic state decreases the lifetime and the overall intensity of the emission. This process is called “quenching.”

A simple energy level diagram in Figure 62.4 serves as the basis for obtaining a simple rate equation in TPs of this type, illustrating (i) that the fluorescence is characterized by single exponential decay and (ii) the temperature dependence of this decay. It is assumed that the state is excited instantaneously with a population of N_0 . The change in number of excited states, dN , is proportional to the number of excited states N and the elapsed time dt such that

$$dN \propto -Ndt \quad (62.1)$$

This model assumes that there is no significant feeding from any other electronic level. A constant of proportionality, κ , is the total decay rate. Therefore,

$$\frac{dN}{N} = -\kappa dt \text{ and } N = N_0 e^{-\kappa t} \quad (62.2)$$

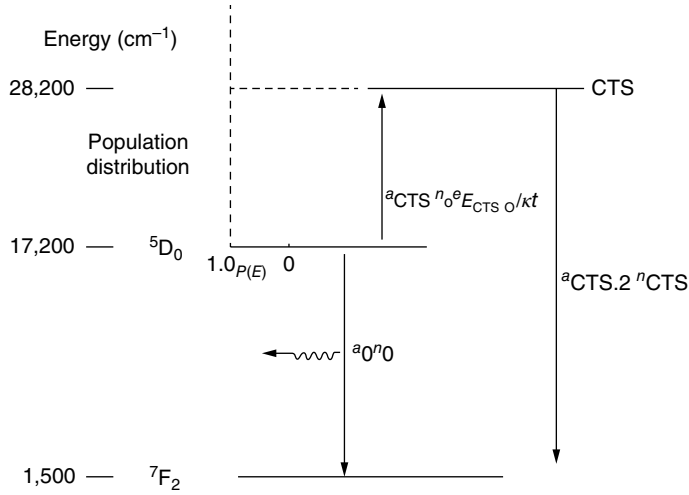


FIGURE 62.4 Simplified energy level diagram for a typical Eu-doped phosphor with Boltzmann distribution, $P(E)$ based on Struck-Fonger Model.

and the decay is exponential. κ is the sum of a radiative component, κ_{rad} , and a nonradiative component, κ_{nonrad} , where energy is lost, in this case, to the CT state. κ_{rad} is the familiar Einstein coefficient. The expression is

$$\kappa = \frac{1}{\tau} = \kappa_{\text{rad}} + \kappa_{\text{nonrad}} = \frac{1}{\tau_0} + \kappa_{\text{nonrad}} \quad (62.3)$$

where τ is the measured lifetime and τ_0 is the low temperature value of the lifetime before the quenching temperature has been reached. To ascertain the temperature dependence of this non-radiative quenching factor, the relative distribution of vibrational states must be considered. Boltzmann's law, a fundamental law of thermodynamics, shows that the population distribution follows an exponential dependence where, for a given level N_i , the ground state population is N_i ,

$$N_i = N_i e^{-(E_i/kT)}. \quad (62.4)$$

Given this functional dependence, the relative population at the vibrational level, whose energy is E_{CTS} , is obtained by substituting into the Boltzmann equation. Figure 62.4 illustrates this distribution qualitatively at low and high temperatures. At sufficiently high temperatures, governed by $e^{-(E_{\text{CTS}}/kT)}$, a significant population exists at an energy E_{CTS} . κ_{nonrad} is therefore proportional to this exponential factor. The constant of proportionality is A , a rate constant typically on the order of 10^{10} or 10^{11} transitions per second. It is related physically to the period of vibration in the lattice and the time it takes for the electron to physically move from the nonmetal ion to the europium ion. We now have

$$\kappa = \frac{1}{\tau} = \frac{1}{\tau_0} + A e^{-(E_{\text{CTS}}/kT)} \quad (62.5)$$

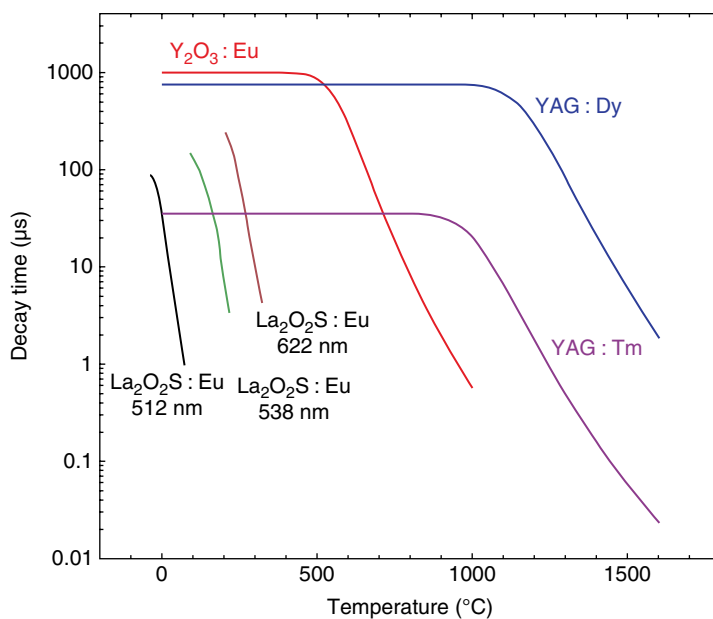


FIGURE 62.5 Fluorescent lifetime versus temperature for various phosphor materials.

or

$$\tau = \frac{\tau_0}{\left(1 + \tau_0 A e^{-(E_{CTS}/kT)}\right)}. \quad (62.6)$$

As a final result, this expression describes the temperature dependence of fluorescence lifetime for this model.

In order for the lifetime measurements to be a practical indicator of temperature, it is desirable for the lifetime to have a linear or a simple logarithmic dependence on temperature. A semilogarithmic plot of models based on experimental data for lifetime τ versus temperature, T , for several TPs is shown in Figure 62.5. To select $\text{Y}_2\text{O}_3:\text{Eu}$, for example, note that in the temperature range from ambient to about 600°C , the logarithm of lifetime is nearly constant with temperature. The logarithmic response permits calibration of the technique with a simple linear relationship over that temperature range.

Returning to our consideration of $\text{La}_2\text{O}_2\text{S}:\text{Eu}$, the functional dependence, for three of the characteristic electronic transition bands in that phosphor, of the lifetime on temperature, illustrated in Figure 62.6, is in agreement with that predicted by Equation 62.6. This agreement and the linear dependence with temperature over the observed range of temperatures can be understood from a simple analysis of Equation 62.6 and

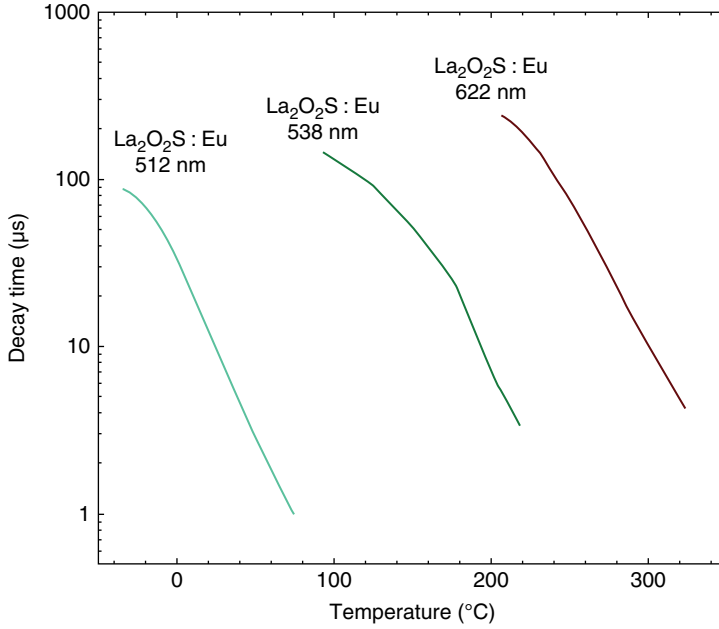


FIGURE 62.6 Decay time (lifetime) versus temperature for three characteristic electronic transition bands in $\text{La}_2\text{O}_2\text{S}:\text{Eu}$.

the following considerations. If the natural logarithms of both sides of Equation 62.6 are taken, Equation 62.6 becomes

$$\ln(\tau) = \ln(\tau_0) - \ln\left(1 + \tau_0 A e^{-(E_{\text{CTS}}/kT)}\right). \quad (62.7)$$

Different cases represent three temperature ranges and can help explain that this equation models the experimental data in Figure 62.7 and that the logarithms of the lifetimes have a linear dependence on temperature over a restricted range of temperatures. These three ranges are (i) the range for small temperatures, $E_{\text{CTS}}/kT \gg 1$; (ii) the range for high temperatures, $E_{\text{CTS}}/kT \ll 1$; and (iii) the midrange temperatures where the response is useful for making measurements, $E_{\text{CTS}}/kT \cong 1$. These limiting cases are discussed in this order below.

CASE 1 $E_{\text{CTS}}/kT \gg 1$

For the case of small temperatures, $-(E_{\text{CTS}}/kT)$ is a large negative number so that $e^{-(E_{\text{CTS}}/kT)}$ is a number near zero. In the limit as the temperature, T , approaches zero, $\tau_0 A e^{-(E_{\text{CTS}}/kT)}$ becomes negligible compared to 1, and the value of $\ln(1 + \tau_0 A e^{-(E_{\text{CTS}}/kT)})$ approaches zero since $\ln(1) = 0$. Therefore, for the range of small values of temperature, $\ln(\tau) = \ln(\tau_0)$ and the lifetime is a constant, τ_0 , as is observed in Figure 62.7.

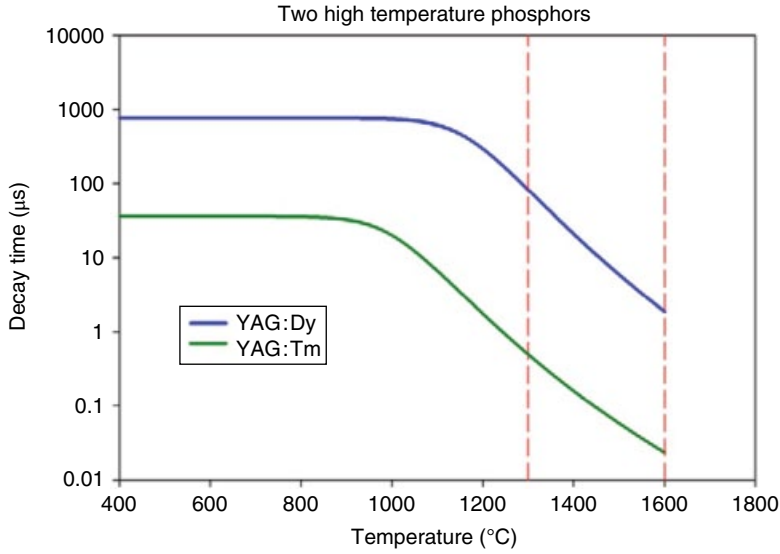


FIGURE 62.7 Temperature dependence of YAG:Dy and YAG:Tm.

CASE 2 $E_{\text{CTS}}/kT < 1$

In the high temperature range, E_{CTS}/kT becomes a small number approaching zero and $e^{-(E_{\text{CTS}}/kT)}$ approaches 1 so that Equation 62.6 gives $\tau = \tau_0/(1 + \tau_0 A)$, a constant value.

CASE 3 $E_{\text{CTS}}/kT \cong 1$

Since the lifetimes can decrease over four decades, this implies $\tau_0 A$ to be large compared to 1 and that there is a range of values for T in which $\tau_0 A e^{-(E_{\text{CTS}}/kT)}$ is greater than 1. In this range, Equation 62.7 can be approximated by

$$\ln(\tau) = \ln(\tau_0) - \ln(\tau_0 A e^{-(E_{\text{CTS}}/kT)}). \quad (62.8)$$

or

$$\ln(\tau) = \ln(\tau_0) - \ln(\tau_0) - \ln(A) + \left(\frac{E_{\text{CTS}}}{kT}\right). \quad (62.9)$$

This equation is of the form

$$\ln(\tau) = B + \left(\frac{E_{\text{CTS}}}{kT}\right). \quad (62.10)$$

The linear relationship observed in Figure 62.7 only holds for small changes in temperature, ΔT , about some large temperature T_0 . Since the temperature is in degrees

Kelvin, T_0 may be on the order of 1150 K and ΔT may vary ± 250 K, as is the case in the example shown in Figure 62.7. As a result, the temperature T can be expressed as $T = T_0 + \Delta T$, and Equation 62.10 may be expanded to yield

$$\ln(\tau) = B + \frac{E_{\text{CTS}}}{k(T_0 + \Delta T)} = B + \frac{E_{\text{CTS}}}{k}(T_0 + \Delta T)^{-1} \quad (62.11)$$

or

$$\ln(\tau) = B + \frac{E_{\text{CTS}}}{kT_0} \left(1 + \frac{\Delta T}{T_0} \right)^{-1}. \quad (62.12)$$

$$\ln(\tau) = B + \frac{E_{\text{CTS}}}{kT_0} \left(1 - \frac{\Delta T}{T_0} \right). \quad (62.13)$$

$$\ln(\tau) = B' + C' \Delta T. \quad (62.14)$$

where $B' = B + (E_{\text{CTS}}/kT_0)$ and $C' = E_{\text{CTS}}/kT_0^2$. This illustrates that the logarithm of the lifetime decreases linearly with small changes in temperature about some moderately large value of temperature.

62.4 LABORATORY CALIBRATION OF TP SYSTEMS

The correlation of TP emissions with temperature requires calibration measurements that associate particular emission properties with the temperature of the TP. This calibration is required since the TP molecular configurations are far too complex in their quantum behavior to allow a theoretical calculation of correspondence with adequate accuracy. However, because these molecular systems, once calibrated, continue to respond with statistical consistency as long as the molecular configuration is not compromised, TP measurement systems require no recalibration or signal drift analysis.

A schematic depiction of a typical TP measurement arrangement is shown in Figure 62.8 (Schematic of a typical TP measurement system). There are three basic components: (i) an interrogation source (UV laser) to stimulate the fluorescence, (ii) a detection system to sense the fluorescence, and (iii) a data analysis system to convert the detector signal to temperature and to estimate accuracy and precision.

Often, TP methods are important to use for very-high-temperature systems. One TP that has been studied for high-temperature use is $\text{Y}_3\text{Al}_5\text{O}_{12}:\text{Dy}$ (YAG:Dy) [8]. Its characteristic emission spectrum from ultraviolet stimulation is shown in Figure 62.9. Note the emission bands between 400 and 600 nm; these are the bands typically used for high temperature correlation.

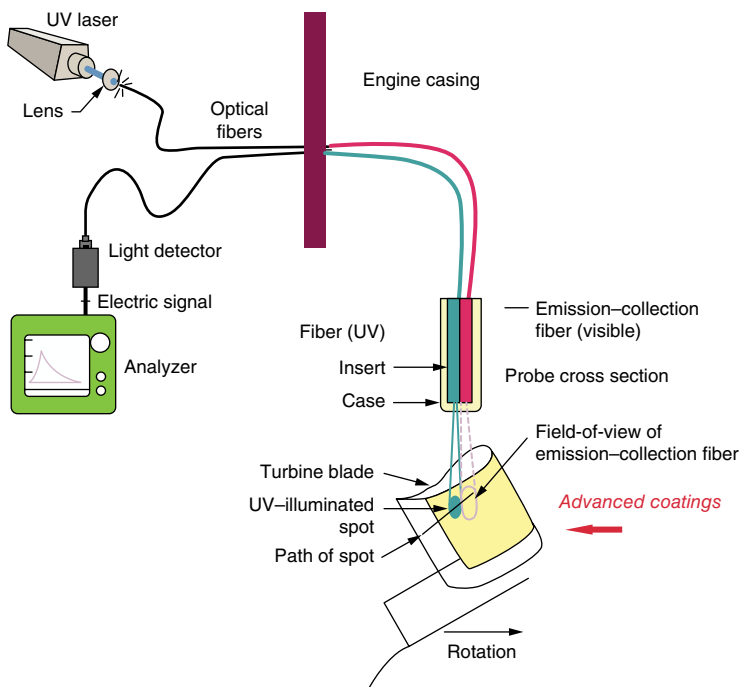


FIGURE 62.8 Schematic of a typical TP measurement system.

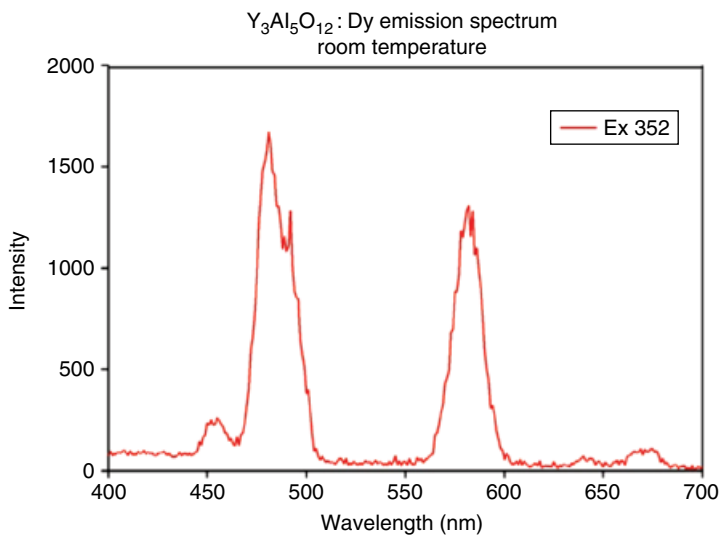


FIGURE 62.9 YAG:Dy emission spectrum.

Figure 62.10 is shown a calibration arrangement to determine the temperature correspondence of the emission of YAG:Dy, a phosphor with excellent high-temperature response. In this arrangement the tripled energy of a YAG laser was used, producing an interrogation wavelength of 355 nm. This beam was optically routed into

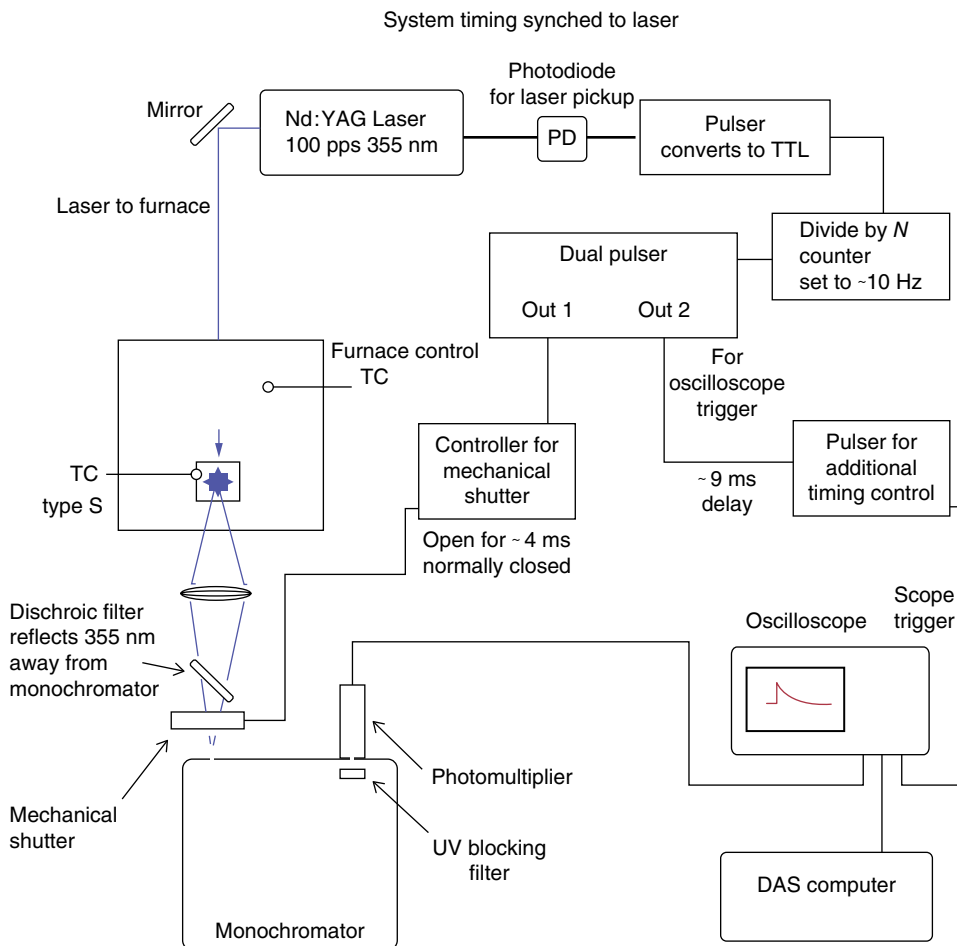


FIGURE 62.10 High-temperature calibration arrangement.

a high-temperature oven, which contained the phosphor sample coated on a ceramic surface. The fluorescent signals were separated optically from the interrogation source, passed through a narrowband filter to collect the appropriate temperature-sensitive emission, and measured by a photomultiplier tube with high gain.

The calibration was done by varying the oven temperature and measuring the phosphor sample temperature with a thermocouple whose junction was in contact with the phosphor surface. Those temperatures were then linked to fluorescent lifetimes, which were measured by repeated laser pulses and signal averaging. Although the blackbody emission background of the sample was minimized by the optical arrangement, those background levels had to be subtracted from the fluorescent signal before the lifetime was determined. With this formulation of YAG:Dy thus calibrated, it could then be applied to various surfaces and measured, yielding the temperature of those surfaces. The a fit of the data for YAG:Dy and YAG:Tm is shown in Figure 62.7.



FIGURE 62.11 EMCO thermographic phosphor LabKit.

Other calibration arrangements are much simpler and often use less intense interrogation sources such as LEDs or laser diodes. This more complex arrangement, however, illustrates the feasibility of developing TP systems for very high-temperature measurement in adverse conditions.

For calibrations near room temperature, and to clearly illustrate the PT method, EMCO has developed a Phosphor LabKit containing all the measurement components and including the basic software required to convert fluorescent signals to temperature. Figure 62.11 is a photograph of the LabKit. Figure 62.12 shows how the signal changes versus temperature. This plot shows the decay only. Each curve is the result of 128 averages. Further information can be obtained by contacting EMCO directly.

62.5 HISTORY OF PHOSPHOR THERMOMETRY

While it has been known for many decades [1] that fluorescent materials have temperature dependence, one of the first major examples of its use in practical non-contact applications was developed in the late 1980s at Oak Ridge, Tennessee's Gaseous Diffusion Plant, K-25, to measure the temperature of rotors spinning at high speed to separate isotopes of uranium. Other uses of the technology were quickly added, and many further development studies and measurements involving TPs were done.

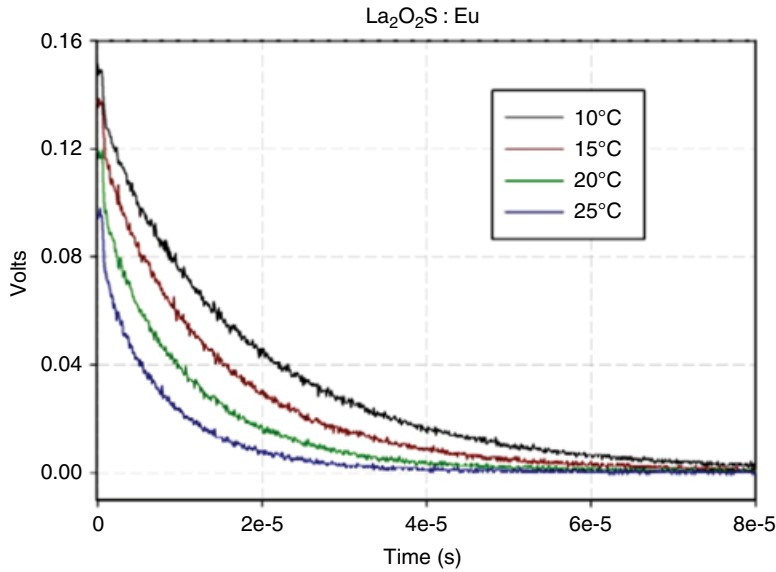


FIGURE 62.12 LED-excited fluorescence versus temperature.

62.6 REPRESENTATIVE MEASUREMENT APPLICATIONS

In this section we will discuss two specific measurement applications to illustrate the utility of TPs in various scenarios where standard temperature measurement approaches would be ineffective or extremely difficult and prone to error. The first is the use of optical fibers and TPs to measure the temperature of a spinning high-speed motor with permanent magnets in the rotor. The second is a temperature study of operating jet turbine components in a measurement test stand.

62.6.1 Permanent Magnet Rotor Measurement

Permanent magnet motors are especially useful for generating high rotational speeds and for being amenable to digital control systems that vary the motor operating parameters. Figure 62.13 is a drawing of the rotor assembly and the optical fiber arrangement used to provide access to the region between the spinning rotor containing permanent magnets and stator below. The underside of the rotor was coated with La₂O₂S:Eu mixed into an epoxy binder. The radial band of phosphor allowed temperature to be measured all around the rotor during operation. Two optical fibers were placed side-by-side, one transporting short pulses from a 337 nm nitrogen gas laser, the other transporting the resultant fluorescence to a photomultiplier filtered to select the appropriate emission band. In this case the filter passed light at about 510 ± 10 nm. The gap between rotor and stator was only about 2 mm, but it was adequate for the fiber optical arrangement.

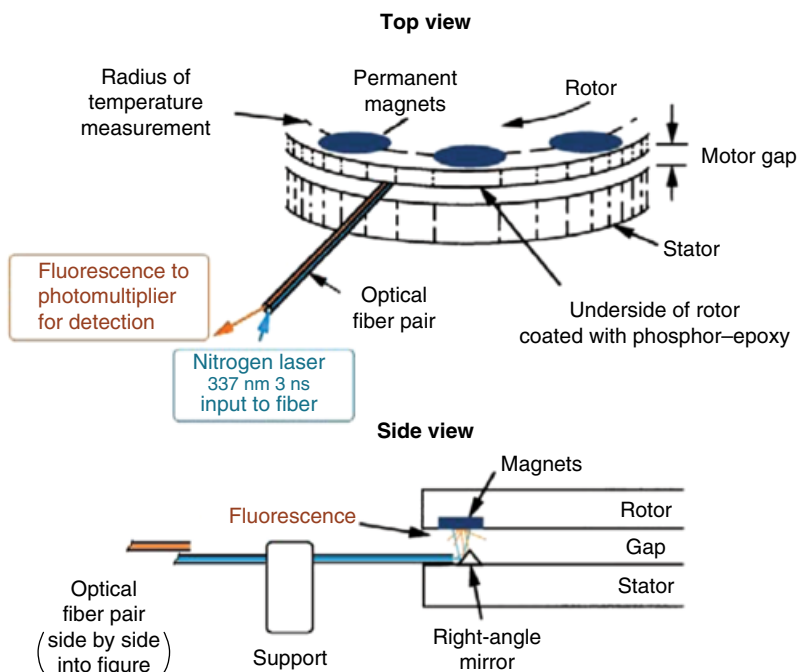


FIGURE 62.13 Permanent magnet motor experimental arrangement.

To make the measurement it was necessary to pulse the laser several hundred times at the same spot on the rotor in order to accurately determine the temperature of that spot. Consequently, the laser was triggered by a circuit that monitored the rotational speed by reflectance from a reference position, the axial position determined by the fraction of the time between reference pulses. That position was determined by a delay time built into the triggering circuit, and when that delay time was changed the entire rotor surface could be measured in small axial steps. Figure 62.14 shows some typical data for the measurement, data revealing the amount of heating of the permanent magnets for particular operating parameters. The two temperature traces were obtained at different times for the same magnet and shows the influence of eddy current heating.

62.6.2 Turbine Engine Component Measurement

An important TP application has been for measurement of temperature in the hot zones of turbine engines and generators. These zones reach temperatures as high as 1300°C and, are difficult to access, and full of reflected light from the combustion process.

Tables 62.1 and 62.2 show a summary of applications of phosphor thermometry over the history of its use pursuant to turbine and other high-temperature applications. The tables are not intended to be exhaustive rather to be an illustration of the many potential uses of the method.

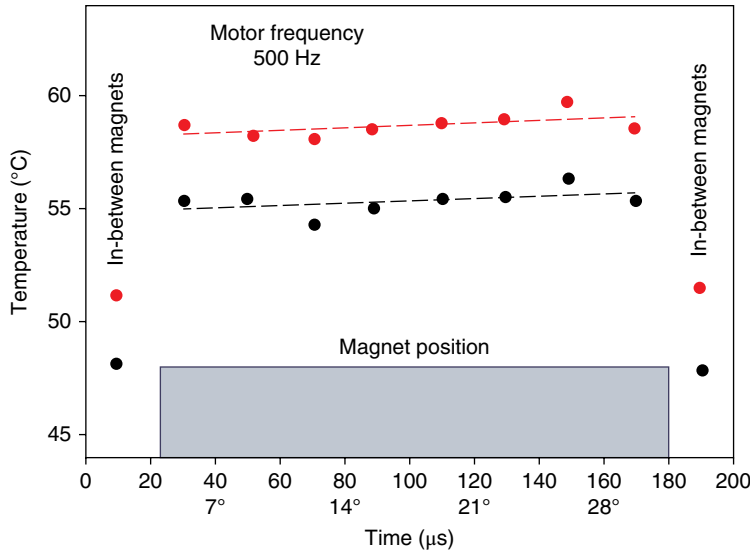


FIGURE 62.14 Heating of magnets in motor.

TABLE 62.1 Various Measurements

Lab Measurements with Flame/Combustion		High Temperatures (Stationary/Slow Surfaces)		High-Speed Surfaces	
Burning wood [9]	840°F	Many tests	To 3100°F	Permanent magnet motor [10]	750 Hz 200°F
Intumescent surfaces [11]	1100°F	Galvanneal steel processing [12]	1300°F	Rail gun armatures [13]	400 m/s 200°F
Pistons intake valves [14]	400°F				

TABLE 62.2 Combustion-Related Measurements

High Temperature and with Combustion	
Burner rig (slow speed rotating) [15]	1650°F
Turbine engine vane [16]	1750°F
Turbine engine afterburner nozzle [17]	1300°F
Turbine engine afterburner flame holder [18]	1100°F

62.7 TWO-DIMENSIONAL AND TIME-DEPENDENT TEMPERATURE MEASUREMENT

TPs can be also be used for complex temperature measurements, such as area temperature mapping and monitoring the temperature of a zone as a function of time. For area measurement the ratio method is often recommended. This approach uses two

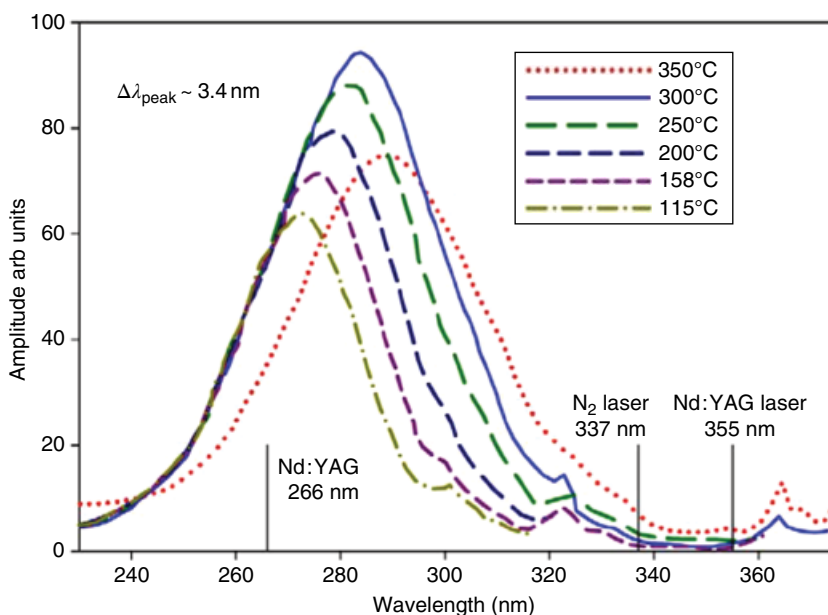


FIGURE 62.15 Excitation band of $\text{Y}_2\text{O}_3:\text{Eu}$ from 115 to 350°C.

wavelengths of fluorescence measured with a video system, with one image divided by the other. Such arrangements can be calibrated ahead of time and can often use broad-band filtering to allow significant light for each image. To increase the measurement sensitivity, image intensifiers can be used for some situations. Another approach is to scan the surface of interest with a one-dimensional system, building the image over multiple scans.

For time-dependent measurements, too, there are various approaches. We discuss a few of them here. A straightforward one is simply to measure fluorescent lifetime as quickly as possible, averaging the measurements in at least one of two ways: extracting temperature from each pulse or maintaining a running average over selected time periods. In any case, the one-sigma limit of time resolution, sometimes called the Sartori limit, is approximately three times the average lifetime measured (Dr. Walter Sartori, personal communication, ca. 1982). Time-dependent temperatures can also be measured by monitoring the intensity of emissions associated with a ratio measurement. The ratio, in this case, would vary as a function of time, with its time resolution associated with the characteristic lifetime of the wavelengths measured. Another approach would take advantage of the shift of an excitation spectrum for a particular phosphor as a function of temperature. Figure 62.15 shows the excitation spectra for $\text{Y}_2\text{O}_3:\text{Eu}$ at various temperatures for emission in one of strong emission bands. As temperature changes, the strength of the excitation changes. Thus, by measuring a spectrum or the intensity change at a given wavelength, a measurement of temperature versus time can be extracted.

62.8 CONCLUSION

In this article we've discussed the use of fluorescent materials to measure temperature. We have included a theoretical description of the method for certain typical types of TPs and have presented several illustrations to make clear the utility of this powerful physical measurement technology.

REFERENCES

1. S. W. Allison and G. T. Gillies, "Remote thermometry with thermographic phosphors: Instrumentation and applications," *Review of Scientific Instruments*, 68, 2615–2650 (1997).
2. C. Knappe, J. Lindén, F. Abou Nada, M. Richter, and M. Aldén, "Investigation and compensation of the nonlinear response in photomultiplier tubes for quantitative single-shot measurements," *Review of Scientific Instruments*, 83, 034901 (2012).
3. A. H. Khalid and K. Kontis, "Thermographic phosphors for high temperature measurements: Principles, current state of the art and recent applications," *Sensors*, 8, 5673–5744 (2008).
4. M. D. Chambers and D. R. Clarke, "Doped oxides for high-temperature luminescence and lifetime thermometry," *Annual Review of Materials Research*, 39, 325–359 (2009).
5. M. Aldén, A. Omrane, M. Richter, and G. Sarner, "Thermographic phosphors for thermometry: A survey of combustion applications," *Progress in Energy and Combustion Science*, 37, 422–461 (2011).
6. J. Brübach, C. Pflitsch, A. Dreizler, and B. Atakan, "On surface temperature measurements with thermographic phosphors: A review," *Progress in Energy and Combustion Science*, 39, 37–60 (2013).
7. W. H. Fonger and C. W. Struck, " $\text{Eu}^{+3} \text{ } ^5\text{D}$ resonance quenching to the charge-transfer states in $\text{Y}_2\text{O}_3\text{S}$, $\text{La}_2\text{O}_3\text{S}$, and LaOCl ," *Journal of Chemical Physics*, 52(12), 6365 (1970).
8. M. R. Cates, S. W. Allison, S. L. Jaiswal, and D. L. Beshears, "YAG:Dy and YAG:Tm Fluorescence to 1700°C," Proceedings of the 49th International Instrumentation Symposium of the ISA (The International Society of Instrumentation, Systems, and Automation), ISA Vol. 443, Orlando, FL, May 4–8, 2003.
9. A. Omrane, *Thermometry Using Laser-Induced Emission from Thermographic Phosphors: Development and Applications in Combustion*, PhD Dissertation, Lund University, Lund, Sweden, 2005.
10. S. W. Allison, G. T. Gillies, M. R. Cates, and B. W. Noel, "Method for monitoring permanent magnet motor heating with thermographic phosphors," *IEEE Transactions on Instrumentation and Measurement*, 37(4), 637–641 (1988).
11. A. Omrane, Y. C. Wang, U. Göransson, G. Holmstedt, and M. Aldén, "Intumescent coating surface temperature measurement in a cone calorimeter using laser-induced phosphorescence," *Fire Safety Journal*, 42(1), 68–74 (2007).
12. W. W. Manges, S. W. Allison, and J. R. Veach, "Galvanneal Thermometry with a Thermographic Phosphor System," 1997 AISE Annual Convention and Iron and Steel Exposition, Cleveland, OH, September 29–October 2, 1997.

13. S. W. Allison, M. R. Cates, S. M. Goedeke, A. Akerman, M. T. Crawford, S. B. Ferraro, J. Stewart, and D. Surls, "In-flight armature diagnostics," *IEEE Transactions on Magnetics*, 43(1), 329–333 (2007). (Part II of two parts of Selected Papers from the 13th International Symposium on Electromagnetic Launch (EML) Technology, Berlin, Germany, May 22–25, 2006.)
14. J. S. Armfield, N. Domingo, J. M. Storey, S. W. Allison, D. L. Beshears, and M. R. Cates, "Powertrain Component Temperature Measurements via Phosphor Thermometry," Presented at the World Car Conference, Ref. No. 97WCC018, Riverside, CA, January 19–22, 1997.
15. K. W. Tobin, S. W. Allison, M. R. Cates, G. J. Capps, D. L. Beshears, and M. Cyr, "Remote High-Temperature Thermometry of Rotating Test Blades Using $\text{YVO}_4:\text{Eu}$ and $\text{Y}_2\text{O}_3:\text{Eu}$ Thermographic Phosphors," AIAA/ASME/SAE/ASEE Proceedings of the 24th Joint Propulsion Conference, AIAA-88-3147, Boston, MA, July 11–13, 1988.
16. B. W. Noel, W. D. Turley, and S. W. Allison, "Thermographic-Phosphor Temperature Measurements: Commercial and Defense-Related Applications," Proceedings of the 40th International Instrumentation Symposium of the ISA, Baltimore, MD, May 1–5, 1994.
17. H. Seyfried, G. Särner, A. Omrane, M. Richter, H. Schmidt, and M. Aldén, "Optical Diagnostics for Characterization of a Full-Size Fighter-Jet Afterburner," Proceedings of the ASME Turbo Expo, Vol. 1, pp. 813–819. ASME Turbo Expo 2005—Gas Turbine Technology: Focus for the Future, Reno-Tahoe, NV: ASME Press.
18. H. Seyfried, M. Richter, M. Aldén, and H. Schmidt, "Laser-induced phosphorescence for surface thermometry in the afterburner of an aircraft engine," *AIAA Journal*, 45(12), 2966–2971 (2007).

VOLTAGE AND CURRENT TRANSDUCERS FOR POWER SYSTEMS

CARLO MUSCAS AND NICOLA LOCCI

Department of Electrical and Electronic Engineering, University of Cagliari, Cagliari, Italy

63.1 INTRODUCTION

Control, management, monitoring, and protection applications on electrical power systems require continuous and trustworthy knowledge of the two fundamental electrical quantities, namely, current and voltage. By considering the paramount importance of these actions, in terms of both safety and economic impacts, when designing a measurement system for such applications, many metrological aspects should be considered carefully so that the overall performance of the system guarantees that the desired parameters are measured with sufficient accuracy.

It is widely acknowledged that, owing to different causes, including faults, power quality (PQ) disturbances, operation of network components, etc., voltages and currents in modern power grids do not usually meet the ideal conditions of rated frequency, rated amplitude, sinusoidal waveform, and positive symmetry in three-phase systems.

All these electrical phenomena, which may have characteristics quite different from each other in terms of amplitude, duration, repetition, and effects on the system elements, as well as perception from users, are nowadays measured by means of digital programmable systems, be they intended for control, management, monitoring, or protection purposes. Digital measurement systems typically receive as input only voltages in the range of a few volts or at most a few tens of volts. Therefore, voltage

and current transducers are needed to convert the electric quantities involved in power grids, either currents or higher voltages, into low voltages suitable for the successive measurement steps.

The main functions of the measurement transducers used in electric power systems can be summarized as follows:

- Adapting the level and/or nature of the primary quantity to a value and/or nature that can be dealt with by measurement instrumentation or protection relays
- Ensuring safety of measurement and protection systems, by either insulating such systems from the power grid or using alternative solutions

In power plants, the most commonly used transducers for voltages and currents are the magnetic core instrument transformers, the current transformer (CT) and the voltage transformer (VT), sometimes referred to as potential transformer (PT). Figure 63.1 shows an example of insertion in a single-phase line operating at voltage V_1 where the current I_1 flows. One terminal of the secondary circuit is usually connected to ground for safety reasons.

Commercial VTs and CTs are usually designed to operate with 50 or 60 Hz (depending on the rated frequency of the system) sinusoidal primary quantities, and their accuracy specifications are defined with respect to these rated conditions. In the presence of either distorted waveforms, and thus of harmonics and interharmonics, or transient events, their performance may significantly decay.

Furthermore, it should be considered that, for historical reasons, the rated output value of these transducers is generally 100 V for the VT and 5 A (or 1 A) for the CT, in order to have sufficient energy to drive electromechanical measurement or protection devices.

On the other hand, modern digital measurement systems and electronic relays have low input power requirement. This leads to the need to introduce low-power current transformers (LPCTs) and low-power voltage transformers (LPVTs), which supply a low-voltage signal (from tens of millivolts to a few volts) as secondary output.

To comply with the above requirements, there are on the market voltage and current transducers, for either laboratory or field applications, based on different principles of operation, that can guarantee good accuracy and a wide measuring range, as well as

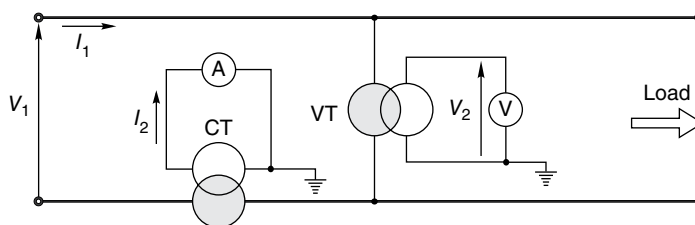


FIGURE 63.1 Insertion of instrument transformers.

a bandwidth extended from a few hertz (or even DC, in some cases) up to the megahertz region. Furthermore, the kind and level of their output quantity are generally compatible with the input of data acquisition systems, and their weight and size are usually lower than for the conventional instrument transformers.

By considering that transducers are often the main source of uncertainty in the entire measurement chain, this chapter is intended to provide the reader with the basic elements to understand the principle of operation and the behavior of the most common transducers for voltage and current in power systems, to assess their metrological characteristics, and finally, to choose the most appropriate devices to be used in the different measurement applications.

Note: In the technical literature about measurement systems, the terms *transducer* and *sensor* are often used as synonymous to define a device that transforms an input physical quantity into a different quantity, with different values and/or different characteristics, which are more suitable for the measurement instrument/system connected at the output of the device itself. In this chapter it has been chosen, only for the sake of clarity and without loss of generality, to use the term transducer for the complete device, which includes all the auxiliary components (power supply, amplifiers, filters, etc.) required to make practically feasible the conversion, while the term sensor will be used sometimes to indicate more specifically the physical “sensing” element, that is, where the physical principle of the conversion acts.

63.2 CHARACTERIZATION OF VOLTAGE AND CURRENT TRANSDUCERS

Choosing the right transducer is one of the most important steps in the design of the entire measurement system. Such choice can be done properly only if the performance of the devices available on the market is known. It is therefore important that manufacturers provide suitable information about the metrological behavior of these devices.

International standards on instrument transformers (e.g., the IEC 61869 series, which will be considered as a reference in the following) provide, for the sinusoidal conditions, the following definitions (in brackets the symbols employed in this chapter for the defined quantities):

- **Actual transformation ratio (K):** the ratio of the actual primary quantity to the actual secondary quantity.
- **Rated transformation ratio (K_r):** the ratio of the rated primary quantity to the rated secondary quantity.
- **Ratio error (η):** the error that a transformer introduces into the measurement of a voltage/current and that arises from the fact that the actual transformation ratio is not equal to the rated transformation ratio.

- **Phase displacement** (ϵ): The difference in phase between the primary and secondary current vectors, the direction of the vectors being so chosen that the angle is zero for a perfect transformer. It is usually expressed in minutes or centiradians ($1 \text{ crad} = 10^{-2} \text{ rad}$).

Under sinusoidal conditions, if we define \mathbf{Q} as the phasor of a generic quantity and use the subscripts in and out for transducer's input and output quantities, respectively, it is

$$\mathbf{Q}_{\text{out}} = \mathbf{Q}_{\text{in}} \frac{1}{K_r} (1 + \eta) \cdot e^{j\epsilon} \quad (63.1)$$

By considering the modulus of both members in the earlier equation, we obtain

$$\eta = \frac{K_r Q_{\text{out}} - Q_{\text{in}}}{Q_{\text{in}}} = \frac{K_r Q_{\text{out}} - K Q_{\text{out}}}{K Q_{\text{out}}} = \frac{K_r - K}{K} \cong \frac{K_r - K}{K_r} \quad (63.2)$$

where Q_{in} and Q_{out} are the root-mean-square (rms) values of the input and output quantities of the transducer. η and ϵ are defined in steady-state sinusoidal conditions, and their value depends on both the frequency and the amplitude of the voltage (or current) on the primary side.

When measuring the distorted quantities existing in modern power systems, other characteristics of the transducers have an importance that is equal to, or sometimes greater than, the previous parameters. The most significant of these characteristics are linearity and bandwidth.

The importance of linearity, which, in this context, mainly represents the ability to ensure sufficient accuracy in a wide measuring range, is evident since it is desirable, especially in protection applications, that the transducer performance does not change even for large variations of the input quantity.

On the other hand, it is clear, by taking into account the characteristics of the distorted signals to be measured, that it is of great interest defining the frequency range for which the uncertainty stays below prefixed limits. To this purpose, the transducer bandwidth is usually defined as the frequency range where the transformation ratio does not differ from its rated value of more than 3 dB (i.e., $\sim 30\%$).

63.3 INSTRUMENT TRANSFORMERS

63.3.1 Theoretical Fundamentals and Characteristics

Instrument transformers (VTs or PTs and CTs) are extensively used in production, transmission, distribution, and utilization of electricity, in association with measuring instruments, meters, and protective or control devices.

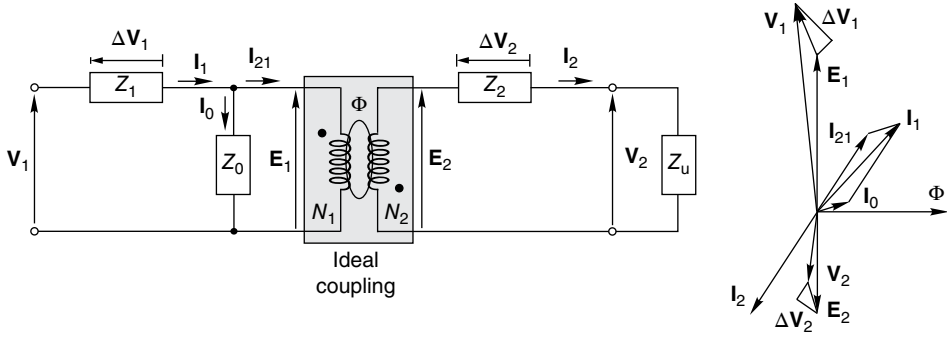


FIGURE 63.2 Equivalent circuit and phasor diagram of a transformer.

Voltage and current instrument transformers consist of a magnetic core on which two windings are wound. The primary and secondary windings have N_1 and N_2 turns, respectively, and are electrically insulated from the core and from each other. Alternatively, the primary circuit of a current transformer can simply consist of the single conductor where the primary current flows. The secondary circuit is loaded by the secondary burden determined by measurement and/or protection devices.

Instrument transformers are characterized by their rated ratio, that is, the ratio between rated input and output quantities:

$$\text{CT: } K_r = \frac{I_{1r}}{I_{2r}} \quad \text{VT: } K_r = \frac{V_{1r}}{V_{2r}} \quad (63.3)$$

As an example, the rated ratio could be $K_r = 200:5$ (A/A) for a CT and $K_r = 20,000:100$ (V/V) for a VT. An ideal instrument transformer should reduce the amplitude of the input signal according to the rated ratio and let the phases unchanged. Actually, this does not happen, for several reasons, which can be shortly explained by recalling the equivalent circuit of a transformer, shown in Figure 63.2 along with the phasor diagram of the involved electrical quantities.

The behavior of the ideal transformer is summarized by the following relationships between vectors:

$$\mathbf{I}_{21} = -\frac{N_2}{N_1} \mathbf{I}_2 \quad \mathbf{E}_1 = -\frac{N_1}{N_2} \mathbf{E}_2 \quad (63.4)$$

It is common practice to define a further equivalent circuit of the transformer (Fig. 63.3), where all the quantities, including the ones related to the secondary circuit, such as the impedances of the secondary winding and of the load, are reported to the primary one, by means of the following expressions:

$$\bar{Z}_{21} = \bar{Z}_2 \left(\frac{N_1}{N_2} \right)^2 \quad \mathbf{V}_{21} = -\mathbf{V}_2 \frac{N_1}{N_2} \quad (63.5)$$

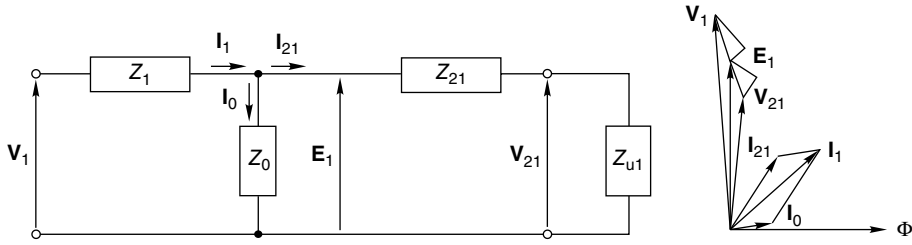


FIGURE 63.3 Equivalent circuit of the transformer referred to the primary side.

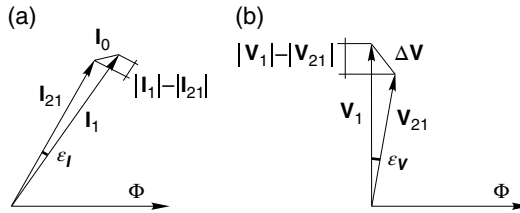


FIGURE 63.4 Phasor diagrams for CTs (a) and VTs (b).

In this way, no ideal transformer is included in the model, and all the quantities refer to the primary voltage V_1 .

In this circuit a T-network is present, which is responsible for the different behavior of the real transformer with respect to the ideal one. In order to study these aspects, Figure 63.4 shows the phasor diagrams for CTs (Fig. 63.4a) and VTs (Fig. 63.4b), respectively, while Equation 63.6 shows the relevant mathematical relations:

$$\begin{aligned} \text{CT: } \mathbf{I}_1 &= \mathbf{I}_{21} + \mathbf{I}_0 \cong \mathbf{I}_{21} = -K_h \mathbf{I}_2 & \left(K_h = \frac{N_2}{N_1} > 1 \right) \\ \text{VT: } \mathbf{V}_1 &= \mathbf{V}_{21} + \Delta \mathbf{V} \cong \mathbf{V}_{21} = -K_v \mathbf{V}_2 & \left(K_v = \frac{N_1}{N_2} > 1 \right) \end{aligned} \quad (63.6)$$

According to (63.6), an approximated value for the amplitude of the primary quantities (I_1 or V_1) can be achieved by multiplying the amplitude of the secondary ones (I_2 or V_2) by the turn ratio (K_h or K_v).

For a CT with a turn ratio $K_h = N_2/N_1$, the vector difference between the currents \mathbf{I}_1 and \mathbf{I}_{21} corresponds to the exciting current \mathbf{I}_0 flowing in the magnetizing branch (see Fig. 63.4a). This difference is thus minimized if the measured current is much larger than the exciting current. To obtain this, besides a proper sizing of the magnetic circuit, the transformer should work close to short circuit conditions, or in other words, the load on the secondary circuit should be as low as possible.

In any case, even in the presence of a null secondary burden, the unavoidable presence of the exciting current would make it impossible to have the actual ratio K equal to the theoretical one K_t . For this reason, in the common praxis, the rated

transformation ratio attributed to the CT is larger than the ratio between the numbers of turns (i.e., $K_r > N_2/N_1$) so that the natural ratio error is compensated at least partially. In such way, once the current I_2 has been measured and multiplied for the rated constant K_r , the obtained value is a better approximation of the actual primary current I_1 . In the practice, once the desired rated transformation ratio for a CT has been established, the device is built with a turn ratio slightly smaller.

Analogously, for a VT with a turn ratio $K_v = N_1/N_2$, the vector difference between the voltages V_1 and V_2 corresponds to the voltage drop ΔV across the series branches of the equivalent circuit (see Fig. 63.4b). To limit this voltage drop, besides a proper design of the windings, in order to have low resistances and low leakage reactance, a very high impedance should be connected at the secondary side of the VT. However, even assuming an ideal condition of null secondary burden (open circuit), the unavoidable exciting current causes a voltage drop on the primary side, thus affecting both the amplitude and the phase of the actual primary voltage. The presence of a nonzero burden implies additional voltage drops on both sides of the transformer, thus making the actual ratio more different from K_v . This also implies that the actual ratio K varies for varying burden. The rated transformation ratio assigned to the VT, K_{vr} , is therefore slightly larger than N_1/N_2 so that the voltage drop in the series branches is at least partially compensated.

Clearly, in both cases, this compensation, whose effects on the ratio error vary with the operative conditions, is definitely ineffective with respect to phase displacement.

The standards about instrument transformers usually define the accuracy of these devices in terms of precision class, which represents the maximum value allowed (in percent) for the ratio error defined in Section 63.2. For each class, the standards also define limits for the phase error. As an example, for a class 0.3 VT, according to standard IEC 61869-3, the maximum ratio error is $\eta = \pm 0.5\%$ and the maximum phase error is $\varepsilon = \pm 0.6$ crad. These maximum errors should be ensured for voltages included in the range 80–120% of the rated value (see Fig. 63.5), when the burden is between 25 and 100% of the rated value.

Similar specifications exist for CTs, but in this case different error limits are defined for different values of the input current to take into account the high variability of the current absorbed by both domestic and industrial loads. As an example, IEC 61869-2 considers a current range between 5 and 120% of the rated value (Fig. 63.5).

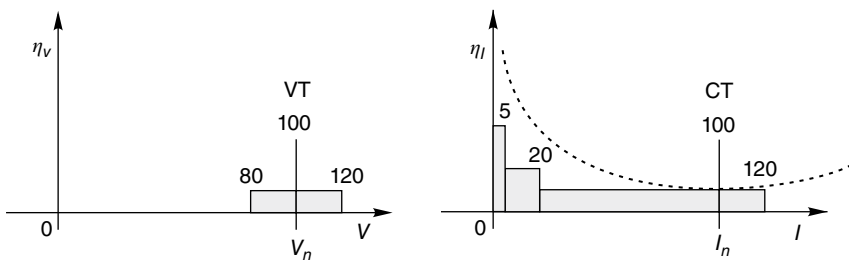


FIGURE 63.5 Error limits for VTs and CTs.

Ratio and phase errors in instrument transformers also vary when the load on the secondary side varies. From this point of view, a limit value for the secondary burden is defined in terms of the maximum apparent power that can be requested to the secondary circuit when the rated quantity is applied on the primary side. As an example, according to IEC 61869-3, for measuring VTs the preferred burden ranges are from 1 to 10 VA (with power factor of 1) or from 10 to 100 VA (with power factor of 0.8, lagging), while IEC 61869-2 specifies for measuring CTs a burden range from 2.5 to 30 VA.

As far as safety is concerned, if the secondary circuit of a CT was opened, all the primary current would become an exciting current (see Fig. 63.3), and the voltage between the terminals at the primary side could reach high values so that a risk for the operator or for the device itself could arise. For this reason, an overvoltage protection may be included in CTs. For dual reasons, overcurrent protection may be instead used in VTs to avoid the problems that could arise if the secondary terminals were connected to a low impedance.

63.3.2 Instrument Transformers for Protective Purposes

When instrument transformers are used with protection relays, their working conditions can be very different from the normal operation. Thus, their metrological characteristics should include some additional specifications.

In the case of VTs, the given accuracy should be ensured for voltage values much higher (typically from 1.2 to 1.9) than the rated one.

The case of current transformers should be analyzed more carefully. Indeed, one of the most common protections in electric grids is the one against overcurrent (arising from either short circuit or overload), usually performed through circuit breakers, whose operation is based on relays supplied by CTs. Thus, the correct and prompt intervention of the protection systems firstly depends on the behavior of the CT. During faults, the transient current is composed of a steady-state component superimposed to a DC decaying component, whose value depends on both the circuit parameters and the instant in which the faults begins. When this fault current is much higher (up to tens of times) than the CT rated current, then saturation occurs and the secondary current may be heavily distorted.

Figure 63.6 shows, for two qualitative examples, the primary current and the secondary one multiplied by the rated transformation ratio: in case (b) the saturation is higher than in case (a). These highly distorted waveforms of the secondary current may represent an issue for the correct operation of the protection relays.

In these situations the usual vector representation cannot be adopted. Thus, besides ratio and phase errors, a new error term can be introduced in the standards for protective CTs (e.g., IEC 61869-2), named the composite error (ε_c), which is usually provided in percent and defined according to the following expression:

$$\varepsilon_c = \frac{100}{I_1} \sqrt{\frac{1}{T} \int_0^T (K_r i_2 - i_1)^2 dt} \quad (\%) \quad (63.7)$$

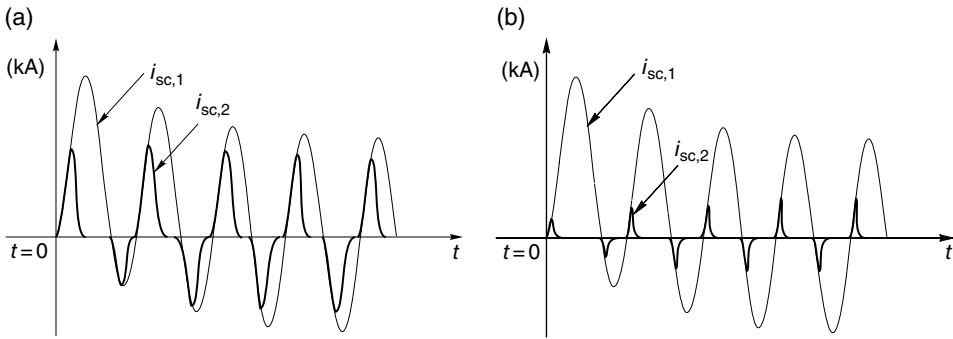


FIGURE 63.6 Examples of possible effects of saturation on the secondary current of a CT: in case (b) the saturation is higher than in case (a).

where i_1 and i_2 are the instantaneous values of the primary and secondary current, respectively, while T is the period of these quantities.

For instance, in a protective CT with a precision class 10P, the maximum allowed composite error is 10%, for an input maximum current that is much higher (typically from 5 to 30 times) than the rated current.

63.3.3 Instrument Transformers under Nonsinusoidal Conditions

In Section 63.1 it has been recalled that voltages and currents in power systems may be characterized by a waveform that is different from the sinusoidal one traditionally considered for such systems, due to both possible steady-state disturbances (e.g., harmonics or interharmonics) and events, such as voltage dips, rapid transients, etc.

Measurement transducers used to measure these quantities must guarantee sufficient accuracy while reproducing the distorted input waveform. As an example, when harmonic and interharmonic components are concerned, it is necessary to accurately transduce the amplitude of each harmonic but also, when power related terms have to be evaluated, its phase.

Magnetic core instrument transformers usually have a limited bandwidth. In order to explain this behavior, let us consider the equivalent circuit of Figure 63.7, which differs from the classical representation of a transformer at low frequency for the presence of the stray capacitances: C_1 and C_{21} refer to the capacitive coupling in the turns of the primary and secondary windings, respectively, while C_{ps} represents the capacitive coupling between the two windings.

At industrial frequency these capacitances are negligible, but their influence increases for increasing frequency. This leads to the possibility of creating resonant circuits, which would result in significant decay of the instrument accuracy.

This problem is particularly evident in instrument transformers for medium- and high-voltage systems. Indeed, in such systems the insulation requirements are stronger and impose the use of larger sizes, thus determining larger stray capacitances and leakage inductances to appear and causing the risk of resonance at lower frequencies.

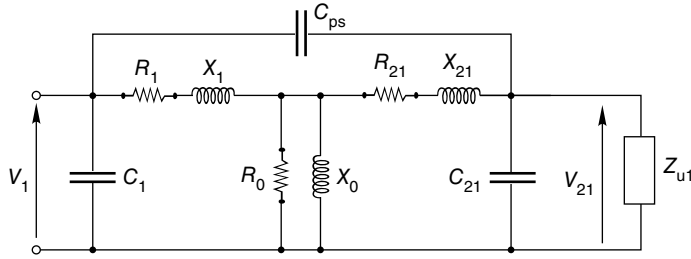


FIGURE 63.7 Equivalent circuit of the transformer for higher frequencies.

As a consequence, if no solutions are adopted to compensate such effects, instrument transformers (especially for medium- and high-voltage systems) are not suitable to measure voltages and currents with significant high-frequency components.

Generally, low-voltage CTs and VTs have a ratio error less than a few percent and phase displacement less than a few degrees for frequency up to a few thousand hertz. For medium-voltage VTs the same errors occur at frequencies lower than 1 kHz, while in high-voltage VTs this happens at about 500 Hz. CTs, owing to their different characteristics, have slightly better performance.

In any case, magnetic core transformers are not appropriate to measure quantities with DC components. Indeed, such components are not transferred to the secondary side, and in addition, they may lead to core saturation, thus affecting significantly also the measurement of alternate components.

63.3.4 Capacitive Voltage Transformer

Insulation issues impose the use of sufficient distances between the elements of a VT. When these devices are used for high voltages, insulation requirements determine large sizes and weights, thus implying difficulties in building them and producing significant impact on the costs, which can be predominant with respect to the accuracy needs. For the preceding reasons, in power systems operated at voltages higher than 150 kV, VTs are replaced by capacitive voltage transformers (CVTs). According to Figure 63.8, in such devices a capacitive divider (composed by capacitances C_1 and C_2) reduces the primary voltage V_{in} to an intermediate level $V_{out,C}$, which is then applied to a transformer that, besides providing the required insulation, further reduces the voltage to the desired secondary value V_{out} . The Thevenin equivalent circuit for the capacitive divider has the equivalent voltage $V_{eq} = V_{out,C} = V_{in} \cdot C_1 / (C_1 + C_2)$ and the equivalent series capacitance $C_{eq} = (C_1 + C_2)$. The voltage drop in the capacitive element C_{eq} can be compensated by an opposite voltage drop across a reactive inductance (totally or partially contained in the transformer), whose value makes the circuit resonant for the working frequency. Since these optimal compensation conditions are verified only for the rated frequency, the nominal accuracy of the CVT is ensured only for frequency variations in a very limited range (e.g., about ± 0.5 Hz) around the rated

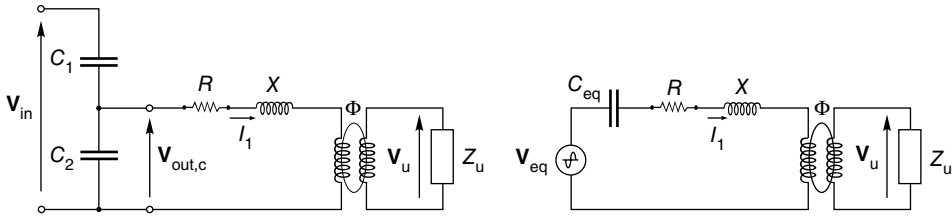


FIGURE 63.8 Capacitive voltage transformer and its equivalent circuit.

value. Out of this range, the lack of the resonance condition between inductance and capacitance may introduce significant errors. As a further consequence of this consideration, in the presence of distorted voltages, the errors in measuring the harmonic components could become unacceptable.

63.4 TRANSDUCERS BASED ON PASSIVE COMPONENTS

Using passive elements (resistors and capacitors) is one of the oldest and most reliable methods to convert currents into voltages and to reduce the amplitude of a primary voltage to a value suitable for measurement instrumentation. These sensors are generally characterized by low cost, easy use, and good accuracy. On the other hand, a major drawback in the use of such devices is that they do not guarantee the insulation between primary and secondary circuits. Thus, alternative solutions are needed to ensure safety of the measurement system.

63.4.1 Shunts

Resistive shunts are based on the Ohm's law to convert the current $i(t)$ flowing into the resistance R_s into the voltage $v(t)$ across its terminals: $v(t) = R_s i(t)$.

On the other hand, this linear relationship is only ideal, whereas in the reality it is affected by a number of influence factors, like variability of the parameters with time, signal frequency, environmental conditions, in particular the temperature, etc.

Shunts can be built with many different techniques, including wound conductors, coaxial resistors, thick film and thin film resistors, metallic plates, etc. Each one of these typologies privileges one or more aspects (robustness, long-term stability, circuit miniaturization, thermal exchange, etc.) over the others, and thus the choice of the most suitable shunt should be done by taking into account the requirements of the specific application.

All resistive shunts have limited bandwidth, owing to the effects of parasitic inductances and capacitances. The inductive reactance is the largest problem in the design of these devices, especially for low resistances. Indeed, when the frequency of the signal components increases, a more appropriate circuitual model of the resistive shunt includes at least a series (undesired) inductance, as shown in Figure 63.9.

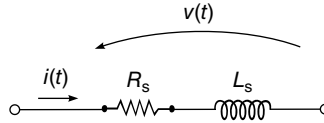


FIGURE 63.9 Equivalent model of a shunt resistor with residual inductance.

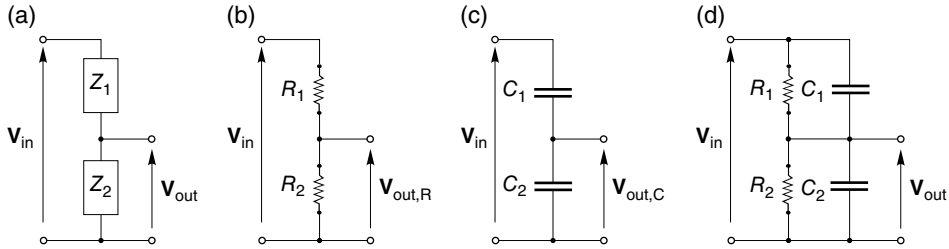


FIGURE 63.10 Voltage dividers: (a) general scheme; (b) resistive divider; (c) capacitive divider; (d) RC divider.

Thus, the relation between current and voltage is actually $v(t) = R_s i(t) + L di(t)/dt$. The relative error caused by the presence of the inductance is higher for higher frequency and for lower value of the resistance.

63.4.2 Voltage Dividers

A voltage divider reduces the voltage applied to its primary terminals of a prefixed ratio. The simplest divider consists of the series of two impedances \bar{Z}_1 and \bar{Z}_2 (Fig. 63.10a). Under sinusoidal conditions and with no load connected on the output terminals, it is

$$\mathbf{V}_{\text{out}} = \mathbf{V}_{\text{in}} \frac{\bar{Z}_2}{\bar{Z}_1 + \bar{Z}_2} = \frac{1}{\bar{K}_r} \mathbf{V}_{\text{in}} \quad (63.8)$$

The divider is said to be compensated when the ratio K_r is a real number independent of the signal frequency. This can be obtained by using either impedances of the same kind, for example, two resistances or two capacitances, as shown in Figure 63.10b and c, respectively, or series/parallel combinations of resistances and capacitances, properly chosen so that the two impedances \bar{Z}_1 and \bar{Z}_2 have the same time constant (in Fig. 63.10d: $\tau_1 = R_1 C_1 = \tau_2 = R_2 C_2$). Inductive components are generally not used for these purposes.

In practical cases, the secondary terminals of the divider are always loaded by a burden, which affects the validity of Equation 63.8. The higher the load impedance, with respect to the equivalent divider's impedance, the lower the influence of the load itself on the transducer behavior.

The divider's components are never pure elements but contain parasitic elements, such as internal or link inductances, capacitive couplings between parts of the device or between the device and neighborhood objects, series or leakage resistances in the

capacitors, etc. These elements make the behavior of the transducer dependent on frequency. For this reason, often resistive and capacitive dividers contain suitable compensation circuits.

The main advantage of voltage dividers, besides the reduced size and weight, is their good linearity. Thus, the same device can be used in wide voltage ranges, for instance, from tens to hundreds kilovolt, thus allowing proper measurements under both normal operating conditions and faults. This makes these devices suitable to be used in MV and HV systems, instead of magnetic VTs.

Compensated dividers can guarantee very high bandwidth (from DC to tens of megahertz).

On the other hand, they have the same drawback seen for the shunts, since they do not guarantee insulation between high-voltage and low-voltage terminals.

However, often the impedance of the elements that compose the divider (which, for instance, in some commercial devices for medium-voltage systems is in the order of $10^8 \Omega$) represents by itself a sufficient level of insulation, thus making voltage dividers the most common alternative to VTs in MV systems. Of course, in this case, for safety reasons, one of the divider's terminals should be connected to ground, and this means that floating voltages, or line-to-line voltages in three-phase systems, cannot be measured with this solution.

In some other cases, the safety of the measurement or protection system may be ensured by either introducing additional devices (e.g., the isolation amplifiers that will be presented in the next section) or implementing alternative solutions, like the ones based on spark gaps or surge arresters.

63.4.3 Isolation Amplifiers

Isolation amplifiers provide the electric insulation between input and output, by means of either magnetic or optical coupling. Therefore, they can be combined with either the shunts or the voltage dividers discussed earlier to realize complete current and voltage transducers, including insulation between primary and secondary circuit. Of course, the overall metrological characteristics of these transducers (accuracy, linearity, bandwidth) are significantly affected by the presence of this new element.

Figure 63.11 shows, as an example, the scheme of an isolation amplifier based on magnetic coupling. Two distinct circuital areas can be noticed, for input and output, respectively. A third section provides power supply for both input and output circuits. All these three sections are mutually insulated through magnetic couplings.

The input signal (which can represent either the voltage drop across the terminals of a current shunt or the output voltage of a voltage divider) is applied to the input buffer.

The useful signal travels across the insulation barrier by means of modulation and demodulation technique. The modulator translates the original baseband signal spectrum to high frequencies. The modulated signal reaches the secondary side by means of the magnetic coupling and is then demodulated into its original baseband. Finally, the output voltage is provided by means of a second buffer. The use of

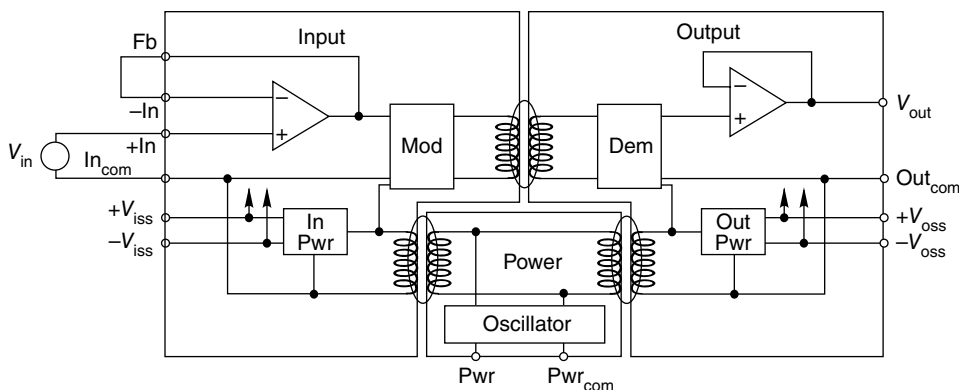


FIGURE 63.11 Isolation amplifier based on magnetic coupling.

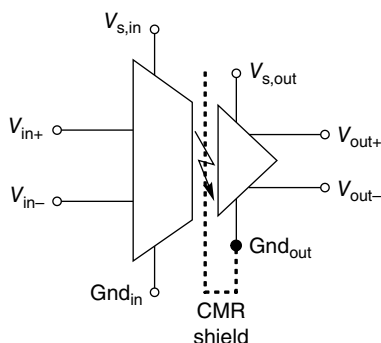


FIGURE 63.12 Isolation amplifier based on optical coupling.

modulation and demodulation allows also DC components to be transferred to the secondary side of the transformer.

A different solution can be implemented by using an optical coupling (Fig. 63.12). In this case, the low voltage at the output of the passive sensor is transformed, by an analog-to-digital converter, into sequence of bits, which is then transmitted across the optical barrier and then, if needed, reconverted in an analog signal.

In many commercial products, the sensing passive element (shunt or divider) and the isolation amplifier are contained in a single device. These transducers are mainly used in low-voltage power electronic applications.

63.5 HALL-EFFECT AND ZERO-FLUX TRANSDUCERS

63.5.1 The Hall Effect

As it is well known, when an electrical current I_p passes through a conducting slab placed in a magnetic field with induction B (Fig. 63.13), a force acts on the charged particles in motion, and consequently, a potential v_H proportional to the current and to

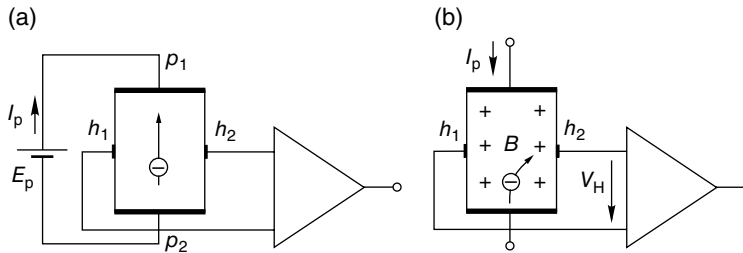


FIGURE 63.13 The Hall effect: path of the electrons without (a) and with (b) magnetic induction.

the magnetic field is developed across the slab in a direction perpendicular to both the current and the magnetic field. This property, known as the Hall effect, can be suitably exploited to measure magnetic induction and related quantities.

The use of Hall-effect transducers is one of the most popular solutions to perform measurements of electric quantities, both in DC and AC, with insulation between input and output. The basic configuration is suitable for current measurements, but as it will be seen, a simple calibrated resistance can allow also voltages to be measured.

63.5.2 Open-Loop Hall-Effect Transducers

In order to explain the behavior of an open-loop Hall-effect current transducer, let us refer to Figure 63.14.

A conductor passes through the hole of a magnetic core, where the Hall sensor is placed. A magnetic field proportional to the current flowing in the conductor is produced around the conductor itself. More in general, a winding composed of a given number of turns may be present in the primary circuit to increase the sensitivity. The lines of the magnetic field concentrate in the core and excite the Hall sensor, giving rise to a voltage proportional to the primary current. This very low voltage is amplified to obtain the voltage at the output terminals.

In the open-loop configuration, accuracy, linearity, and bandwidth of the transducer depend directly on the characteristics of the single components. Furthermore, the saturation of the magnetic core must be avoided in the normal operation of these devices. As a consequence, their metrological performance is usually limited, especially as far as the frequency behavior is concerned. It should be however taken into account that also DC currents can be measured with these devices.

On the other hand, their cost is generally low, and the low power consumption makes them suitable for portable devices where the supply voltage is provided by batteries.

63.5.3 Closed-Loop Hall-Effect Transducers

Closed-loop Hall-effect transducers use feedback technique to improve the performance, by nullifying the magnetic flux in the core. To do this, a secondary winding is added to the components of the open-loop transducer (see Fig. 63.15).

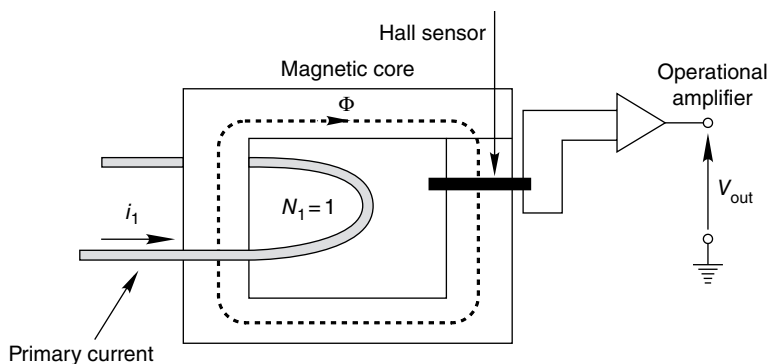


FIGURE 63.14 Open-loop Hall-effect current transducer.

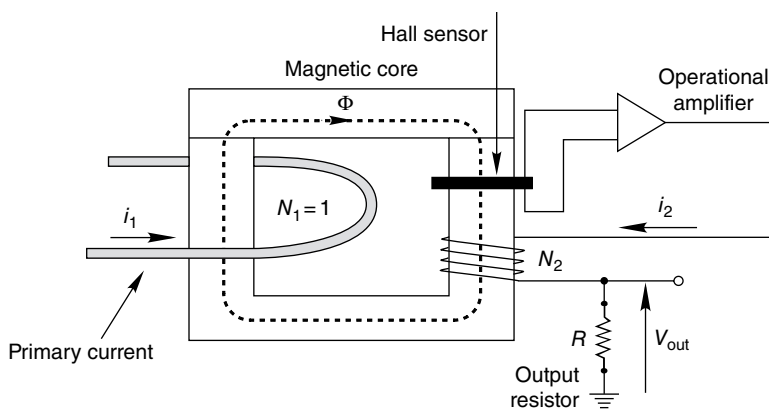


FIGURE 63.15 Closed-loop Hall-effect current transducer.

This compensation winding is so placed that the current flowing in it generates a magnetic field opposite to that produced by the current in the primary conductor. The Hall sensor actually operates as a “zero detector,” insofar as it amplifies the “error signal” (a nonzero flux in the core) to drive the feedback action. The component is therefore designed to have the highest sensitivity. By combining the Hall sensor with a high-gain amplifier, a current is generated in the compensation winding to contrast the magnetic field produced by the primary current. This compensating current can be directly used as output quantity (current-to-current transducer). The transformation ratio is determined by balancing the ampere-turns in the core. As an example, in order to have a 1000:1 ratio in a transducer where the primary current passes directly across the window of the magnetic core ($N_1 = 1$), the secondary circuit will have $N_2 = 1000$ turns. Alternatively, the secondary current can be converted into a proportional voltage through a calibrated resistor (current-to-voltage transducer).

The frequency response can be analyzed by considering two partially overlapped regions (Fig. 63.16). In the first one, from DC to low frequencies, the behavior depends mainly on the electronic operation of the Hall sensor. In the second region the

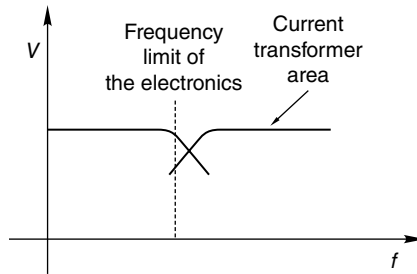


FIGURE 63.16 Bandwidth of a closed-loop Hall-effect current transducer.

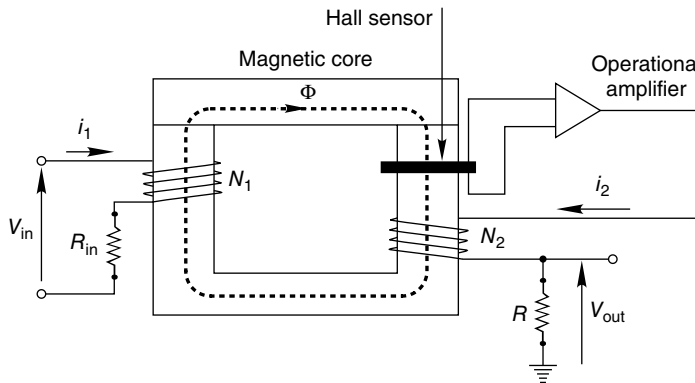


FIGURE 63.17 Hall-effect closed-loop voltage transducer.

compensating winding works practically as the secondary winding of a current transformer, thus extending the frequency range of the device. Obviously, the device should be so designed that the transition between the two regions is gradual to guarantee that the frequency response is sufficiently flat in the overall bandwidth of the transducer (which can be up to hundreds of kilohertz).

As emphasized before, Hall-effect transducers are intrinsically suitable for current measurement. However, voltage transducers can be also built, by adding a calibrated resistance R_{in} across the terminals, which draws a current that can then be measured with a Hall-effect sensor. Figure 63.17 shows the case of a closed-loop voltage transducer. In this case, the additional resistance should be sufficiently low so that, when the measured voltage is applied to its terminals, the current absorbed can be appreciated by the Hall sensor, avoiding sensitivity issues, but also sufficiently high so that the system under test is not subject to excessive loading effects. A trade-off is therefore required.

Note that the voltage drop v_{in} in the primary winding is in phase with the current i_1 , due to resistive part of the input impedance, being the inductive one absent, owing the zero flux in the core.

Hall-effect transducers are usually designed for low-voltage systems (below 1000V), even though devices for medium-voltage systems exist. Their field of application is

mainly in industry automation (variable speed drives, power supplies, filters, etc.), traction systems, and, more in general, in those situations where their ability to measure both DC and AC components can be of fundamental importance. Their use in substations or in switchboards of electric distribution grids is much less frequent.

63.5.4 Zero-Flux Transducers

The principle of operation of the closed-loop Hall-effect transducer can be generalized by employing different systems to detect the flux in the magnetic core and obtaining the so-called electronically compensated current transformers (ECCTs) or zero-flux transducers.

These transducers are conceptually and physically similar to current transformers. Indeed, the sensing element is a magnetic circuit where the flux generated by the measured current is nullified by the flux generated by the secondary current. However, if we take into account that in traditional current transformers the main source of uncertainty is the exciting current needed to magnetize the magnetic core, the accuracy of such transducers can be improved by reducing to zero the flux. This can be achieved by supplying the secondary winding through an amplifier that receives at its input a signal proportional to the core flux. In this way, the voltage across the core's magnetizing impedance is reduced and so is for the exciting current, thus improving the transducer's accuracy.

This approach also eliminates the problems related to the saturation of the magnetic core.

The different typologies based on this principle use different magnetic flux detectors, which can be realized by means of additional windings, Hall sensors (Fig. 63.15), or more complex solutions.

As for the Hall effect, the zero-flux technique can be used also to build voltage transducers by adding a calibrated resistor that absorbs the current to be measured by the ECCT.

Compensated transducers have reduced size, with respect to traditional instrument transformers, bandwidth up to hundreds of kilohertz, and very good accuracy. They are mainly used in special applications or in laboratories, even though on the market some solutions for MV systems exist.

Finally, it should be mentioned that in the last years, several digital compensation techniques for instrument transformers have been introduced. These techniques require a mathematical or experimental model of the transformer to be known. On this basis, once the secondary quantity has been acquired, the primary quantity is known with better accuracy by implementing suitable compensation routines on a microprocessor.

63.6 AIR-CORE CURRENT TRANSDUCERS: ROGOWSKI COILS

The possibility of measuring alternate currents by means of air-core transducers (often referred to as Rogowski coils) has been known since the beginning of the twentieth century. However, such devices are receiving great popularity only in these decades,

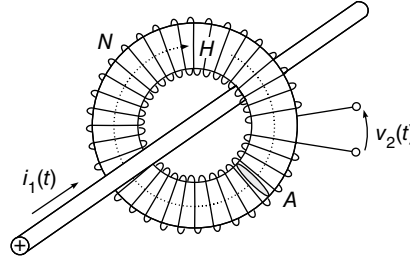


FIGURE 63.18 Rogowski coil.

thanks to the developments of the electronics, which is a necessary complement of these transducers.

Besides ensuring insulation between input and output circuits, Rogowski coils offer wide measurement range (up to hundreds of kiloamperes), wide bandwidth (from a few hertz up to the megahertz region), and excellent linearity.

The Rogowski coil is essentially a coil made of conducting material, uniformly wound around a ring made of a nonferromagnetic material ($\mu \approx \mu_0$), in whose central window the conductor bringing the measured current is placed (see Fig. 63.18).

The current flowing in the conductor produces a magnetic field around it. According to the Ampere's law, the integral of the magnetic field H around a closed path L equals the surrounded current i_1 :

$$\int_L \vec{H} \cdot d\vec{l} = \int_L H \cdot dl \cos \alpha = i_1 \quad (63.9)$$

where α is the angle between the direction of the magnetic field H and the infinitesimal element of length dl .

If n is the number of turns for unity of length, the number of turns in a portion of the coil having length dl is ndl . Assuming that the cross area A of the turns is constant, the magnetic flux concatenated with such portion is

$$d\Phi = \mu_0 H n d l A \cos \alpha \quad (63.10)$$

Therefore, the flux concatenated with the entire coil is

$$\Phi = \int_L d\Phi = \mu_0 n A \int_L H \cos \alpha dl = \mu_0 n A i_1 \quad (63.11)$$

According to the Faraday's law, the voltage $v_2(t)$ at the coil's terminals is given by

$$v_2(t) = -\frac{d\Phi(t)}{dt} = -\mu_0 n A \frac{di_1(t)}{dt} \quad (63.12)$$

An example, if a current $i_1(t)$ with a triangular waveform is considered, as in Figure 63.19, the output of the Rogowski coil is a square wave voltage $v_2(t)$.

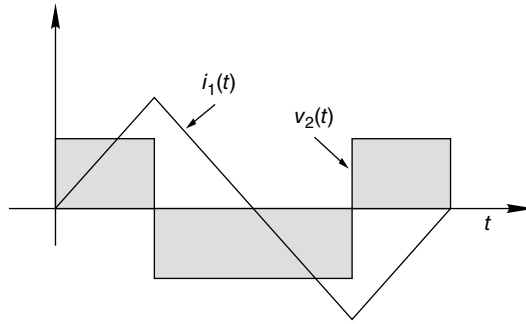


FIGURE 63.19 Output voltage of a Rogowski coil with a triangular wave input current.

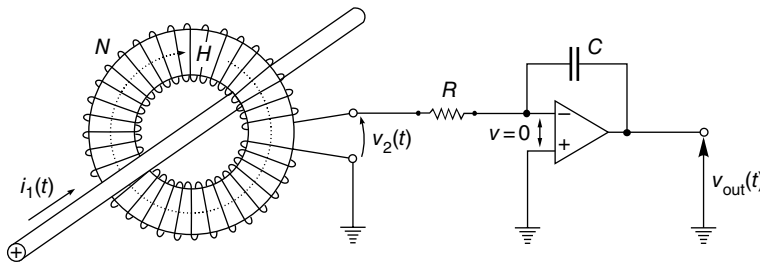


FIGURE 63.20 Rogowski coil with integrator.

Obviously, since the Rogowski coil is sensitive to the current's derivative, it cannot be used to measure DC currents. However, differently from magnetic CTs, the presence of a possible DC component in the primary current does not adversely affect the accurate measurement of the alternate components superimposed to it.

As seen before, the voltage at the coil's output terminals is proportional to the current's derivative. Thus, in order to have an actual transducer's output proportional to the primary current, such voltage must be integrated. This can be done by means of either active or passive integrating circuits. In Figure 63.20 the analog integrator is built through an operational amplifier with a capacitance C in its feedback path. As an example, if the triangular current $i_1(t)$ of Figure 63.19 is considered, the integration of the square wave output voltage $v_2(t)$ provides a voltage $v_{out}(t)$ that reproduces, for each time instant, the triangular wave of the current $i_1(t)$.

As an alternative solution, the voltage $v_2(t)$ could be acquired and digitized, and the integral could be calculated by means of digital signal processing.

The Rogowski coil is an extremely versatile current transducer. Its sensitivity can be regulated in a wide range by acting on both the turn density and their cross area. In this way, currents from a few milliamperes to several megaamperes can be measured.

Two possible physical typologies exist: flexible coil and rigid coil. The former is more versatile and more suitable for high-frequency currents but is less accurate than the latter. Rigid coils are more suitable for low-frequency and low currents.

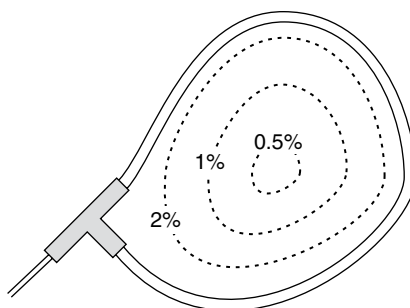


FIGURE 63.21 Possible impact of the conductor placement on the accuracy of a Rogowski coil.

The main uncertainty sources of the Rogowski coils are:

- *Assembly tolerance:*

Theoretically, the transducer's output should not depend on the position of the conductor with respect to the coil. Actually, the nonperfect setting up of windings and insulating supports gives rise to errors that can be up to some percent. The largest error arises when the conductor passes close to the junction between the two extremes of the coil. The smallest error is found, obviously, when the conductor passes near the center of the coil (Fig. 63.21).

- *Temperature variations:*

The dependency on the temperature can be minimized by using special materials with very low temperature coefficient. As an alternative, the temperature can be measured and its effects compensated.

- *Cross talk:*

The cross talk effect, due to the influence of currents in neighboring conductors, mainly depends on the phase-to-phase distance. Such disturbance can be minimized in the design stage. A fundamental solution is that the return link should be placed internally to the nonmagnetic core of the support where the turns are wound (Fig. 63.22). If this was not done, possible magnetic fields having direction perpendicular to that of the coil plane would be concatenated with the coil and would affect seriously the measurement results.

Differently from current transformers and other transducers, Rogowski coil does not need magnetic cores, and thus no saturation occurs. This intrinsically linear behavior makes this device suitable to measure high fault currents. Thanks to this wide measurement range (Fig. 63.23), a single current transducer can be used for both measurement and protection purposes. On the contrary, when traditional CTs are used, different windings should be considered for the different measurement ranges.

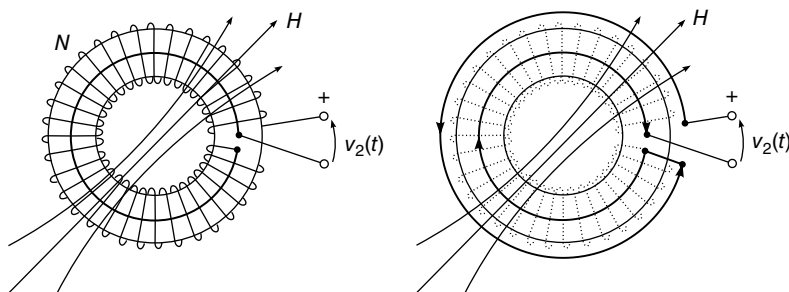


FIGURE 63.22 Compensation of cross talk in a Rogowski coil: position of the return path (left) and equivalent circuit (right).

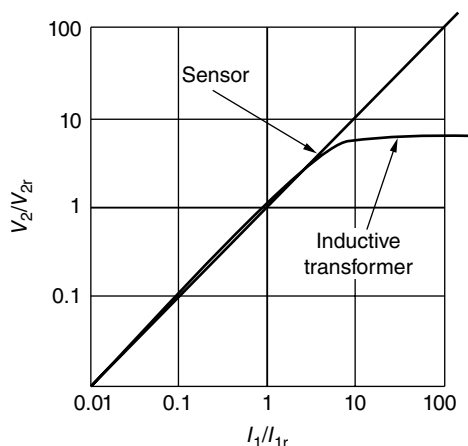


FIGURE 63.23 Input–output characteristic of air-core sensor and inductive CT.

Furthermore, the possible safety issues that could occur in CTs if the secondary circuit was opened (recalled in Section 63.3.1) are not present in Rogowski coils, because in this case no overvoltage would be generated.

The main limitations to the current range arise from the integrator. The measurement range depends on both amplitude and frequency of the measured current. Low currents at low frequency would produce very low voltages at the coil's terminals, thus leading to sensitivity problems. On the other hand, high currents rapidly changing (i.e., with high time derivative) can lead to high voltages that the electronics of the integrator could not tolerate.

The integrator also limits the bandwidth, which, in practical cases, varies for a few hertz to hundreds of kilohertz, which is enough for most applications in power systems. Special productions allow wider bandwidth to be reached, up to hundreds of megahertz.

Thanks to their attractive practical and metrological features, Rogowski coils are used in several applications on power grids, such as fault current detection, PQ monitoring, power and energy measurements, measurement of small currents superimposed

on large DC (e.g., capacitor ripple), large AC currents (e.g., arc furnaces), measurement of bearing and shaft currents in large machines, etc. All these possible applications make this device the most valid alternative solution to CTs, in particular in low- and medium-voltage systems. Some applications for high-voltage systems have been also proposed.

63.7 OPTICAL CURRENT AND VOLTAGE TRANSDUCERS

Optical methods for measuring voltages and currents have received increasing attention in the last years, especially for high-voltage systems. This is due to the advantages they allow with respect to instrument transformers: high immunity to electromagnetic interferences, excellent insulation, lightweight and reduced size, no saturation, wide measurement range, and wide bandwidth.

An optical transducer for voltage or current is a complex system that includes an optical sensing element, an optical fiber communication path, and an electronic module for signal processing and interfacing with measurement and protection equipment.

The sensing element is placed close to the quantity to be measured and generally produces a light modulation.

Some fundamental concepts of optics should be known to understand the principle of operation of optical voltage and current transducers. In order to make this section “self-consistent,” in the following the most important concepts will be recalled in an extremely concise and descriptive way, which must be considered neither complete nor scientifically rigorous. For deeper analysis the reader should refer to specialized texts.

Light *polarization* is a phenomenon specific of the propagation of the light waves, characterized by the oscillation direction of the waves in a given plane, known as polarization plane. This is the plane defined by the oscillation direction (conventionally determined by the direction of the electric field) and the propagation direction (see, for instance, Fig. 63.24, where the shaded gray area defines the polarization plane and c is the propagation speed). When the polarization plane holds a fixed direction,

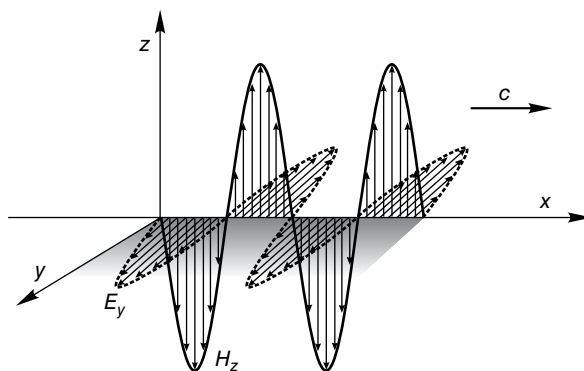


FIGURE 63.24 Propagation of an electromagnetic wave (horizontal E polarization).

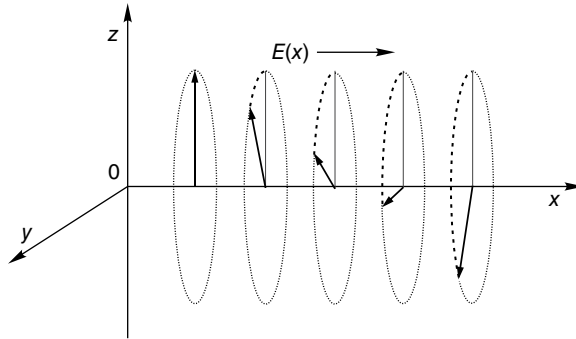


FIGURE 63.25 Rotation of a linearly polarized wave.

the light waves are said to be linearly polarized. This occurs when the electric field components with respect to the axis of a plane orthogonal to the propagation direction are in phase to each other. If the polarization plane rotates around the propagation direction with constant angular speed, the polarization is said to be elliptical (electric field components are not in phase) or circular (when the phase shift between the electric field components is $\pi/2$).

Optical transducers used in power systems are mainly based on two characteristics of optical materials: *optical activity* and *birefringence*.

The *optical activity* is a property of some materials in which, when a light wave passes through them, its polarization plane rotates (Fig. 63.25). If a linearly polarized light beam crosses an optically active matter, the transmitted wave is still linearly polarized but, on a different plane, shifted of a given angle with respect to the incidence polarization plane.

Birefringence is a property exhibited by some materials, in which the index of refraction depends on the propagation direction of light. In particular, the index has two different values for mutually orthogonal light polarizations. Birefringence can be intrinsic, in anisotropic materials like crystals, or induced by either mechanical, electrical, or magnetic stimulus. The different propagation speed of the light in the materials introduces a phase shift between the two orthogonal components. Thus, for instance, a linearly polarized light becomes elliptically polarized.

63.7.1 Optical Current Transducers

The principle of operation of optical current transducers (OCTs) is based on the magneto-optic Faraday's effect, which is essentially a modulated optical activity. When an optical material, subjected to a magnetic field, is crossed in the field direction by a linearly polarized light beam, the polarization plane rotates with an angle proportional to the field intensity:

$$\theta = \mu V \int_L H \, dl \quad (63.13)$$

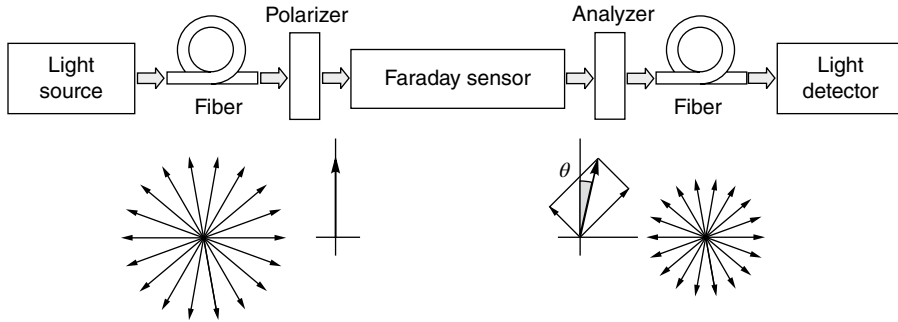


FIGURE 63.26 Main components of an optical current transducer and relevant light polarization status.

where θ is the rotation of the polarization, μ is the magnetic permeability of the material, H is the magnetic field component parallel to the propagation direction of the beam, L is the path length of the light, and V is the Verdet constant, which depends on material, wavelength, and temperature.¹

An OCT generally consists of the following elements (Fig. 63.26):

- Light source (usually a diode)
- Optical fibers, which provide the link between the sensing element and the electronic components
- Polarizer, which could be thought as a device that selects only one kind of light polarization of all the incident optical energy
- Sensing element, built with either silicon crystals or optical fibers
- Photodetector (photodiode), which converts optical signals into electrical signals.
- Digital signal processor (DSP), which implements the function that relates the physical phenomena to the measured parameters. In addition, the processor allows performing some compensation (temperature, linearity, etc.) and establishing digital communication with other devices.

In most recent and accurate OCTs, the sensing element has an optical path surrounding the conductor. In this way, according to the Ampere's law, if the light is uniformly sensitive to the magnetic field H in the closed path L around the conductor, then the rotation θ of the polarization plane is proportional to the current flowing in the conductor:

$$\theta = \mu V \int_L H \, dl = \mu V i \quad (63.14)$$

This solution can be practically implemented in different ways.

¹In some texts the Verdet constant includes the magnetic permeability μ , which, consequently, does not appear in the expression of the rotation. The substance does not change, but it is necessary to consider carefully the definition used, in order to have a correct dimensional interpretation of Equation 63.13.

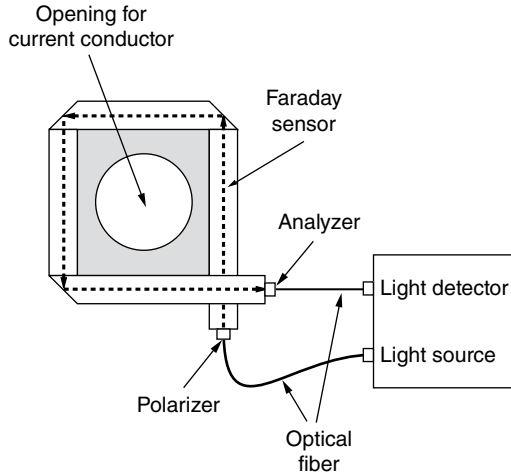


FIGURE 63.27 Example of OCT.

As an example, the sensing element can consist of crystals linked to each other to form a closed ring that surrounds the conductor (Fig. 63.27).

In a second solution the path of the light is developed in an optical fiber wound around the conductor. The use of optical fibers allows more flexibility to be achieved. Since optical fibers have usually a low Verdet constant, the required sensitivity is obtained by increasing the number of turns n around the conductor, according to the expression:

$$\theta = \mu n V i \quad (63.15)$$

Actually, especially when fiber optic OCTs are concerned, the Faraday's magneto-optic effect may be exploited in a different way, that is, involving different polarization conditions of the light, with respect to the previously described solution.

As far as the analysis of the output signal of the sensor is concerned, it should be considered that the variation of the polarization status, for example, the rotation of the polarization plane in the scheme of Figure 63.26, cannot be measured directly, since photodetectors are not sensitive to the light polarization but to the optical power, which is proportional to the square of the electric field. Therefore, suitable methods should be implemented to measure indirectly the rotation of the polarization.

As an example, by referring again to the solution described in Figure 63.26, one of the simplest methods consists of processing the light leaving the sensing element with a second polarizer, named analyzer, whose axis forms typically an angle of $\pi/4$ with respect to the axis of the first polarizer.

If α is the angle between the direction of light polarization at the sensor output and the polarization axis of the analyzer, under the assumption of no losses, the optical power P_{det} received by the detector can be expressed as a function of the input power P_{in} :

$$P_{\text{det}} = P_{\text{in}} \cos^2 \alpha = \frac{1}{2} P_{\text{in}} [1 + \cos(2\alpha)] \quad (63.16)$$

where $\alpha = \pi/4 + \theta$.

This means that

$$P_{\text{det}} = \frac{1}{2} P_{\text{in}} [1 - \sin(2\theta)] = P_{\text{dc}} - P_{\text{ac}} \quad (63.17)$$

The rotation θ varies with time following the time variability of the input current. Thus, the signal at the detector has a constant term $P_{\text{dc}} = 1/2 P_{\text{in}}$ and a variable term $P_{\text{ac}} = 1/2 P_{\text{in}} \sin(2\theta)$, which represents the modulation caused by the Faraday effect. By taking into account that the rotation angle θ is small, the variable component can be expressed as

$$P_{\text{ac}} = \frac{1}{2} P_{\text{in}} \sin(2\theta) = P_{\text{in}} \theta + \text{higher-order terms} \quad (63.18)$$

By neglecting the higher-order terms, normalizing the alternating component of the P_{det} with respect to the constant one and making explicit the time variability of the quantities, it results to

$$\frac{P_{\text{ac}}}{P_{\text{dc}}} = 2\theta(t) = Ki(t) \quad (63.19)$$

where $i(t)$ is the measured current and K is a constant term that depends on the characteristics of the sensor. It is evident that this method does not allow the DC components of the current to be measured. To overcome this limit, different solutions, often involving more than one analyzer, can be implemented.

As far as the metrological behavior is concerned, OCTs today available can have good performance under both sinusoidal and distorted conditions. Accuracy specifications are ensured in a wide dynamic range, from 1% up to 200% of the rated current. This allows the same transducer to be used for both measurement and protection purposes, thus avoiding the need to have two instrument transformers.

As for the frequency response, that of the sensing element could be very high, but that of the complete transducer is limited to some tens of kilohertz, mainly because of the presence of sampling and conversion electronic devices.

Further advantages of the OCTs are reduced weight and size, which lead to reduction of transportation and installation costs and low maintenance costs.

63.7.2 Optical Voltage Transducer

Most optical voltage transducers (OVTs) base their operation on the electro-optic Pockels effect: when a crystal is subjected to an electric field parallel to the light direction, a birefringence proportional to the electric field, and thus to the applied voltage, is induced.

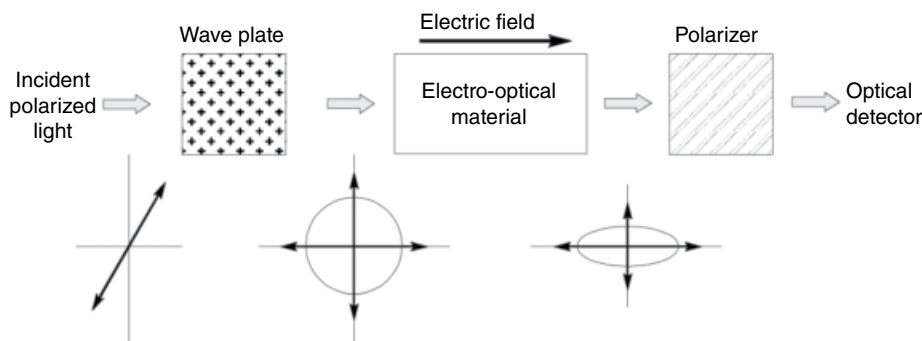


FIGURE 63.28 Main components of an optical voltage transducer and relevant light polarization status.

Figure 63.28 shows the main optical elements that compose an OVT. After a polarizer (not shown in Fig. 63.28) has converted the incident nonpolarized light into a linearly polarized beam, which can be considered composed by two in-phase orthogonal components, the light enters the quarter wave plate, which introduces a $\pi/2$ shift between the two components, thus giving rise to a circularly polarized light beam.

This beam enters the Pockels cell, where an electric field in the light direction is generated by the voltage applied to a couple of electrodes. The electric field induces the birefringence, which causes a further phase shift between the two light components, thus transforming the circular polarization into an elliptical one. The phase difference induced by the birefringence, which is proportional to the applied voltage, is converted by an analyzer into a modulation of the optical power. This power is measured by means of a photodetector, which converts the optical signal into an electrical one, and suitable data acquisition and processing systems.

Some commercial OVTs employ capacitive dividers to reduce the voltage applied to the optical sensor.

As for the metrological behavior, most of the considerations done for OCTs hold for OVTs. Their accuracy class is comparable to that of instrument transformers, and in addition, thanks to wide bandwidth and excellent linearity, these instruments can accurately measure both DC and AC voltages in a range from 20 to 200% of the rated value.

63.7.3 Applications of OCTs and OVTs

Manufacturers often propose commercial solutions that include in a single device the functionalities of optical voltage and current transducers, thus resulting in more compact units that can be advantageous in those installations where space is a critical issue.

As far as the field of application is concerned, it has been already stated that optical current and voltage transducers have been originally designed and built for high-voltage systems. However, the optical technology has evolved so that more recently, optical transducers have become available also for medium-voltage grids. Their characteristics may be in fact very useful in a measurement system designed to monitor, control, and protect

active distribution networks. The high dynamic range allows measuring very accurately both low and high currents, thus eliminating the need for parallel current transformers. At the same time, this flexibility may allow reducing inventory and simplifying maintenance. Their bandwidth extends to several kilohertz, thus allowing a correct reproduction of the distorted quantities present in these systems. The possibility of a digital output means a direct compatibility with digital measurement and protection devices. The lightweight design and intrinsic insulation enable the sensors to be easily and economically installed along feeders. For all the previous reasons, optical transducers are considered among the most promising solutions for next-generation distribution grids.

REFERENCES AND FURTHER READING

Journal and Conference Papers

- Emerging Technologies Working Group and Fiber Optic Sensors Working Group: "Optical current transducers for power systems: a review," *IEEE Transactions on Power Delivery*, Year: 1994, Volume: 9, Issue: 4, Pages: 1778–1788, DOI: 10.1109/61.329511.
- Kojovic, L.: "Rogowski coils suit relay protection and measurement," *IEEE Computer Applications in Power*, Year: 1997, Volume: 10, Issue: 3, Pages: 47–52.
- Kucuksari, S.; Karady, G.G.: "Experimental comparison of conventional and optical current transformers," *IEEE Transactions on Power Delivery*, Year: 2010, Volume: 25, Issue: 4, Pages: 2455–2463.
- Locci, N.; Muscas, C.: "Comparative analysis between active and passive current transducers in sinusoidal and distorted conditions," *IEEE Transactions on Instrumentation and Measurement*, Year: 2001, Volume: 50, Issue: 1, Pages: 123–128.
- Locci, N.; Muscas, C.; Sulis, S.: "Experimental comparison of MV voltage transducers for power quality applications," *IEEE Instrumentation and Measurement Technology Conference*, Year: May 5–7, 2009, Pages: 92–97, DOI: 10.1109/IMTC.2009.5168422.
- Minkner, R.; Schweitzer, E.O. III: "Low Power Voltage and Current Transducers for Protecting and Measuring Medium and High Voltage Systems," 26th Annual Western Protective Relay Conference, Washington State University, October 1999.
- Oates, C.D.M.; Burnett, A.J.; James, C.: "The design of high performance Rogowski coils," *International Conference on Power Electronics, Machines and Drives*, Year: June 4–7, 2002, Pages: 568–573, DOI: 10.1049/cp:20020179.
- Ray, W.F.; Hewson, C.R.: "High performance Rogowski current transducers," *IEEE Industry Applications Conference*, Year: 2000, Volume: 5, Pages: 3083–3090.
- Sawa, T.; Kurosawa, K.; Kaminishi, T.; Yokota, T.: "Development of optical instrument transformers," *IEEE Transactions on Power Delivery*, Year: 1990, Volume: 5, Issue: 2, Pages: 884–891.
- Ward, D.A.: "Measurement of current using Rogowski coils," *IEE Colloquium on Instrumentation in the Electrical Supply Industry*, Year: 1993, Volume: 1, Pages: 1–3.
- Xiao, C.; Zhao, L.; Asada, T.; Odendaal, W.G.; van Wyk, J.D.: "An overview of integratable current sensor technologies," *38th IAS Annual Meeting. Conference Record of the Industry Applications Conference*, Year: 2003, Volume: 2, Pages: 1251–1258.

Application Notes and Technical Brochures

ABB: “*Instrument Transformers—Application Guide*,” 4th edition, ABB, Pinetops, NC, 2015.

LEM Components: “*Isolated Current and Voltage Transducers: Characteristics—Applications—Calculations*,” 3rd edition, LEM Corporate Communications, Geneva, Switzerland, 2005.

Analog Devices: “Analog Isolation Amplifiers”, MT-071 Tutorial, 2009.

International Standards

IEC 61869-1:2007:

Instrument transformers—Part 1: General requirements

IEC 61869-2:2012:

Instrument transformers—Part 2: Additional requirements for current transformers

IEC 61869-3:2011:

Instrument transformers—Part 3: Additional requirements for inductive voltage transformers

IEC 61869-4:2013:

Instrument transformers—Part 4: Additional requirements for combined transformers

IEC 61869-5:2011:

Instrument transformers—Part 5: Additional requirements for capacitor voltage transformers

IEC 60044-7:1999:

Instrument transformers—Part 7: Electronic voltage transformers

IEC 60044-8:2002:

Instrument transformers—Part 8: Electronic current transformers

IEEE Std 3004.1-2013:

IEEE Recommended Practice for the Application of Instrument Transformers in Industrial and Commercial Power Systems

Year: 2013, DOI: 10.1109/IEEESTD.2013.6512522

IEEE Std 1601-2010:

IEEE Trial-Use Standard for Optical AC Current and Voltage Sensing Systems

Year: 2010, DOI: 10.1109/IEEESTD.2010.5674139

IEEE Std C57.13-2008:

IEEE Standard Requirements for Instrument Transformers

Year: 2008, DOI: 10.1109/IEEESTD.2008.4581634

IEEE Std C37.235-2007:

IEEE Guide for the Application of Rogowski Coils Used for Protective Relaying Purposes

Year: 2008, DOI: 10.1109/IEEESTD.2008.4457884

ELECTRIC POWER AND ENERGY MEASUREMENT

ALESSANDRO FERRERO AND MARCO FAIFER

Dipartimento di Elettronica, Informazione e Bioingegneria, Politecnico di Milano, Milano, Italy

64.1 INTRODUCTION

Power and energy measurements are probably the most important and critical measurements in power systems, since they are used to quantify every energy transactions and assign an economical value to energy flowing through a given section of the grid.

No wonder then that measurement methods and instruments aimed at quantifying energy flow and power have represented a challenge since the very beginning of the commercial exploitation of electricity. At first, when DC systems were the only available systems to distribute electricity, and the distance between generators and loads was so short that voltage was assumed to be constant, only current was integrated by ampere hour meters. It was soon realized, however, that the assumption of constant voltage was not correct and that voltage variations had a significant impact on the amount of delivered energy. The first DC energy meters were then designed and installed.

Things became even more complex when the first AC distribution systems began to be used as an alternative and more efficient way to distribute electricity and, in the end, replaced the DC systems almost completely. Not only voltage magnitude contributed to the amount of useful energy transfer but also the current phase shift with respect to

voltage. Methods and instruments to measure active and reactive power and energy had to be developed.

Today instruments are significantly different from the early ones and also from those who had been in use since about a decade ago. Therefore, the following sections will focus only on the modern ones.

On the other hand, the theoretical concepts that are behind the definition of electric power and energy are always the same and often not fully perceived. Since they may have a nonnegligible impact on the correct interpretation of the measurement results, the next section will briefly recall these concepts.

64.2 POWER AND ENERGY IN ELECTRIC CIRCUITS

It is generally thought that power and energy definitions are the same for DC and AC circuits. This assumption is only an approximation of a stricter derivation. To prove this, the DC and AC conditions will be covered separately.

64.2.1 DC Circuits

Let us consider a constant electric field, represented by a constant electric field vector \vec{E} . Let us also suppose that a constant free electric charge q is present in the field. The field will manifest itself with a force acting on charge q , such as

$$\vec{F} = q\vec{E} \quad (64.1)$$

Since the charge is free, it will move in the electric field along the same direction x as the orientation of the electric field, and the elementary work done by the field on the charge will be

$$dW = Fdx \quad (64.2)$$

According to (64.1), this elementary work can be expressed in terms of the electric field and charge values as

$$dW = qEdx \quad (64.3)$$

It is well known that, according to Stokes theorem, the infinitesimal electric potential difference is related to the electric field as $dV = Edx$. It is also known that power is defined as the first time derivative of work. Therefore, differentiating (64.3) and considering the electric potential yields

$$\frac{dW}{dt} = q \frac{dV}{dt} \quad (64.4)$$

Taking into account that the electric current I is defined as the flux of an electric charge in time, (64.4) becomes

$$\frac{dW}{dt} = IdV \quad (64.5)$$

Equation 64.5 is the well-known definition of power in a section of a DC electric circuit:

$$P = VI, \quad (64.6)$$

where I is the current flowing through the section and V is the potential difference (voltage) across the section.

The energy flowing through the same section is, of course, the integral of P over time and is given, therefore, by

$$W = \int_{t_0}^t VI dt = VI(t - t_0) \quad (64.7)$$

being V and I constant.

It can be readily perceived, from (64.6) and (64.7), that power and energy, in a DC circuit, can be measured by measuring V and I and the elapsed time.

64.2.2 AC Circuits

64.2.2.1 General Case It is generally thought that (64.5) can be applied also to circuits where voltages and currents vary with time (such as the AC circuits) by simply replacing the constant values of V and I by their functions of time. However, this is incorrect because (64.3) assumes that only the electric field exists, and this is not true if the field is not constant in time. Under time-varying conditions, electric and magnetic fields exist in the same point of space and are related by Maxwell equations.

Under these time-varying conditions, the power associated with the transverse electromagnetic wave has to be considered. Without entering into too many theoretical details, for which the reader is addressed to [1–3], let us only remind that the instantaneous power of the transverse electromagnetic wave is given by the flux of the Poynting vector through a given closed surface Σ . In mathematical terms, the Poynting vector is given by

$$\vec{\rho} = c^2 \epsilon_0 (\vec{E} \times \vec{B}) \quad (64.8)$$

where

c is the speed of light

ϵ_0 is the permittivity of free space

\vec{E} is the electric field vector

\vec{B} is the magnetic field vector

Its flux through a given closed surface Σ (i.e., the power associated with the electromagnetic wave) can be decomposed as

$$\oint_{\Sigma} \vec{\phi} \cdot \vec{u}_n d\Sigma = \oint_{\Sigma} v \vec{j} \cdot \vec{u}_n d\Sigma + \oint_{\Sigma} v \frac{\partial \vec{d}}{\partial t} \cdot \vec{u}_n d\Sigma + \oint_{\Sigma} \vec{B} \times \frac{\partial \vec{A}}{\partial t} \cdot \vec{u}_n d\Sigma \quad (64.9)$$

where

\vec{u}_n is the orthogonal versor to surface Σ

v is the electric potential

\vec{j} is the current density vector

\vec{d} is the displacement current vector

\vec{A} is the magnetic potential vector

Let us assume, without losing generality, that surface Σ is a spherical surface with diameter d , and let us assume that the wavelength $\lambda|_{\max(f)}$ ¹ of the highest-frequency component in the electromagnetic quantities is such that $d \ll \lambda|_{\max(f)}$. Under these conditions, it can be proved that the last two integrals in the right side of (64.9) become negligible with respect to the first one. Therefore

$$\frac{dW}{dt} = \oint_{\Sigma} \vec{\phi} \cdot \vec{u}_n d\Sigma \cong \oint_{\Sigma} v \vec{j} \cdot \vec{u}_n d\Sigma = v(t)i(t) \quad (64.10)$$

This last equation proves that the generally used definition of electric power, under variable conditions,

$$p(t) = v(t)i(t) \quad (64.11)$$

obtained as the product of the instantaneous voltage and current in a section of the electric circuit is only an approximation and provides correct results if and only if the dimensions of the circuit are negligible with respect to the wavelength of the electromagnetic quantities.

Of course, this is true for AC circuits operated under sinusoidal conditions at mains (50 or 60 Hz) fundamental frequency. It still holds when the signals are distorted, provided that the frequency bandwidth of the signal remains well below a few hundreds of kilohertz. Should this be not the case, the second and third integrals in (64.9) might become significant enough to make (64.11) not sufficiently accurate.

¹It is worth reminding that the wavelength of a waveform is related to its frequency by $\lambda = c/f$, c being the speed of light.

64.2.2.2 The Sinusoidal Conditions Nowadays, sinusoidal AC systems are the most widely used to transport and distribute electric energy. It is therefore important to analyze (64.11) under these conditions.

Let us assume that the voltage and current waveforms are sine waves and are described, respectively, by

$$v(t) = \sqrt{2}V \sin(2\pi ft) \quad (64.12)$$

$$i(t) = \sqrt{2}I \sin(2\pi ft + \varphi), \quad (64.13)$$

where

V and I are the rms values of voltage and current, respectively

f is the frequency at which the AC system operates

φ is the phase displacement between the current and voltage waveforms and is considered conventionally positive if the current is leading and negative if the current is lagging

According to (64.11), the instantaneous power is given by

$$p(t) = v(t)i(t) = 2VI \sin(2\pi ft) \sin(2\pi ft + \varphi) \quad (64.14)$$

Taking into account that $\sin \alpha \sin \beta = \frac{1}{2} [\cos(\alpha - \beta) - \cos(\alpha + \beta)]$, (64.14) becomes

$$p(t) = 2VI \left\{ -\frac{1}{2} [\cos(4\pi ft + \varphi) - \cos \varphi] \right\} = VI \cos \varphi - VI \cos(4\pi ft + \varphi) \quad (64.15)$$

$v(t)$, $i(t)$, and $p(t)$ are plotted in Figure 64.1 for $V = 1$ V, $I = 1$ A, $f = 50$ Hz, and $\varphi = -\pi/6$.

It can be readily perceived, from (64.15) and Figure 64.1, that the instantaneous power shows an average value $P = VI \cos \varphi$ that differs from zero if $\varphi \neq \pm \pi/2$ and shows a variable term that oscillates about P with a frequency that is twice the voltage and current frequency. The average power P represents the useful power transfer.

The energy flow in a circuit section is given by

$$W = \int_{t_0}^{t_0+t} v(\tau) i(\tau) d\tau \quad (64.16)$$

It can be immediately recognized that, when the integration time t is equal to the signal period T or an integer multiple of T ($t = kT$, k an integer), (64.16) becomes

$$W = \int_{t_0}^{t_0+kT} P d\tau \quad (64.17)$$

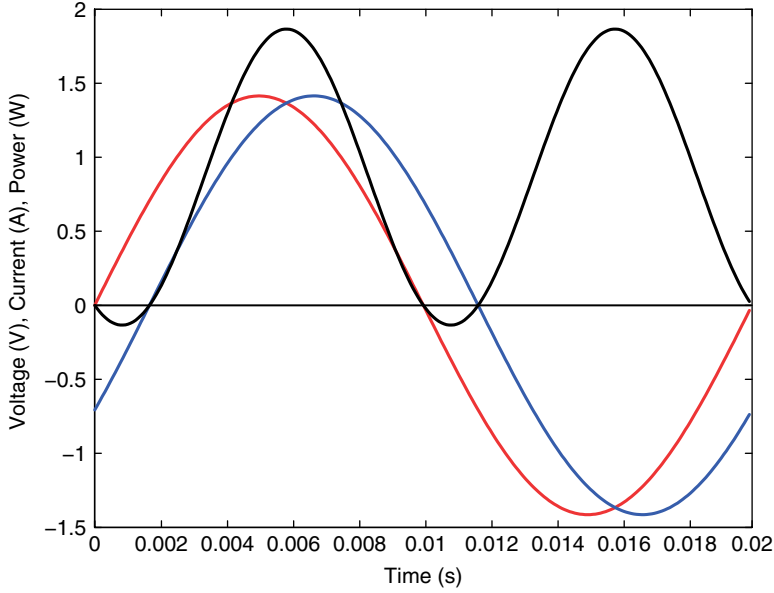


FIGURE 64.1 Voltage (dark gray), current (light gray), and instantaneous power (black) waveforms in a sinusoidal AC system. $V = 1 \text{ V}$, $I = 1 \text{ A}$, $\varphi = -\pi/6$.

It is now interesting to analyze the oscillating part of the instantaneous power to understand whether it shows interesting properties. To do so, let us decompose the current waveform into a component in phase and a component in quadrature ($\pi/2$ phase shift) with the voltage waveform. It is expanding (64.13):

$$\begin{aligned} i(t) &= \sqrt{2}I \sin(2\pi ft) \cos \varphi + \sqrt{2}I \cos(2\pi ft) \sin \varphi \\ &= \sqrt{2}I \sin(2\pi ft) \cos \varphi + \sqrt{2}I \sin\left(2\pi ft + \frac{\pi}{2}\right) \sin \varphi \end{aligned} \quad (64.18)$$

Using the first line of (64.18) in (64.14), we get

$$p(t) = 2VI \cos \varphi \sin^2(2\pi ft) + 2VI \sin \varphi \sin(2\pi ft) \cos(2\pi ft) \quad (64.19)$$

Taking into account that

$$\sin^2 \alpha = \frac{1}{2} - \frac{1}{2} \cos 2\alpha$$

and

$$\sin \beta \cos \beta = \frac{1}{2} \sin 2\beta,$$

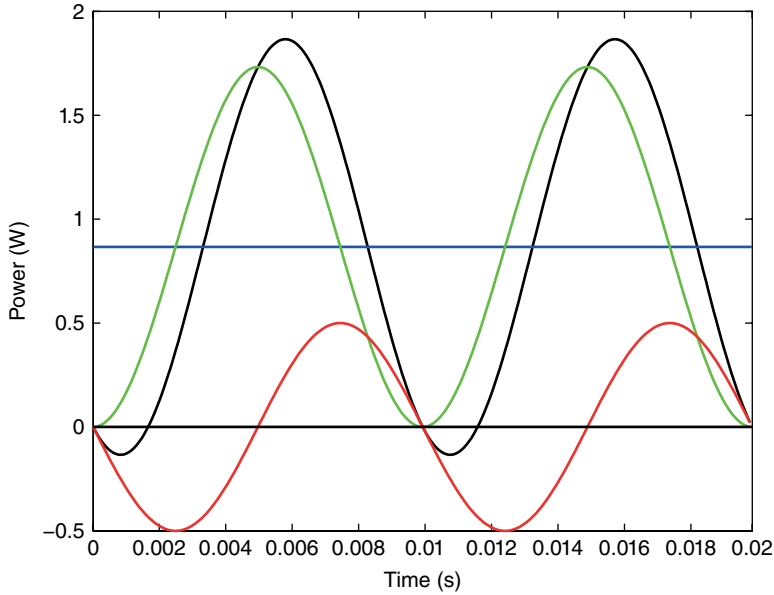


FIGURE 64.2 Instantaneous power components in a sinusoidal AC system: instantaneous power (black), average power (constant gray line), component originated by the in-phase current summed to the average power (light gray), component originated by the quadrature current (dark gray). $V=1\text{ V}$, $I=1\text{ A}$, $\varphi=-\pi/6$.

(64.19) becomes

$$p(t) = VI \cos \varphi - VI \cos \varphi \cos(4\pi ft) + VI \sin \varphi \sin(4\pi ft) \quad (64.20)$$

The three components of power evidenced by (64.20) are plotted in Figure 64.2, again for $V=1\text{ V}$, $I=1\text{ A}$, $f=50\text{ Hz}$, and $\varphi=-\pi/6$. In particular, the black line shows the instantaneous power $p(t)$, the constant line shows the average power P , the light gray line shows the sum of the first two components in (64.20), and the dark gray line shows the third component.

It can be readily checked that if $\varphi=0$, the third term in (64.20) is nil, the average power becomes $P=VI$, and the amplitude of the second oscillating term in (64.20) is VI . This is the situation of a single purely resistive load, supplied by a sinusoidal voltage. For this reason, and since it represents useful power transferred from the supply to the load, the average power P is also called *active power* [4].

On the other hand, it can be readily checked that if $\varphi = \pm \pi/2$, the first and second term in (64.20) are nil and the total instantaneous power is given by the third term and oscillates with zero mean value and peak value equal to VI . This is the situation of a single purely reactive load (inductor or capacitor) supplied by a sinusoidal voltage. For this reason, the peak value $Q = VI \sin \varphi$ of this component of the instantaneous power is called *reactive power*. Under sinusoidal conditions the reactive power has the property

of quantifying the amount of useless power transferred due to the presence of reactive elements. It can be used to design and size passive reactive compensators to minimize this power component [4].

Therefore, the power properties of an electric device can be fully represented by its active and reactive powers:

$$P = VI \cos \varphi \quad (64.21)$$

$$Q = VI \sin \varphi \quad (64.22)$$

The following quantity can be also defined:

$$S = \sqrt{P^2 + Q^2} = VI \quad (64.23)$$

S is called *apparent power* and represents the maximum active power that can be transferred under $\varphi = 0$ conditions. It is often referred to as the *design power*, since it can be obtained as the product of the rated voltage by the rated current.

P is measured in *watts* (W), Q in *reactive volt-amperes* (VA_r), and S in *volt-amperes* (VA). It is worth noting that the dimension is always that of a power, but different names have been given to the units to reinforce the different physical properties of the considered power components.

The following quantity is also defined:

$$\lambda = \frac{P}{S} = \cos \varphi \quad (64.24)$$

and is called *power factor*. It can be readily checked that it is always $\lambda \leq 1$ and that λ represents an index of how efficiently power is transferred for given values V and I of voltage and current.

64.3 MEASUREMENT METHODS

64.3.1 DC Conditions

64.3.1.1 Measurement Method According to (64.6), power in DC circuits can be measured by measuring voltage and current by means of a voltmeter and an ammeter, as shown in Figure 64.3. Under ideal conditions, the two instruments can be connected, as shown in Figure 64.3, with the voltmeter connected before the ammeter, or with the voltmeter connected after the ammeter, directly in parallel with the load, indifferently.

However, real instruments feature an internal resistance that, depending on the employed connection, yields a measured value of power different from the expected one. If the connection shown in Figure 64.3 is considered again, the load resistance and those of the employed instruments are connected as shown in Figure 64.4.

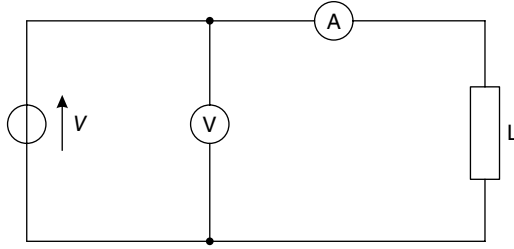


FIGURE 64.3 Instrument connection to measure the DC power drawn by a DC load L supplied by a DC voltage V . Power is measured by means of a voltmeter (V) and an ammeter (A).

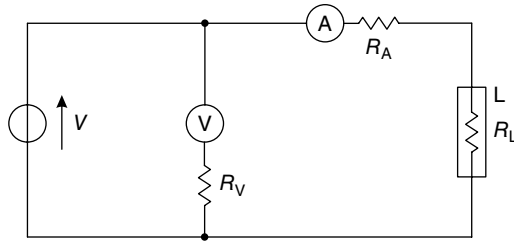


FIGURE 64.4 Instrument connection to measure the DC power drawn by a DC load L supplied by a DC voltage V . The ammeter is connected in series with the load. The load resistance, R_L , and the ammeter and voltmeter internal resistances, R_A and R_V , respectively, are also shown.

It can be readily perceived that the ammeter internal resistance R_A is connected in series with the load resistance R_L . Therefore, the current I_m measured by the ammeter is still the current I flowing in the load, but the voltage V_m measured by the voltmeter differs from the voltage V across the load and is given by

$$V_m = V + R_A I \quad (64.25)$$

Therefore, the measured power P_m is given by

$$P_m = V_m I_m = VI + R_A I^2 = P + P_A, \quad (64.26)$$

that is, the measured power is the actual power taken by the load plus the power dissipated by the internal circuits of the ammeter. This means that the internal resistance of the ammeter is responsible for a systematic contribution that becomes significant when this internal resistance is not much lower (some orders of magnitude) than that of the load. If this is the case, a correction must be applied for this systematic contribution.

On the other hand, if the voltmeter is connected directly in parallel with the load, as shown in Figure 64.5, voltage V_m measured by the voltmeter is voltage V on the load. The current measured by the ammeter (I_m) is given by

$$I_m = I + \frac{V}{R_V} \quad (64.27)$$

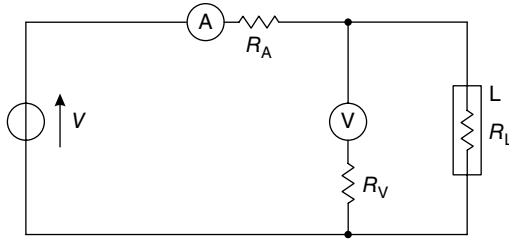


FIGURE 64.5 Instrument connection to measure the DC power drawn by a DC load L supplied by a DC voltage V . The voltmeter is connected in parallel with the load. The load resistance, R_L , and the ammeter and voltmeter internal resistances, R_A and R_V , respectively, are also shown.

Therefore, the measured power P_m is given by

$$P_m = V_m I_m = VI + \frac{V^2}{R_V}, \quad (64.28)$$

that is, the measured power is the actual power taken by the load plus the power dissipated by the internal circuits of the voltmeter. This means that the internal resistance of the voltmeter is responsible for a systematic contribution that becomes significant when this internal resistance is not much higher (some orders of magnitude) than that of the load. If this is the case, a correction must be applied for this systematic contribution.

64.3.1.2 Uncertainty Evaluation After having applied the required corrections, if needed, for the above systematic effects, uncertainty evaluation is a typical example of combined uncertainty evaluation.

Let us assume that the standard uncertainty values $u(V)$ and $u(I)$ have been evaluated for the measured values of voltage (V_m) and current (I_m) respectively, following either a type A or a type B evaluation method, as suggested by the Guide to the Expression of Uncertainty in Measurement (GUM) [5]. According to the GUM, the combined standard uncertainty $u(P)$ is given by [5]

$$u(P) = \sqrt{\left(\left.\frac{\partial P}{\partial V}\right|_{I_m}\right)^2 u^2(V) + \left(\left.\frac{\partial P}{\partial I}\right|_{V_m}\right)^2 u^2(I) + 2 \left.\frac{\partial P}{\partial V}\right|_{V_m} \left.\frac{\partial P}{\partial I}\right|_{I_m} u(V)u(I)r(V,I)}, \quad (64.29)$$

where $r(V,I)$ is the correlation coefficient that takes into account the possible correlation among the voltage and current measurements. By developing the partial derivatives in (64.29), we get

$$u(P) = \sqrt{I_m^2 u^2(V) + V_m^2 u^2(I) + 2I_m V_m u(V)u(I)r(V,I)}. \quad (64.30)$$

64.3.2 AC Conditions

64.3.2.1 Single-Phase Circuits According to (64.21), (64.22), and (64.23), the power exchange in a section of a circuit operated under sinusoidal AC conditions cannot be fully defined by measuring only the rms values of voltage and current. Dedicated instruments must be employed, and the most important is the wattmeter.

The schematics of the instrument connections are shown in Figure 64.6, where a wattmeter W is used to measure the active power (64.21). This instrument has a current port, connected in series with the load, and a voltage port, connected in parallel. Like the ammeters and voltmeters, its internal circuits feature a finite impedance, different from the ideal one, which is zero for the ampermetric circuit and infinite for the voltmetric circuit. Therefore a correction should be applied, as seen in the case of DC power measurement, to subtract the power drawn by the wattmeter internal circuit to the measured one. In general, the internal impedance of the voltmetric circuits is almost purely resistive, so that its effect can be more easily compensated. For this reason, the connection shown in Figure 64.6, where the voltmeter and the voltmetric circuits of the wattmeter are connected in parallel with the load, is preferred to the connection with the voltmetric circuits connected before the ampermetric ones.

According to Figure 64.6, the wattmeter provides the measured value P_m of the active power, the voltmeter the measured value V_m of the voltage, and the ammeter the measured value I_m of the current. Therefore, the measured value of the apparent power is obtained as $S_m = V_m I_m$, and the absolute value of the measured reactive power is obtained as $Q_m = \sqrt{S_m^2 - P_m^2}$. If the nature of the load (inductive or capacitive) is known, the sign of Q_m is also known. Otherwise, another instrument, called *varmeter*,² has to be connected, in the same way as the wattmeter, to measure the reactive power in a direct way.

The voltmeter and ammeter in Figure 64.6 are used not only to measure the apparent power but also to control the values of voltage and current in order not to overload the internal circuits of the wattmeter.

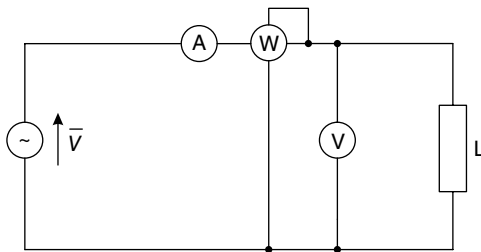


FIGURE 64.6 Instrument connection to measure the AC power drawn by a load supplied by a sinusoidal AC voltage \bar{V} . The active power is measured by a wattmeter (W).

²When sinusoidal AC conditions are considered, varmeters are generally wattmeters with an additional circuit that delays the input voltage by $\pi/2$.

64.3.2.2 Three-Phase Circuits Nowadays, most electric systems operated under sinusoidal AC conditions are three-phase systems. They can be three-wire systems, such as the one represented in Figure 64.7, or four-wire systems. The system in Figure 64.7 can be fully characterized by the three line-to-line voltages $\bar{V}_{ab}, \bar{V}_{bc}, \bar{V}_{ca}$ and the three line currents $\bar{I}_a, \bar{I}_b, \bar{I}_c$.

The most efficient way to operate such systems is when the three line-to-line voltages have the same amplitude and are phase-shifted by $\pi/3$ with respect to each other and the load is composed by three equal single-phase loads Y or Δ connected, so that the three line currents have the same amplitude and are phase-shifted by $\pi/3$ with respect to each other, as shown in the phasorial diagram of Figure 64.8. This situation is called *symmetrical* and *balanced*, and, under this condition, three-wire and four-wire systems behave in the same way, since the current in the fourth wire is always zero.

The diagram in Figure 64.8 shows also the line voltages $\bar{V}_a, \bar{V}_b, \bar{V}_c$, referred to the theoretical neutral point O (the center of gravity of the triangle of the line-to-line currents). The sum of these three line voltages, in phasorial terms, is nil, and it is $V_{ab} = \sqrt{3}V_a$.

It can be readily proved [4] that, under symmetrical and balanced conditions, the active power of a three-phase system is given by

$$P = 3V_a I_a \cos \varphi = \sqrt{3}V_{ab} I_a \cos \varphi, \quad (64.31)$$

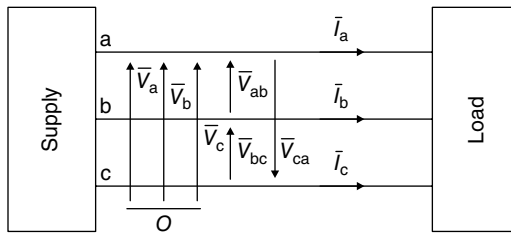


FIGURE 64.7 Three-wire, three-phase system.

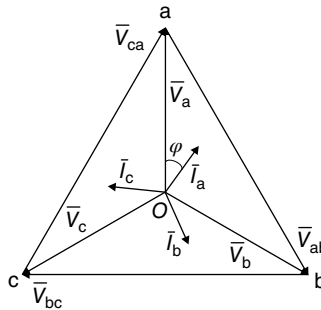


FIGURE 64.8 Phasorial diagram of voltages and currents in a symmetrical and balanced three-wire, three-phase system.

where φ is the phase angle between the line voltage phasor \bar{V}_a and the line current phasor \bar{I}_a . In phasorial terms, (64.31) can be rewritten as

$$P = 3(\bar{V}_a \cdot \bar{I}_a). \quad (64.32)$$

It can be also proven that, if the symmetrical and balanced conditions are not met, as it may happen under practical situations, the active power is given by

$$P = \bar{V}_a \cdot \bar{I}_a + \bar{V}_b \cdot \bar{I}_b + \bar{V}_c \cdot \bar{I}_c \quad (64.33)$$

Since the system is a three-wire system, it is also $\bar{I}_a + \bar{I}_b + \bar{I}_c = 0$, and, hence $\bar{I}_b = -\bar{I}_a - \bar{I}_c$. Equation 64.33 can be, therefore, written as

$$P = \bar{V}_a \cdot \bar{I}_a - \bar{V}_b \cdot (\bar{I}_a + \bar{I}_c) + \bar{V}_c \cdot \bar{I}_c = (\bar{V}_a - \bar{V}_b) \cdot \bar{I}_a + (\bar{V}_c - \bar{V}_b) \cdot \bar{I}_c. \quad (64.34)$$

According to the diagram in Figure 64.8, it can be seen that, under whatever conditions, it is $\bar{V}_{ab} = \bar{V}_a - \bar{V}_b$ and $\bar{V}_{cb} = \bar{V}_c - \bar{V}_b$. Therefore, (64.34) becomes

$$P = \bar{V}_{ab} \cdot \bar{I}_a + \bar{V}_{cb} \cdot \bar{I}_c. \quad (64.35)$$

This last equation shows that the total power across a section of a three-wire, three-phase system can be measured by means of only two wattmeters, connected as shown in Figure 64.9. This connection is generally known as the Aron connection and can be generalized to the four-wire systems, where three wattmeters are needed, with the current circuit connected to three wires and the voltage circuits connected across the same wires as those of the current circuits and the fourth wire, to measure the total power.

A similar theorem can be proven, under sinusoidal conditions, for the total reactive power. It is

$$Q = \|\bar{V}_a \times \bar{I}_a\| + \|\bar{V}_b \times \bar{I}_b\| + \|\bar{V}_c \times \bar{I}_c\| = V_a I_a \sin \varphi_a + V_b I_b \sin \varphi_b + V_c I_c \sin \varphi_c, \quad (64.36)$$

where $\varphi_a, \varphi_b, \varphi_c$, are the phase angles between the line voltages and currents of phases a, b, and c, respectively. Therefore, the total reactive power can be measured with two varmeters connected in the same way as the wattmeters in Figure 64.9.

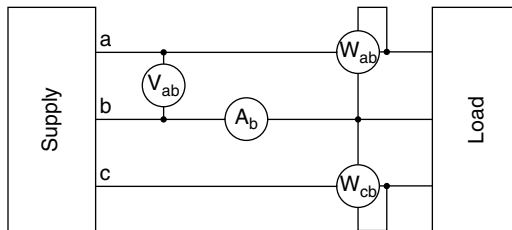


FIGURE 64.9 Aron wattmeter connection in a three-wire, three-phase system.

64.4 WATTMETERS

64.4.1 Architecture

Wattmeters and energy meters are probably among the oldest instruments built since the industrial exploitation of electric energy. The first wattmeters and energy meters were electromechanical instruments.

Electrodynamical actions were exploited to realize wattmeters [6], and this operating principle remained the same until the electronic circuits replaced, in the second half of the twentieth century, the electromechanical structures. Wattmeters based on thermal principles and analog multiplier structures have replaced the electrodynamic ones, until the more modern architectures based on digital signal processing (DSP) techniques have replaced also those instruments.

Energy meters, on the other hand, were based on the same principle as the magnetic rotating field obtained by Galileo Ferraris and used to design and realize AC electric machines. In the energy meters based on this principle, a thin metallic disk acts as the motor rotor and rotates at a speed proportional to the electric active power. The number of turns, counted by a mechanical system, is proportional to the active energy flowing in the metering sections [6]. These energy meters, known as *Ferraris meters*, were first produced at the very beginning of the twentieth century, represented the most used and widespread instruments all over the world, and millions of them are still in use, though they are now being slowly replaced by more modern instruments based on DSP structures.

Since the electromechanical and analog electronic structures are slowly disappearing, only the modern DSP-based structures will be briefly discussed in the following. The block diagram of their architecture, for a single-phase instrument, is shown in Figure 64.10.

The voltage $v(t)$ and current $i(t)$ signals are sensed by transducer circuits and signal conditioning circuits. The transducers play the important role of adjusting the signal

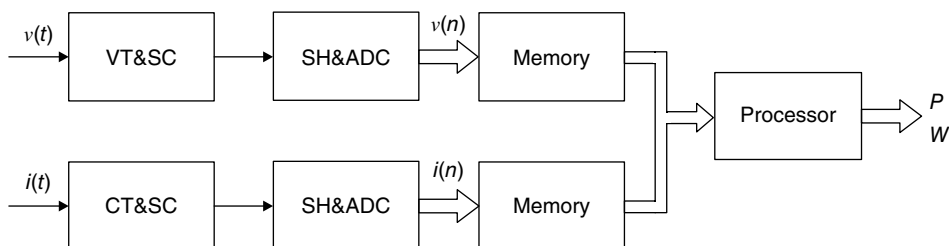


FIGURE 64.10 Structure of a modern instrument for power and energy measurement. CT&SC, current transducer and signal conditioning; SH&ADC, sample and hold and analog-to-digital converter; VT&SC, voltage transducer and signal conditioning. $v(t)$ and $i(t)$ voltage and current time signals, respectively. $v(n)$ and $i(n)$ sequences of the voltage and current samples, respectively.

level to that of the instrument's electronics and guarantee the necessary galvanic insulation with given accuracy over a given frequency band. Since their performance can be critical in ensuring the desired accuracy of the whole instrument, the next paragraph will be specifically dedicated to the voltage and current transducers and their structure and performance.

The signal conditioning circuits are needed to further adjust the input signals to the features of the subsequent blocks of the instrument. Among these circuits an important role, also in determining the final accuracy is played by the antialiasing filters [7].

The voltage and current signals are then sampled and converted into digital by means of sample and hold devices and analog-to-digital converters [7]. These two devices convert the original, continuous-time signals into sequences of samples, that is, the discrete-time signals $v(n)$ and $i(n)$ [7] that can be stored into the instrument memory and processed by the processor unit, according to dedicated signal processing algorithms. These algorithms are the core of the modern instruments and allow one to implement different measurement functions on the same instrument, such as active, reactive, and apparent power measurements, as well as their integration over time, so that active and reactive energy can be measured for billing purposes.

Three-phase instruments can be readily obtained from this structure simply by adding more input channels, to acquire the voltage and current signals as shown in previous Section 64.3.2.2, and by implementing the required algorithm to process them.

64.4.2 Signal Processing

As mentioned in the previous section, the signal processing algorithms represent the core of the modern instruments for power and energy measurement. Therefore, they are worth a dedicated analysis, since an incorrect choice of these algorithms or an incorrect choice of the sampling strategy may affect the accuracy of the obtained measurement results dramatically.

Let us start from the correct selection of the sampling frequency, since this is the most important point to ensure correct results. It is well known that, to be correctly sampled, the input signals must be upper limited in frequency [7]. Let us suppose, accordingly, that the voltage and current signals are periodic in time over a period T ; that the voltage signal $v(t)$ is upper limited, in frequency, by the M th harmonic component; and that the current signal $i(t)$ is upper limited, in frequency, by the N th harmonic component.

According to the mathematical derivations reported in the previous sections, all relevant power quantities are obtained from the instantaneous power $p(t) = v(t)i(t)$,³ whose samples are obtained from the samples of voltage and current as $p(n) = v(n)i(n)$. The

³This holds in general and especially under nonsinusoidal conditions. Since the modern instruments are more and more used for power quality measurements, when the signals are distorted, correct sampling of the instantaneous power is a critical issue with these instruments and must be carefully considered.

employed sampling frequency must therefore satisfy the sampling theorem for the instantaneous power $p(t)$.

In the frequency domain, the instantaneous power is obtained as the convolution of the frequency-domain versions $V(f)$ and $I(f)$ of $v(t)$ and $i(t)$, respectively. Therefore, the maximum harmonic component of $p(t)$ has order $M+N$. The ideal method for obtaining the correct number of samples of $p(t)$ is that of sampling $v(t)$ and $i(t)$ with a sampling period T_s so that [7]

$$T = [2(M+N) + 1]T_s \quad (64.37)$$

where $2(M+N)+1$ is the minimum number of samples required to sample $p(t)$ correctly.

Assuming $K=M+N$, the active power can be obtained as the average value of the instantaneous power as

$$P = \frac{1}{j(2K+1)} \sum_{k=0}^{j(2K+1)-1} v(k)i(k) \quad (64.38)$$

$j=1, 2, 3, \dots$ being an integer.

If $j=1$ is chosen, the obtained active power value is averaged over a single period. If higher values are taken for j , a longer averaging period can be considered. In particular, if $j=50$ is taken for systems operated at 50 Hz fundamental frequency and $j=60$ is taken for systems operated at 60 Hz fundamental frequency, the active power value P_1 is obtained, averaged over a 1 s interval of time.

It is then possible to easily obtain the total active energy flowing through the metering section in a time interval of k seconds as

$$W = \sum_k P_{1k} \cdot \quad (64.39)$$

Apparent and reactive powers can be also obtained from the samples of voltage and current by implementing one of the definitions given in Section 64.2.

It is worthwhile emphasizing the high flexibility of this modern DSP-based structure, which allows the implementation of different possible definitions. This is a great advantage, especially when power components under nonsinusoidal conditions have to be measured. On the other hand, since, as it will be shown later in this chapter, different definitions may yield, under nonsinusoidal conditions, different measured values for the same conditions, it is important that the adopted algorithm is always declared to avoid gross mistakes.

64.5 TRANSDUCERS

Power and energy measurements are performed in power systems at all voltage levels, from low-voltage to ultrahigh-voltage systems. It can be immediately perceived that these voltage levels are not compatible with the input dynamics of the employed

instruments, especially the electronic ones. Moreover, galvanic insulation must be ensured, for obvious safety reasons, between the power system and the instruments.

This requires the use of voltage and current transducers that adjust the voltage and current values to those compatible with the instrument dynamics and ensure the required insulation level.

The impact of these transducers on the metrological performance of the whole measuring system is so significant that it is worth covering this topic in a dedicated section of this chapter.

64.5.1 Current Transformers

The most common current transducer in power applications is the current transformer (CT). This transducer performs a reduction of the value of the AC current to be measured in order to adjust it to the input dynamics of the employed ammeter. Another important feature of this transducer is that it provides galvanic insulation between the power network and the measurement equipment, thus granting operator safety.

From an ideal point of view, the CT is an ideal transformer whose primary is series connected to the line carrying the current to be measured. Its secondary winding is short-circuited by means of an ideal ammeter as shown in Figure 64.11.

The current measured by the ammeter, \bar{I}_A , depends on the primary current of the transformer, \bar{I}_1 , according to the theoretical ratio K_T :

$$\bar{I}_A = -\bar{I}_2 = -\left(-\frac{N_1}{N_2}\bar{I}_1\right) = \frac{1}{K_T}\bar{I}_1 \quad (64.40)$$

where N_1 and N_2 are the numbers of turns of the primary and secondary windings, respectively.

In ideal CTs, K_T is a real constant. Consequently, (64.40) shows that the measured current \bar{I}_A is in phase with current \bar{I}_1 and is simply scaled by a constant factor.

Since the CT is usually employed to perform a reduction of the current value, K_T is a number much greater than one. This implies that if the secondary winding is left

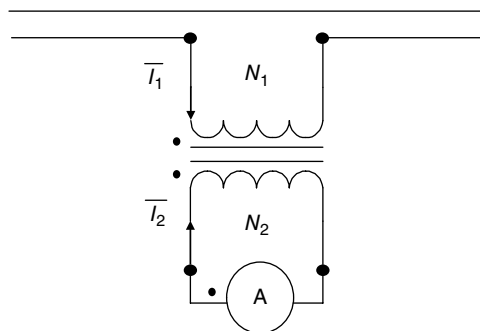


FIGURE 64.11 Ideal CT.

open, the voltage induced at the secondary terminal will be $\bar{V}_2 = K_T \bar{V}_1 \gg \bar{V}_1$. That is a very dangerous condition that must be avoided since it can result in electric arcs or explosions. Always for safety reasons, one of the terminal of the secondary winding must be connected to the ground in order to force its potential to zero. The absence of this connection can result in the presence of a dangerous voltage at the secondary terminals due to the capacitive coupling between the primary and secondary windings. The rated secondary output of CTs is 1 A or 5 A.

64.5.1.1 Measurement Errors Ideal CTs feature, of course, zero measurement errors. This is not the case for real CTs, and, in order to evaluate their errors and, hence, their contribution to the uncertainty of the whole measurement chain, let us now take into account the model of a real CT (Fig. 64.12).

This model considers the resistance and leakage reactance of the two windings, as well as the magnetizing branch. An impedance \bar{Z}_b representing the electric load of the CT has been also added. In order to evaluate the actual relationship between the currents at the primary and secondary windings of the CT, a phasor analysis shall be done. For the sake of simplicity, let us suppose that the CT operates with a load Z_b equal to zero. Under these conditions the phasor graph representing the CT is reported in Figure 64.13. In the graph, current \bar{I}_0 flowing in the magnetizing branch has been splitted into the magnetizing current \bar{I}_m , and the core and iron losses current \bar{I}_p .

Figure 64.13 shows that (64.40) is no longer valid. In fact, it can be noticed that there are a phase displacement and a difference in amplitude between \bar{I}_1 and $-K_T \bar{I}_2$.

For this reason two errors can be defined: the phase error ε and the ratio error η .

By considering that the phase error is generally small, it can be written as

$$\varepsilon \cong \sin \varepsilon = \frac{\overline{AB}}{\overline{OA}} = \frac{I_0}{I_1} \sin \theta \quad (64.41)$$

$$\theta = \varphi_0 - \varphi_2 \quad (64.42)$$

where φ_2 is the angle between \bar{I}_2 and \bar{E}_2 , the secondary no-load voltage, and φ_0 is the angle between \bar{I}_0 and \bar{E}_1 , the voltage applied to the magnetizing branch.

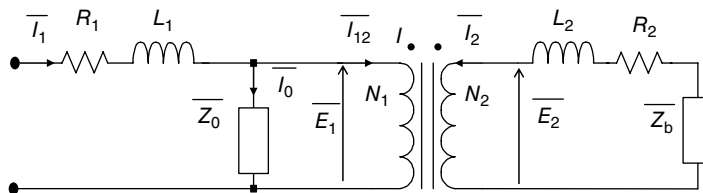


FIGURE 64.12 Model of a real CT.

For this reason, CTs are characterized by a rated load, expressed in VA at a given rated current, that represents the maxim value of the impedance that can be connected to the CT in order to guarantee the errors for the rated accuracy class. The common accuracy classes for CTs employed in power and energy measurements are 0.1, 0.2, 0.5, and 1. The accuracy specifications for these classes are listed in Table 64.1.

64.5.1.2 Saturation A typical problem of CTs is the saturation due to the nonlinearities of the magnetic core. Two kinds of saturations can be defined. The first one can be defined as symmetrical or AC saturation. This saturation occurs when the value of the AC current to be transduced forces the core to work with high values of magnetic flux so that it leaves the linear range of the $B-H$ curve (Fig. 64.14).

An example of AC saturation is reported in Figure 64.15. The primary current is a sine wave at the mains frequency, but its amplitude results in a too high magnetic flux that, due to the nonlinearity of the $B-H$ curve, is distorted in the peak area. Since

TABLE 64.1 Accuracy Specifications for CTs

Accuracy Class	\pm Percentage Ratio Error				\pm Phase Error (Minutes)			
% of rated the current	5	20	100	120	5	20	100	120
0.1	0.4	0.2	0.1	0.1	15	8	5	5
0.2	0.75	0.35	0.2	0.2	30	15	10	10
0.5	1.5	0.75	0.5	0.5	90	45	30	30
1	3	1.5	1	1	180	90	60	60

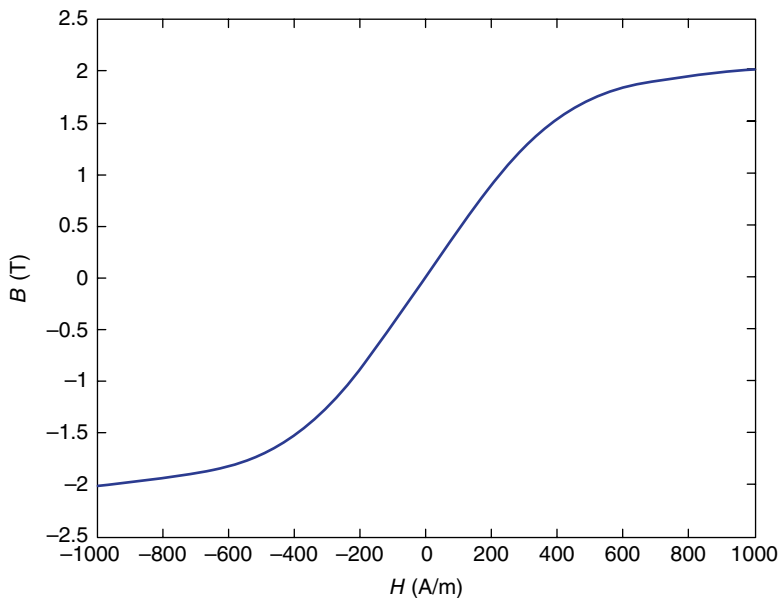


FIGURE 64.14 $B-H$ characteristic of the core.

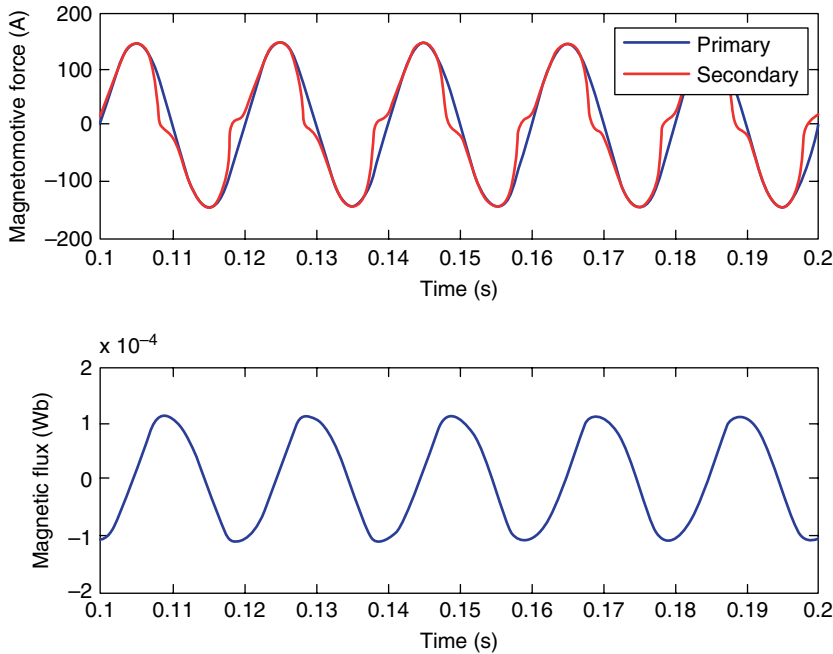


FIGURE 64.15 AC saturation.

current and flux are almost in quadrature, the effect of this distortion can be seen in the area of the zero crossing of the secondary current.

The second kind of saturation can be defined as asymmetrical saturation and is due to the presence of unidirectional components in the primary current or to the remanence in the magnetic core. In this condition, the magnetic core works on an asymmetrical loop in the B - H plane. Because of this, if the unidirectional component is not too high, the AC component of the current will cause a magnetic flux that is distorted only for positive or negative values. An example of asymmetrical saturation due to a unidirectional transient of the primary current is reported in Figure 64.16.

64.5.1.3 Bandwidth A CT is usually designed in order to properly work at the mains frequency. The definition of its errors, as previously reported, was derived by considering the CT working with a sinusoidal current at the mains frequency. It is evident that by changing the frequency of analysis, the values of the CT parameters change, so its behavior. In particular the core losses will change, as well as the B - H magnetization loop. The amount of these variations strongly depends on the CT design, in particular on materials and geometry. Moreover it has to be considered that a CT is a nonlinear system. The definition of its transfer function, which could be theoretically used in order to compensate for the errors, suffers from model uncertainty. In fact it does not take into account the effect of the intermodulation due to the CT nonlinearities. This effect can be significant and may lead the specifications for the rated accuracy class to be exceeded. In general, the -3 dB bandwidth of a CT does not exceed 1 kHz.

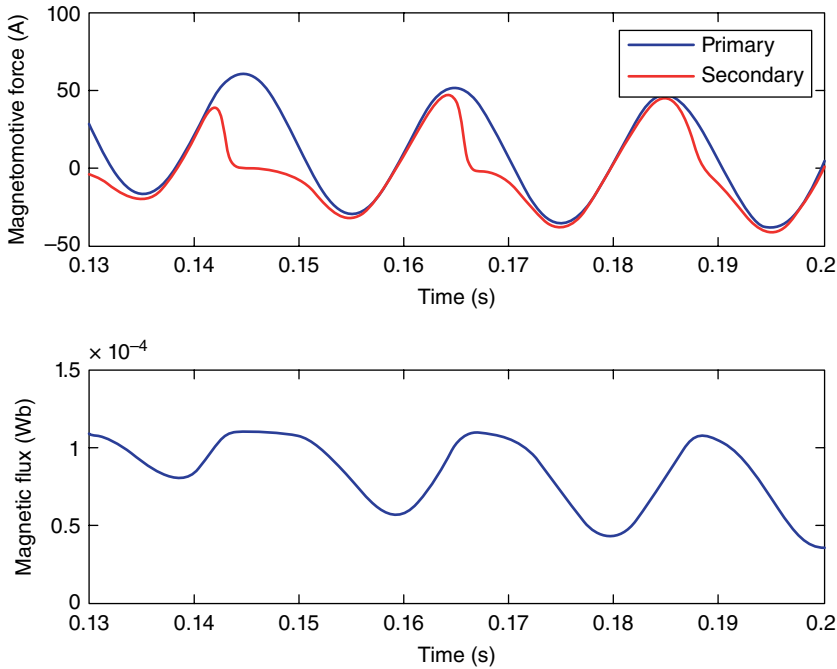


FIGURE 64.16 Asymmetrical saturation.

64.5.2 Hall-Effect Sensors

In order to overcome the typical problems of traditional CTs, a good solution is represented by the zero-flux transducers and in particular those based on Hall-effect sensors. A Hall-effect sensor is a transducer that converts a magnetic field B into a voltage V_H . In particular, due to the Hall effect, a voltage difference V_H is generated at the terminals of a semiconductor⁴ plate, when a current I flows in the plate and the plate itself is placed inside a magnetic field B , according to the topology shown in Figure 64.17. It is

$$V_H = B \cdot I \cdot K_H \quad (64.46)$$

where K_H is a constant that depends on the sensor dimensions and material.

This sensor is used to implement a zero-flux transducer in a closed-loop configuration. The typical solution is that shown in Figure 64.18. The Hall-effect sensor is placed inside an air gap in the magnetic core and senses the field generated by the current I_1 . The sensor output voltage controls a current generator that provides the

⁴Actually, the Hall effect is present also in conductor plates, but it is generally much lower than in semiconductors, so that only semiconductors are employed in Hall-effect sensors.

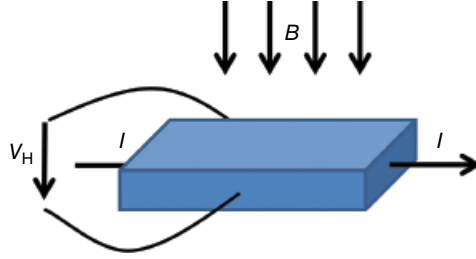


FIGURE 64.17 Hall effect.

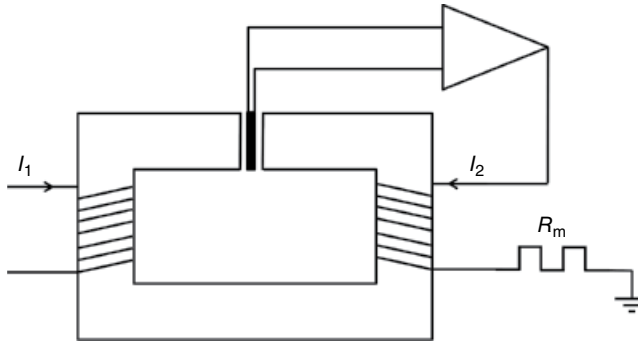


FIGURE 64.18 Zero-flux configuration based on a Hall-effect sensor.

current I_2 required to zero the magnetic field in the core. In this condition the voltage drop V_m on resistance R_m is

$$V_m = I_1 \cdot R_m \cdot \frac{N_1}{N_2} \quad (64.47)$$

The advantage of this transducer, with respect to the traditional CTs, is that it does not suffer from saturation problems, the dynamic performances are basically those of the current generator, and it has a much wider bandwidth. In general, a bandwidth from DC up to 100kHz can be obtained with these transducers, with an accuracy of 0.1% of the rated current for normal applications.

64.5.3 Rogowski Coils

A Rogowski coil is a simple coil wound on a toroidal nonferromagnetic core. The working principle of Rogowski coils is based on the Ampere law (Fig. 64.19):

$$\oint_L \vec{H} \cdot d\vec{l} = \iint_S \vec{J} \cdot d\vec{s} \quad (64.48)$$

where \vec{H} is the magnetic field, L is a closed line, S is the surface enclosed in L, and \vec{J} is the current density vector.

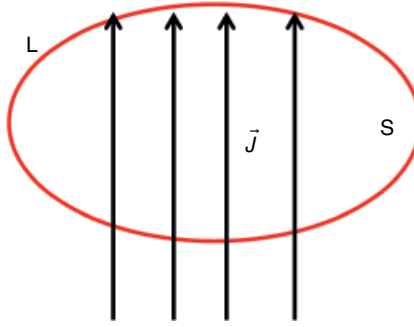


FIGURE 64.19 Ampere law.

Equation 64.48 can be rewritten as

$$\oint_L \frac{\vec{B}}{\mu_0} \cdot d\vec{l} = I_s \quad (64.49)$$

where I_s is the current flowing through surface S . Therefore by integrating the magnetic field along a closed line around a conductor, it is possible to evaluate the current flowing in that conductor.

Let us now consider a toroidal winding wound along line L . The voltage induced at the winding terminals will be (Fig. 64.20)

$$v_c = M \frac{di}{dt} - R_c i_c - (M + L_c) \frac{di_c}{dt} \quad (64.50)$$

where M is the mutual inductance between the winding and the conductor carrying current i , R_c and L_c are the resistance and the inductance of the coil, and i_c is the current flowing in it (Fig. 64.20).

Therefore if voltage v_c is measured with an ideal voltmeter, having infinite impedance, (64.50) can be rewritten as follows:

$$v_c = M \frac{di}{dt} \quad (64.51)$$

Hence by integrating the voltage induced in the coil, it is possible to evaluate the current i flowing in a conductor crossing the toroid.

The Rogowski coil can suffer from crosstalk due to a magnetic field produced by currents not crossing the coil. In fact an external magnetic flux can couple the coil through the central hole of the toroid. In order to reduce this effect, a compensating turn is placed into the toroid and connected in series with the main coil to compensate for the voltage induced by external fields (Fig. 64.21). Another, more accurate method to reduce crosstalk requires to wound an even number of layers of turns and connecting them in series.

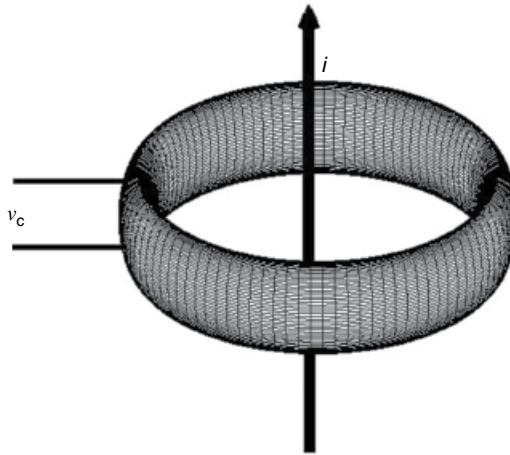


FIGURE 64.20 Rogowski coil.

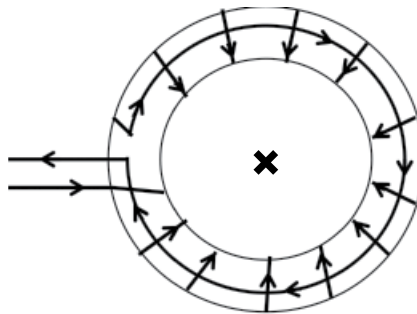


FIGURE 64.21 Compensation loop.

The Rogowski coil does not saturate and is characterized by wide bandwidth and range of measurement. Moreover it guarantees galvanic insulation. Typical values are a bandwidth up to 1 MHz, primary currents up to MA, and insulation levels up to several hundreds of kilovolt.

64.5.4 Voltage Transformers

The most widely used voltage transducer is the voltage transformer (VT). Similarly to the CT, this transducer is basically a transformer that performs a reduction of the value of the quantity to be measured. In particular it performs a known reduction of the amplitude of an AC voltage, in order to adjust it to the input dynamics of the employed voltmeters. Like the CT, the VT provides galvanic insulation between the power network and the measurement equipment.

The primary winding of the VT is connected in parallel to the power line, and its secondary winding, in ideal conditions, is left open (Fig. 64.22).

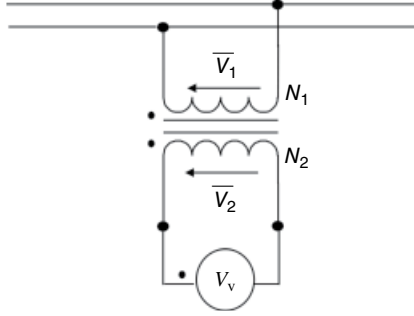


FIGURE 64.22 Ideal VT.

The voltage V_v measured by an ideal voltmeter is

$$\bar{V}_v = \bar{V}_2 = \bar{V}_1 \frac{N_2}{N_1} = \frac{1}{K_T} \bar{V}_1 \quad (64.52)$$

where N_1 and N_2 are the numbers of turns of the primary and secondary windings, respectively. In ideal VTs, the theoretical constant K_T is a real constant, and, therefore, the secondary voltage is a scaled replica of the primary voltage. For safety reasons, VTs require the connection to ground of one of the secondary terminals to avoid the presence of dangerous common mode voltages on the secondary windings. The rated secondary output of VTs is 100 V.

64.5.4.1 Measurement Errors The analysis of the VT measurement errors can be done by considering the electrical model of a real VT under no-load conditions (Fig. 64.23).

In a similar way as that followed for the CTs, two kinds of errors can be defined by means of the phasor analysis: the phase error ε and the ratio error η . In fact, by composing the voltage drops characterizing the VT, the graph reported in Figure 64.24 can be easily plotted.

By analyzing the graph, under the condition of small phase error ε , this same error can be defined as

$$\varepsilon \cong \sin \varepsilon = \frac{|\bar{Z}_1 \cdot \bar{I}_0|}{V_1} \sin \theta \quad (64.53)$$

$$\theta = \varphi_0 - \varphi_1 \quad (64.54)$$

$$\bar{Z}_1 = R_1 + j\omega L_1 \quad (64.55)$$

where φ_1 is the phase angle of \bar{Z}_1 and φ_0 is the angle between \bar{I}_0 and $K_T \bar{V}_2$.

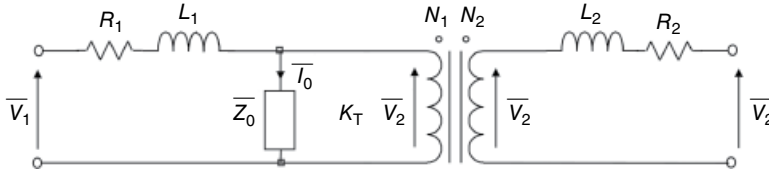


FIGURE 64.23 Electrical model of a VT.

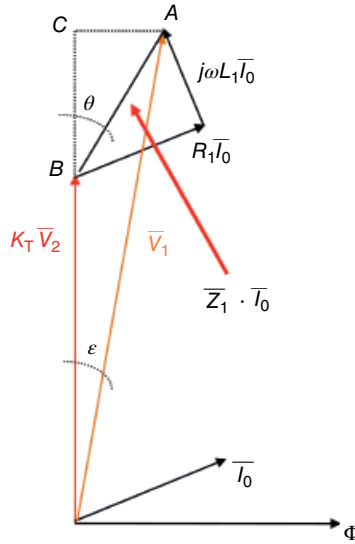


FIGURE 64.24 Phasor graph of a VT.

As for the CT, also the ratio error can be defined as

$$\eta \cong \frac{K_r - K_T}{K_r} - \frac{|\bar{Z}_1 \cdot \bar{I}_0|}{V_1} \cos \theta \quad (64.56)$$

$$K_r = \frac{V_{1r}}{V_{2r}} \quad (64.57)$$

Equations 64.53 and 64.56 show that the VT errors depend on its magnetizing current and the primary series parameters of the transformer. Therefore the improvement of the VT performances requires the optimization of the magnetic core, the numbers of turns of its windings, and the primary winding parameters. Equation 64.56 also shows that the ratio error can be zeroed, for a given voltage, by properly defining the rated ratio K_r of the VT.

In the performed analysis, the ideal no-load condition was considered. However, the measurement of voltage \bar{V}_2 implies the connection of a voltmeter, which can be represented by its internal impedance \bar{Z}_b (Fig. 64.25).

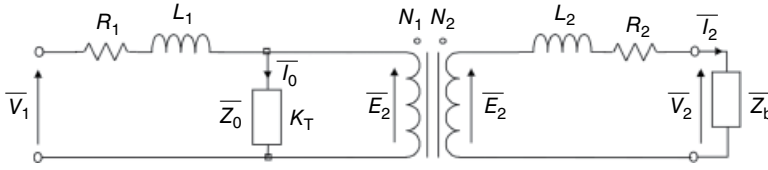


FIGURE 64.25 Electrical model of a loaded VT.

It is clear that the presence of a load modifies the working condition of the VT. In particular there will be now a secondary current \bar{I}_2 that causes a voltage drop on the secondary series parameters, thus increasing the error:

$$\bar{V}_2 = \bar{E}_2 - \bar{I}_2 \cdot (R_2 + j\omega L_2) = \bar{E}_2 \cdot \left(\frac{\bar{Z}_b}{\bar{Z}_b + (R_2 + j\omega L_2)} \right) \quad (64.58)$$

In addition the secondary current causes an increment in the primary current, thus increasing the voltage drop on the primary series parameters.

Similarly to CTs, also VTs suffer from the problems of the core saturation and limited bandwidth. In general, their bandwidth is upper limited to 1 kHz.

64.5.5 Electronic Transformers

In the last decade, stricter requirements in terms of bandwidth, related to power quality measurements, as well as the need of new functions, have pushed toward the development of new voltage and current transducers. Nowadays, thanks to the use of electronics, it is possible to guarantee the performances in terms of insulation and accuracy avoiding the use of the traditional CTs and VTs. Moreover the introduction of electronics opens the way to the implementation of functions such as measurement synchronization, diagnostics, measurement preprocessing, etc. For these reasons, two standards have been issued: IEC 60044-8 for electronic current transformers (ECTs) [8] and IEC 60044-7 for electronic voltage transformers (EVTs) [9]. These standards provide the requirements and the characteristics that the new transducers must satisfy, regardless of the employed technology.

One of the main differences between CTs/VTs and ECTs/EVTs is related to their output signals. Electronic transformers (ETs) can have two kinds of output signals: analog and digital. Moreover the analog output can have a rated value different from that of the classical CTs and VTs.

Another significant difference is related to the introduction of the concept of delay. The ET output can have a rated time delay different from zero. The ET errors are computed taking into account the rated delay. This new concept is related to the digitalization and transmission of the information. The maximum phase and amplitude errors are specified not only for the mains frequency but also for its harmonics. The errors are defined according to the aimed function: measurement or protection.

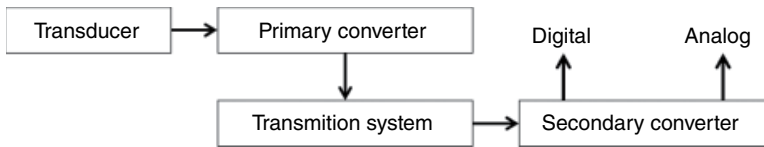


FIGURE 64.26 ET general architecture.

The typical architecture of ETs is shown in Figure 64.26.

By means of this architecture many solutions can be obtained, according to the chosen technology. As an example, by choosing as primary converter an analog-to-digital converter, the transmission system can be based on fiber optics, thus assuring the required insulation in a rather immediate way. In the secondary converter, a digital-to-analog converter can be present, as well as a digital interface. Computation capabilities can be placed in different parts of the ET, in order to perform signal preprocessing as, for instance, filtering, parameter computation, etc.

The architecture of Figure 64.26 can be modified by adding as many transducers as necessary, for instance, for implementing a combined three-phase measurement transformer. In this case the transformer will be able to measure simultaneously three currents and voltages as well as to compute active power, power factors, etc.

The kinds of transducers that can be used for the implementation of an ET are significantly higher than those used with the traditional transducers, thanks to the described architecture. In fact, by using a primary converter based on electronics characterized by stable and easily tunable input impedance, it is possible to use wideband sensors such as resistive shunts and Rogowski coils for the current and capacitive voltage dividers for the voltage. The insulation is ensured by the communication channel.

ETs can easily feature bandwidth up to 100 kHz, rated input voltage up to 500 kV and rated input current up to 10 kA, accuracy of 0.1% of the rated primary value, and an insulation level up to hundreds of kilovolts.

64.6 POWER QUALITY MEASUREMENTS

The previous sections have mostly dealt with power and energy measurements under AC sinusoidal conditions, since nowadays electric systems are generally operated under these conditions. However, the situation is slowly changing, due to the widespread use of power electronic devices in power systems. High-power power electronic converters are expected to increase in number and power, in the new smart-grid scenario, when energy production from renewables will play a significant role, and, consequently, generation will become more and more distributed.

There are two consequences of this new scenario: the energy flow will become bidirectional in almost every branch of the grid, and the voltage and current signals will become distorted, thus causing the sinusoidal conditions to be abandoned.

The first problem can be easily solved with a structure such as the one shown in Figures 64.10 and 64.26. This structure can be readily equipped with proper algorithms capable of accounting energy flowing in both directions. Moreover, being based on a digital processor, more functions, such as connections to a remote center, can be implemented, thus obtaining the so-called smart meters, which are the natural evolution of wattmeters and energy meters in the smart-grid scenario.

The second problem is definitely more critical and gives rise to the so-called power quality problems.

It is well known that power electronic devices are strongly nonlinear and also, sometimes, time variant. Therefore, they inject periodic disturbances on the line current, and these disturbances can be synchronous with the fundamental frequency (harmonic disturbances) or asynchronous with the fundamental frequency (subharmonic or interharmonic disturbances).

Since the short-circuit power in a given network section is never infinite, current disturbances causes voltage drops on the equivalent source impedance, so that also the voltage, in the same given section, shows the same kind of distortion as the current [10]. This causes power components to be associated with the frequency-domain components (harmonic and nonharmonic) of voltage and current.

While the active power, being defined as the average value of the instantaneous power, keeps its physical meaning and can still be measured by instruments based on (64.38), provided they feature enough bandwidth,⁵ all other power components lose the physical properties they show under sinusoidal conditions [10].

The first attempts to extend the reactive power definition to the nonsinusoidal conditions go back to the early decades of the twentieth century and are due to Budeanu [11] and Fryze [12]. These two approaches to power components definition under nonsinusoidal conditions have been widely discussed in the second half of the twentieth century and sometimes strongly criticized [13], but no final word has been said on this issue yet [10, 14]. A recent IEEE standard, the IEEE Std. 1459-2010 [15], gives some interesting definitions, but it still clearly states that “There is not yet available a generalized power theory that can provide a simultaneous common base for: Energy billing, Evaluation of electric energy quality, Detection of the major sources of waveform distortion, Theoretical calculations for the design of mitigation equipment such as active filters or dynamic compensators.”

Unfortunately, most definitions available in the literature⁶ try to extend the reactive power concept to the nonsinusoidal conditions and call the newly defined quantities always “reactive power.” The consequence is that different values can be measured,

⁵The bandwidth is not only ensured by a correct sampling frequency, as it is generally thought, but also by a correct choice of the voltage and current transducers. As shown in Section 64.5, transducers may have a dramatic impact on the performances of the instruments connected to their output terminals, both in terms of bandwidth and accuracy.

⁶A short survey of these definitions can be found in [14].

under the same distortion conditions, for, apparently, the same quantity, depending on the adopted definition.

Moreover, all definitions have been derived under the assumption that the instantaneous power can be defined by (64.11). It is worth recalling that this is true, provided that the highest-frequency components in the signal have wavelength far lower than the dimensions of the circuit. Considering that the switching frequency of power electronic devices is ever increasing, the presence of high-frequency components with nonnegligible energy cannot be excluded, in the future, on power systems.

It is then possible to conclude that the modern instruments for measuring electric power and energy, such as those based on the architecture shown in Figures 64.10 and 64.26, can virtually measure every newly defined quantity. The problem, as also reported in [16], is that there is no agreement, yet, on what to measure, on which frequency band, and with which accuracy.

Therefore, we limit this contribution to warn the readers that power quality measurements, unlike voltage and current quality measurements, are still largely undefined and require a significant additional research work to be agreed upon by the scientific and technical community. The key features, in this kind of measurements, are the measurement algorithms and bandwidth and accuracy of the voltage and current transducers. They are an important part of the whole measurement equipment and cannot be neglected when assessing the metrological performance of the employed equipment.

REFERENCES

1. R. A. Serway, J. W. Jewett, *Physics for Scientists and Engineers*, 6th Edition, Thomson Brooks/Cole, Salem, OR, 2004.
2. P. A. Tipler, G. Mosca, *Physics for Scientists and Engineers*, 6th Edition, W.H. Freeman and Company, New York, 2008.
3. M. Alonso, E. J. Finn, *Fundamental University Physics: Vol. 2 Fields and Waves*, Addison-Wesley Publishing Co., Reading, MA, 1971.
4. J. Bird, *Electrical Circuit Theory and Technology*, 5th Edition, Routledge, Oxon, 2014.
5. JCGM 100:2008, *Evaluation of Measurement Data—Guide to the Expression of Uncertainty in Measurement (GUM 1995 with Minor Corrections)*, Joint Committee for Guides in Metrology, 2008. Available online: <http://www.bipm.org/en/publications/guides/gum.html> (accessed November 7, 2015).
6. M. B. Stout, *Basic Electrical Measurements*, Prentice-Hall, Englewood Cliffs, NJ, 1960.
7. G. D'Antona, A. Ferrero, *Digital Signal Processing for Measurement Systems: Theory and Applications*, Springer, New York, 2006.
8. IEC, "Instrument Transformers—Part 8: Electronic Current Transformers—Edition 1.0", *IEC Std. 60044-8*, 2002.
9. IEC, "Instrument Transformers—Part 7: Electronic Voltage Transformers—Edition 1.0", *IEC Std. 60044-7*, 1999.

10. A. Ferrero, Measuring electric power quality: problems and perspectives, *Measurement*, Vol. 41, no. 2, 2008, pp. 121–129.
11. C. I. Budeanu, *Puissances reactives et fictives*, Inst. Romain de l'Energie, Bucharest, 1927.
12. (a) S. Fryze, Active, reactive and apparent power in circuits with non-sinusoidal voltage and current, *Przegl. Elektrotech.*, Vol. 7, 1931, pp. 193–203 (in Polish); (b) S. Fryze, Active, reactive and apparent power in circuits with non-sinusoidal voltage and current, *Przegl. Elektrotech.*, Vol. 8, 1931, pp. 225–234 (in Polish); (c) S. Fryze, Active, reactive and apparent power in circuits with non-sinusoidal voltage and current, *Przegl. Elektrotech.*, Vol. 22, 1932, pp. 673–676 (in Polish).
13. L. S. Czarnecki, What is wrong with the Budeanu concept of reactive and distortion power and why it should be abandoned, *IEEE Trans. Instrum. Meas.*, Vol. 36, 1987, pp. 834–837.
14. A. Ferrero, Definitions of electrical quantities commonly used in non-sinusoidal conditions, *Eur. Trans. Electr. Power*, Vol. 8, no. 4, 1998, pp. 235–240.
15. IEEE, “IEEE Standard Definitions for the Measurement of Electric Power Quantities Under Sinusoidal, Nonsinusoidal, Balanced, or Unbalanced Conditions,” *IEEE Std 1459-2010 (Revision of IEEE Std 1459-2000)*, 2010.
16. A. Ferrero, “Measurements on Electric Power Systems: Are We Prisoners of Tradition?” *IEEE International Workshop on Applied Measurements for Power Systems (AMPS)*, Aachen, Germany, September 25–27, 2013, pp. 120–125.

PART VIII

CHEMISTRY

AN OVERVIEW OF CHEMOMETRICS FOR THE ENGINEERING AND MEASUREMENT SCIENCES

BRAD SWARBRICK¹ AND FRANK WESTAD²

¹*Quality by Design Consultancy, Sydney, New South Wales, Australia*

²*CAMO Software AS, Oslo, Norway*

65.1 INTRODUCTION: THE PAST AND PRESENT OF CHEMOMETRICS

The term chemometrics was coined by Svante Wold in 1971 in a grant application [1], and soon after, the International Chemometrics Society (ICS) was formed. The ICS defines chemometrics as follows: “Chemometrics is the science of relating measurements made on a chemical system or process to the state of the system via application of mathematical or statistical methods.” It must be noted here in the definition that chemometrics is an area of chemistry, not a branch of abstract mathematics, and it is by this definition that gives chemometrics a practical nature.

Chemometrics is fast becoming a commonplace set of tools used in many industrial and research environments. Its rise over the past two decades can be attributed to two major factors:

1. The speed of computers has improved exponentially and personal computers can store large quantities of data for analysis.
2. Analytical technologies, such as spectrometers and chromatographs, can generate large volumes of data per measurement that requires highly sophisticated methods of analysis.

Chemometric methods are based on the larger discipline of multivariate analysis (MVA), also known as multivariate data analysis (MVDA). MVA methods that are most applicable to chemometrics can be divided into three main categories:

1. **Exploratory data analysis (EDA):** Mathematical methods used to investigate the natural patterns that exist in a specified data set by the application of typically linear methods of analysis, whereby sample and variable relationships may be established and used for further applications or enhanced data insights. Exploratory methods are also known as unsupervised classification methods.
2. **Regression methods:** These methods aim to define a model that relates one set of variables, known as independent variables to a set of responses, or dependent variables. The quality of the model is determined by its predictive ability, and this is determined based on comprehensive statistics generated by chemometric models. The set of independent variables may itself be a compilation of variables from various types of sensors or those generated by a single multichannel instrument, such as a spectrometer.
3. **Classification methods:** Multivariate classification is also known as supervised classification and requires the definition of classification rules. These rules can be submodels developed in the EDA phase of data collection, and the quality of such a classification schema is determined again on predictive ability.

A key theme of the chemometric methods defined earlier is the model's predictive ability. It must be remembered that chemometric models are not hard models applicable in a general way but are empirical (soft) models. An empirical model is, by definition, an approximation made based on a limited set of available data. This gives rise to two of the most important concepts of chemometric modeling: representation and validation. These two topics will be discussed in great detail in this chapter.

Prior to the widespread availability of fast personal computers, chemometrics was a topic reserved for the academic community. What are simple calculations by today's standards that are complete in a matter of minutes or seconds could have previously taken hours or days to complete. Even in the early days of the personal computer, the analysis of relatively small data sets could take hours to generate results. The advent of the Microsoft® Windows platform also contributed to the acceleration of the usage of chemometrics in the past two decades. This is because of the better graphics capabilities of current computer operating systems. Since chemometrics is a highly visual analysis method, high definition and superior graphical capabilities enhance the interpretability of chemometric models.

The current challenge that needs to be addressed by chemometric methods is how to handle the ever-growing data sets being generated by modern instrumentation and manufacturing systems. Algorithms that can better utilize processor cores and grid computing will better handle large data sets; however, many chemometric methods were initially implemented with algorithms based on deflation. By this it meant that in order to complete an entire analysis, the process must be performed in a sequential

manner; thus, the calculation of the next step of the process is dependent on the last one completing. There is no increase in speed obtained as the calculation proceeds, so if the data set is large, the calculation time will increase as a function of size. Over the years, algorithms tailored to specific matrix dimensions have been developed [2, 3].

A related data analysis methodology to chemometrics is design of experiments (DoE), which has received much attention in the literature and is seen as its own discipline distinctly different to chemometrics, although the methods share some commonality. As many DoE studies also involve several responses and other variables (covariates) that may be correlated internally and with the design factors, the generic multivariate methods play a role in a more holistic approach. The interested reader is referred to the extensive literature that is available on DoE [4–7].

Terminology is also a major challenge in chemometrics and Section 65.8 provides a detailed review of the most commonly used terms in chemometrics.

65.2 REPRESENTATIVE DATA

It cannot be stressed more that good data results in good models provided there is information in the data to be modeled in the first place! Put another way, the old adage “garbage in–garbage out” holds well when applied to chemometric modeling. By good data, it is meant that they are representative of the situation to be modeled. Multivariate methods aim to describe the variance that exists in a selected data, and this variation comes from two main sources:

1. **Explainable:** Systematic variations within the data that can be potentially modeled using multivariate (or other) techniques.
2. **Unexplainable:** Random variations in the data that cannot be modeled.

The terms explainable and unexplainable are commonly used in the statistical process control (SPC) literature [8] for detecting common cause events that are different from the random (noise) within the measurement system. This also relates back to a highly important method of statistical analysis called analysis of variance (ANOVA). One-way ANOVA is used to determine whether the between treatment averages are significantly different than the between sampling averages. Put into mathematical terms,

$$SS_T = SS_{\text{Between}} + SS_{\text{Within}}$$

where the following equation terms are defined:

SS_T = Total variance within the data set that can be explained (i.e., 100% total)

SS_{Between} = The variance explained by the model that can differentiate between two or more treatment levels

SS_{Within} = The variance between the sample replicates within a treatment level, that is, the precision of the measurements

It naturally follows on from this equation that the total variability in a data set is the sum of what can be explained and what cannot be explained by the model. Therefore, the ANOVA expression can be rewritten in nonmathematical terms as

$$\text{DATA} = \text{INFORMATION} + \text{NOISE}$$

How much information can be extracted from the data is a function of:

1. How representative the data set selected is of the problem to be solved.
2. The quality of the measurement system used to measure the samples in the data set.
3. The form of the model used (i.e., if a linear model is used to measure nonlinear data, then the information will be lower due to what is known as lack of fit).

Sample representation can take on a number of forms when developing a chemometric model. First and foremost, great attention is required when selecting a sample set to model that is going to be characteristic of future samples. If a crystal ball was a reality, then this would be a simple process; however, the development specialist does not have such a luxury, so in many cases, models have to be developed in an iterative manner. In particular, when a model is to be developed on natural samples, such as agricultural products, biological specimens, or soils, these samples must be selected to span a range of one or more properties. This may take some time to build a reliable and robust calibration and validation set of data since new samples are typically found by chance during model development. Another challenge of natural samples is their inherent heterogeneity, that is, the sample has different characteristics depending on where it is measured.

One excellent example of heterogeneous sampling is encountered when developing predictive models of agricultural products using near-infrared (NIR) spectroscopy. In particular, if a model for predicting the protein content of wheat is to be developed, an analyst should aim to collect a number of samples of various protein levels, over a number of growing seasons and regions, etc. Once the pool of samples is collected, they must be split in some way as to obtain a representative split and then each sample must be scanned numerous times to average out packing differences and local heterogeneity. Again, it is stressed that representative sampling is the most critical part of chemometric model building.

In industries where samples can be artificially manufactured, the development of calibration samples that span large regions is possible through careful experimental design. Industries such as the pharmaceutical sector have the ability to develop robust samples that exceed the normal tolerances of typical manufacturing targets, and using pilot scale equipment (or using production equipment at the end of a batch run) can develop representative samples that can be used to develop reliable models.

The next step in model development requires the selection of samples suitable for calibration development. In chemometrics there are two main strategies used to test the reliability of a model for future use on new samples:

1. **Test set validation:** This requires a rational sample selection method to separate the sample pool into a calibration (training) set and a validation (test set). More details on this approach are provided in Section 65.6.1.
2. **Cross validation:** Typically, cross validation is used when there are not enough samples available in the pool to create a robust model and test it against a representative set. There are a number of ways cross validation can be used and these are further discussed in Section 65.6.2.

Finally, when the samples have been found to span a suitable region for model development, the measurement system has been found to provide reliable data, and a method of rational sample selection has been decided upon, the next step is to ensure that the samples represent the entire space of the model developed and this is dependent on the modeling strategy to be employed. In the case of exploratory models (see Section 65.3.3), only the independent X -variable space needs to be spanned in order to develop robust models. When regression models are to be developed, the calibration and validation set must representatively span both the X - and Y -variable (response) space. More details of this will be presented in Section 65.6 on validation.

65.2.1 A Suggested Workflow for Developing Chemometric Models

Representation is not only applicable to sample selection but is also applicable to the measurement systems used to measure the samples. If a sensor or spectrometer response is unreliable, no matter how representative the samples are, the data are bad and therefore a reliable model cannot be developed. Table 65.1 provides a quick checklist to follow when selecting a representative data set and developing a robust chemometric model.

65.2.2 Accuracy and Precision

In the development of any analytical procedure, the concepts of accuracy and precision are extremely important for assessing the model's ability to perform its task. The International Conference on Harmonization (ICH) [9] has developed a document that is a useful guide when developing quantitative methods of analysis. Of the main aspects of model validation, the following are deemed to be critical for reliable model development:

1. **Accuracy:** How close the predicted value of the new method is to the reference value (when a validated reference method exists).

TABLE 65.1 Suggested Workflow for Developing Chemometric Models

Task	Suggested Approach
Select samples to build the model such that they span future sample ranges	<p>Start with a small set, when the nature of the samples is unknown, and use a nondestructive (where possible) method to find samples different from the initial set before much effort is put into model development</p> <p>For situations where artificial samples can be produced, create a small set of samples, measure them, and check that the samples made to target specifications have similar characteristics to samples made using the actual manufacturing process</p>
Ensure measurement systems are reliable	<p>When using measurement systems, perform gage R&R or some form of measurement systems analysis (MSA) to check the quality of measurements made [8]</p> <p>Ensure instruments such as spectrometers and chromatograms have been calibrated and qualified before use</p>
Ensure that measurements made are repeatable and reproducible	<p>Measure a single sample multiple times and visually assess results. Perform ANOVA where needed in order to determine if multiple measurement with averaging is required before a model is developed</p> <p>After a suitable sample averaging method has been defined, compare multiple sample averages for consistency of results</p> <p>If the sample is unstable or requires an exact measurement window for reliable results, automate the sample collection process as much as possible and send the samples for reference analysis as soon as possible (if required)</p>
Reference method analysis	<p>The reference method is the “gold standard” when developing quantitative methods of analysis. Theoretically, the secondary method cannot be more accurate than the reference method; however, the secondary method can be more precise (see Section 65.2.2 for more details)</p> <p>It is important to remember that the sample measured by the secondary method is the one analyzed by the reference method. In cases where the sample measured by the secondary method is too large to be measured by the reference method, a suitable sample splitting method with replicate analysis may be required, depending on the heterogeneity of the sample being measured</p>
Validate the model and interpret it	<p>Depending on the context the model is to be used for, it is always suggested to validate a model using an independent test set. There are a number of sample selection strategies available to create representative training and test sets (see Section 65.6.1.2), and this will provide the most reliable estimate of the models future performance</p>

TABLE 65.1 (Continued)

Task	Suggested Approach
	Parsimony is a key term used in model development, and in general terms, the simpler the model, the more interpretable it is. A model that cannot be interpreted on a physical, chemical, or biological level should not be used for practical purposes
Implement the model	A model is only of use if it can be interpreted and, more importantly, implemented for practical usage in a production or research environment

2. **Precision:** How close a set of replicate measures are to each other. Precision is further subdivided into,
 - (a) **Repeatability:** How close a set of predicted values is to each other when the same sample is measured multiple times (with replacement) using the secondary method. Repeatability is a measure of the inherent noise in the measurement system.
 - (b) **Intermediate precision:** This is a measure of reproducibility. Intermediate precision is a measure of the sampling error between different analysts or different measurement systems (or a combination of both) and estimates the simplicity of the method usage.
3. **Robustness:** A measure of how sensitive the method is when small but deliberate changes are made to the system. This may include turning on/off a lamp in a spectrometer during analysis or even changing the lamp to assess the impact on predicted values.

Other criteria important to model development are linearity, range, limit of detection (LOD), and limit of quantification (LOQ). These must be assessed based on criticality (particularly LOD and LOQ when predictions are to be made close to the dynamic limits of the measurement system). These points will be expanded upon in Section 65.4.4.7.

ICH Q2(R1) [9] also states that once intermediate precision has been established, then accuracy is inferred.

What does this all mean in terms of chemometric model development? Without accurate and precise reference methodology, it is highly unlikely that the model will be reliable. In business critical operations such as in pharmaceutical or biopharmaceutical applications, the inability to establish accuracy and precision invalidates a model completely. Accuracy and precision are easily visualized using the “dartboard” principle. Figure 65.1 shows the principles of accuracy and precision and why they are important for reliable model development.

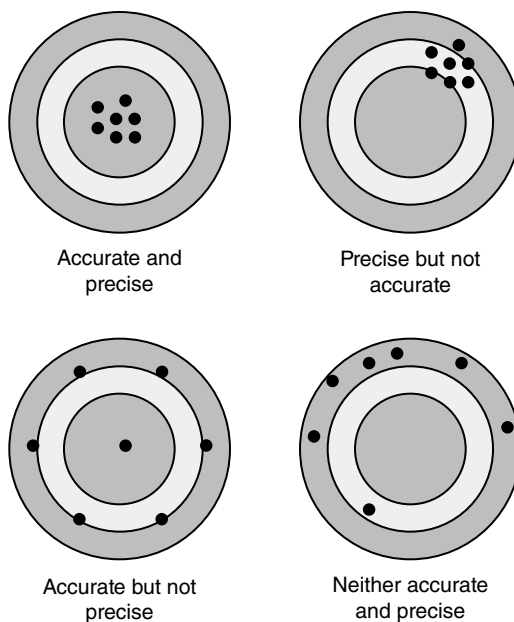


FIGURE 65.1 Diagrammatic representation of accuracy and precision.

65.2.3 Summary of Representative Data Principles

It was the intent of Section 65.2 to stress the importance of sound sampling and preparation before any attempt is made to develop reliable chemometric models. Ninety percent of the effort required to build a model was described in this section and a proven approach to development was presented.

Overall, reliable models are simple models, are easy to interpret, and can be easily root cause analyzed in the event of an outlying result. This is the result of good planning, understanding of the system being investigated, and putting in the effort to build robustness into the model from the start. As mentioned earlier, the principle of garbage in–garbage out must be acknowledged and this is where subject matter expertise is required to minimize the risk of modeling pitfalls.

The following points summarize the prerequisites of reliable model development:

1. Select a secondary method that is fit for purpose, that is, run a small feasibility set to ensure the system is capable of measuring the properties of interest.
2. Understand the complexity of the sample being measured. Heterogeneous samples will require more replicate measurements to be made for both the reference and the secondary methods in order to generate accurate and reliable results.
3. Ensure all equipment used is calibrated and in good working order. Sound obvious? It is rarely established in practice!

4. Use sound validation principles to generate the simplest models and interpret the model on either a physical, chemical, or biological level (this is the topic of Section 65.6).
5. Implement the model and learn from its usage.

65.3 EXPLORATORY DATA ANALYSIS

EDA is typically the first step in any data analysis problem. It allows an analyst to get an initial feed for a data set, particularly the distribution of the data for each of the variables. Section 65.3.1 discusses the univariate approach to data analysis (and some of its pitfalls) and describes why a multivariate approach is to be preferred. There are a number of methods available for EDA; however, the most commonly used methods are discussed in further depth in the following sections. The key point to remember is that EDA is unsupervised in nature, that is, the analyst is looking for natural patterns in the data. Once these patterns are established, rules can be developed and used to classify new samples. This is called supervised analysis or pattern recognition. Supervised methods are discussed further in Section 65.5.

65.3.1 Univariate and Multivariate Analysis

Prior to the development of many chemometric models, an investigative analysis is usually performed to gain some initial insights into the data structure. Scientists and engineers are typically taught in undergraduate programs to investigate raw data using simple charting techniques to look for “obvious” trends. While these simple plotting tools are powerful when analyzing variable 1 (or maybe 2 and 3) at a time, they become highly cumbersome when dealing with multivariate data.

Analyzing data, one variable at a time has been defined as the “scientific approach” to analysis. This only works if the variables are independent of each other, that is, there is little to no correlation between the variables. In Figure 65.2, two situations are provided that only take into account two variables at a time. This is a classical example and has been presented many times in the literature [8, 10] but serves as the best way of showing that for even the simplest of systems, the failure to analyze a data set multivariately may lead to false and sometimes fatal conclusions. It is assumed in this example that the reader is familiar with control charts used for SPC.

In case 1, two control charts, one for the temperature and one for the feed rate of a particular system, are provided. The control charts show that the variables appear to be in a state of statistical control. When the two sets of data are plotted point for point as a scatter plot, it can be seen that there is no linear (or other) relationship between the points, that is, the correlation is close to zero and therefore the two variables can be considered to be independent. In this case, the two sets of control

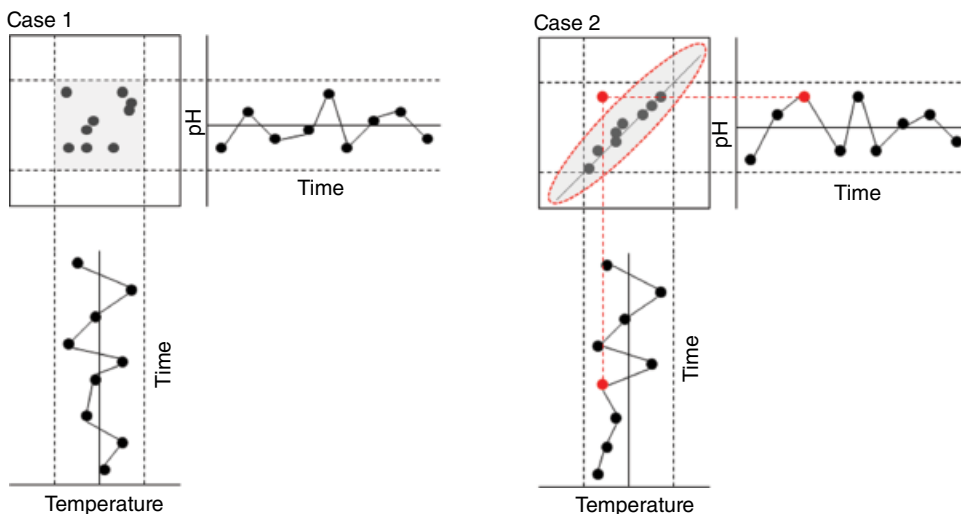


FIGURE 65.2 Simple data representations showing the multivariate nature of data.

limits bound the variability of both variables and anywhere inside the limit box are deemed to be in control.

Looking at case 2, in this case, two control charts, one of temperature and one of pH, are presented. As was the situation for case 1, the variables appear to be in a state of statistical control; however, there is an outlier in the data that cannot be seen univariately. When the data are plotted together as a scatter plot, note there is a linear relationship between the variables. Thinking about this from a chemical point of view, this makes sense as there is a scientific justification for why pH changes with temperature (i.e., use subject matter expertise to interpret the system). Note now in case 2, the limit box is no longer representative of the data and, more importantly, a visual outlier can be detected.

This occurs because the variables cannot be analyzed in isolation of each other. This is a simple case that shows a failure of the scientific approach when the variables are correlated. The control limits now change from being box shaped to be ellipse shaped. This is a key principle in MVA, that is, joint confidence intervals, and for this reason multivariate statistical process control (MSPC) (Section 65.7) is being widely adopted by industry for early event detection.

65.3.2 Cluster Analysis

The main goal of cluster analysis is to detect inherent patterns within a data set in an unsupervised manner. For the purposes of clarification:

1. Unsupervised methods look for natural patterns within the data in order to define possible classes and groups can be assigned based on known characteristics of the data set.

2. Supervised methods use so-called classification rules to separate new samples into predefined groups. Supervised methods are described in more detail in Section 65.5.

There are many types of cluster analysis methods available to the development scientist or engineer. These methods aim to find the similarity (dissimilarity) between samples in a data set, based on the variables used to measure the samples. Two of the simplest methods of cluster analysis available are known as K -means and hierarchical cluster analysis (HCA).

65.3.2.1 K -Means K -means is one of the most conceptually easy-to-understand approaches to unsupervised data analysis. It aims to separate samples into K -predefined classes with grouping based on the points being closest to a class centroid. As new points are added to a class, the centroid is recalculated and the points in the sample set are reassessed for class membership. Therefore, if the nearest neighbor distance is statistically exceeded, the sample is assigned to a new class and so on until all samples have been classified into the number (K) of predefined classes [11]. The algorithm works by minimizing the within-cluster sum of squares [12], thus allowing the definition of statistical limits for sample acceptance/rejection from a defined class. In all cases, K -means aims to partition all samples into one class only.

Adams [12] describes a four-step process for the K -means algorithm as follows:

Step 1: Define K clusters (K usually being a small integer) to group the data into and define any initial samples per cluster (should such class knowledge exist). Calculate the cluster means and the initial partition error.

Step 2: For the first sample, calculate the increase/decrease of the partition error by moving the sample into the classes defined. If the error is reduced by moving the sample to a particular class, keep it in that class; otherwise, leave it in its original class. Recalculate the class means each time a sample is moved to a new class.

Step 3: Repeat step 2 for all samples in the data set.

Step 4: If no samples have been moved, stop the process; otherwise, go to step 2.

Disadvantages of K -Means Clustering There are a number of disadvantages encountered when using K -means (or the related K -medians) methods. Firstly, the method requires an analyst to define the number of clusters to partition the samples before analysis begins. This means that if the first analysis is unsuccessful, many iterations of cluster definition may be required. Secondly, the final grouping of samples reflects the initial choice of clusters or initial samples chosen to define the first cluster centroids. Other disadvantages revolve around the distance measures used to determine the similarity/dissimilarity of samples for class assignment. The most commonly used distance metric is the Euclidean distance, but other methods such as the city block or

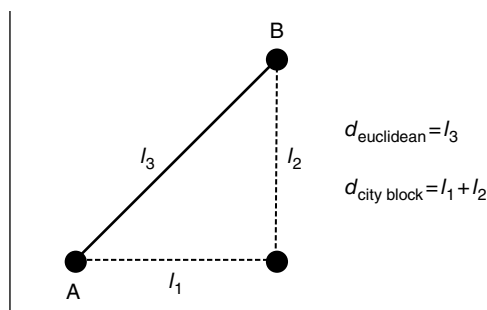


FIGURE 65.3 Euclidean and city block distance metrics.

correlation are available. Figure 65.3 provides a graphical example of the difference between the Euclidean and city block distance metrics.

It is not the intention of this chapter to discuss the algorithm details of the *K*-means algorithm, and the interested reader is referred to the text by Everitt [13] for more details.

An Example of K-Means Clustering One of the classical experiments performed for assessing the ability of clustering methods is Fisher's iris classification data set, first published by Sir Ronald Fisher in 1936 [14]. This is a simple multivariate classification problem and shows in simple terms how the *K*-means algorithm works but, at the same time, shows the limitation of the method, when the number of variables exceeds three.

For the purposes of simplicity, the data set is not repeated here and only a description is provided. The main aim of the experiment was to develop an objective method of classifying three types of iris, namely, *Iris setosa*, *Iris versicolor*, and *Iris virginica*, based on four easy-to-measure variables:

1. Sepal length
2. Sepal width
3. Petal length
4. Petal width

Data was collected on 150 samples (50 of each type) to generate a data table of dimension 150 rows by 4 columns. For each sample, the class name was assigned; therefore, the number of classes to define is $K=3$. Using the Euclidean distance measure, the classification rate is presented in Table 65.2.

Using the *K*-means method of classification with Euclidean distance, it can be seen from Table 65.2 that:

1. Setosa can be uniquely classified from versicolor and virginica.
2. Versicolor can be uniquely classified from setosa but can be confused 4 times in 100 with virginica.

TABLE 65.2 Confusion Matrix for Iris Classification Using *K*-Means (Euclidean Distance)

Predicted/Actual	Versicolor	Virginica	Setosa	Classification Rate (%)
Versicolor	48	2	0	96
Virginica	14	36	0	72
Setosa	0	0	50	100

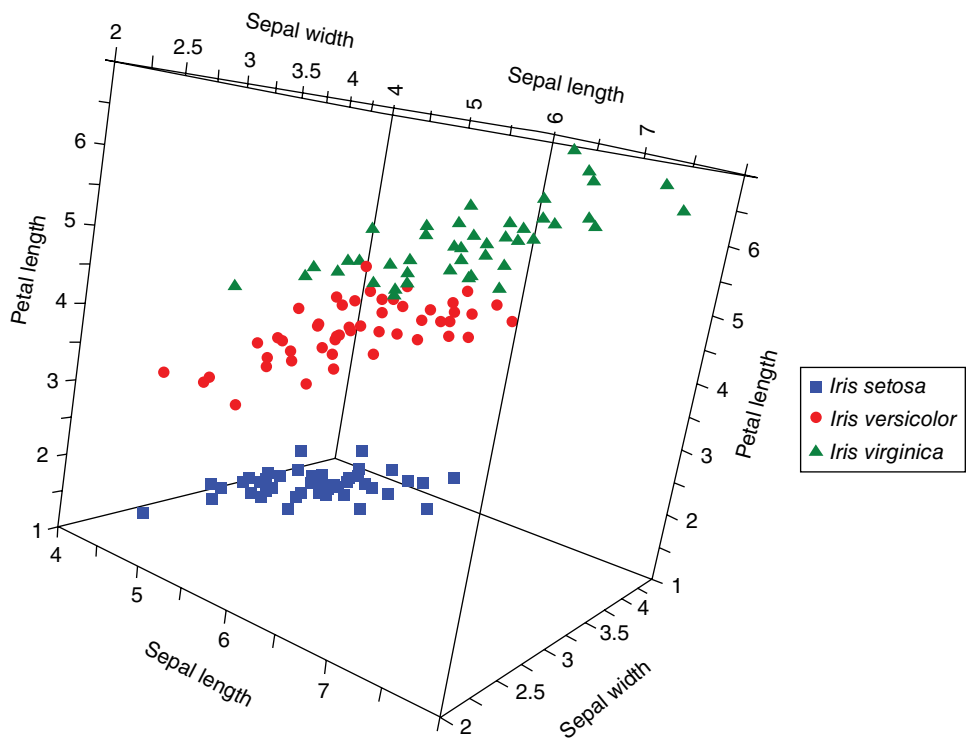


FIGURE 65.4 3D scatter plot of Fisher's iris data grouped by *K*-means clustering.

3. Virginica like versicolor can be uniquely classified from setosa but can be confused 28 times in 100 with versicolor.

The results presented in Table 65.2 can be improved by using correlation as the distance measure; however, these results are not presented here. To visualize the results of a *K*-means analysis, it is possible to plot the complete analysis as long as there are only three variables measured. In the Iris example, there are four variables; therefore, only three variables at a time can be shown. It is possible to plot multiple scatter plots; however, the process becomes extremely cumbersome when the number of variables becomes larger (typically >10). Figure 65.4 provides a three-dimensional (3D) scatter

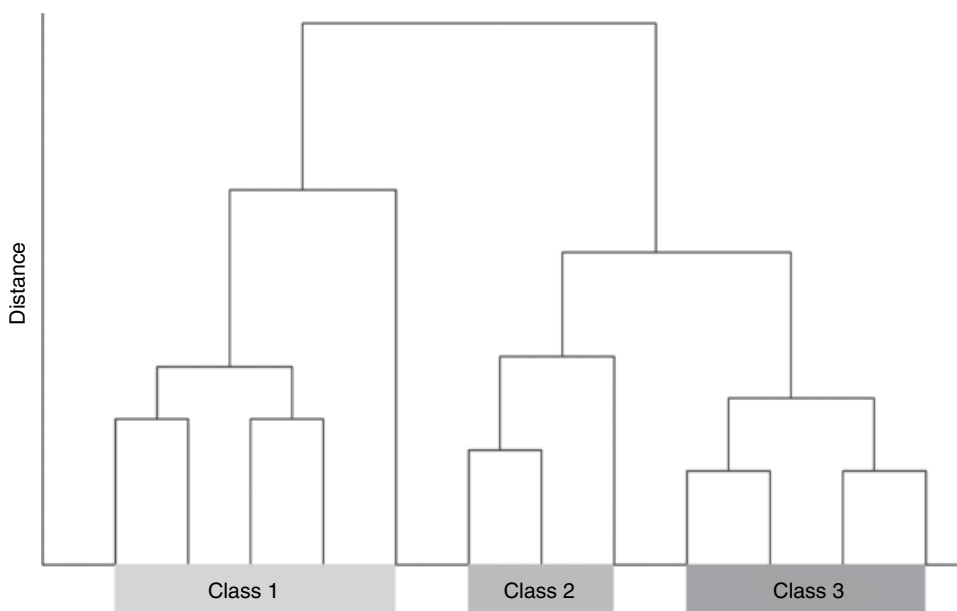


FIGURE 65.5 Example dendrogram.

plot of the variables sepal length, sepal width, and petal length with the samples assigned to their clusters.

It can now be seen in Figure 65.4 that the species *versicolor* and *virginica* are close to each other in properties and why the algorithm cannot completely separate the two classes (refer to the confusion matrix in Table 65.2).

65.3.2.2 Hierarchical Cluster Analysis Hierarchical methods aim to separate the original data into a few classes by either agglomerative or divisive methods [12]. Agglomerative methods fuse together smaller subclusters of samples and successively build to larger groups of samples, whereas divisive methods start with a single cluster and divide it into smaller clusters of similar samples.

As with *K*-means, a disadvantage of HCA is that the distance method has to be decided, and there are many to choose from as there are available HCA methods, including the well-known Ward's method [13]. The major advantage of HCA over *K*-means is that it provides a graphical display of the clusters known as a dendrogram. A dendrogram is a tree structure showing the linkages and similarity/dissimilarity of the samples. Figure 65.5 provides an example of a dendrogram.

It is not the intent of this chapter to provide extensive details on the principles of *K*-means and HCA, and the interested reader is referred to the excellent texts available by Adams [12] and Everitt [13] for more detailed discussions of these methods.

65.3.3 Principal Component Analysis

Principal component analysis (PCA) is a bilinear modeling method [14] that provides an interpretable overview of the main information contained in a multidimensional table. It is also known as a projection method, because it takes information carried by the original variables and projects them onto a smaller number of latent (or hidden) variables called principal components (PCs). Each PC explains a certain amount of the total information contained in the original data, and the first PC contains the greatest source of information in the data set. Each subsequent PC contains, in order, less information than the previous one. For PCA, the objective function is to maximize the variance for each subsequent PC. PCA is one of the most powerful EDA methods known, particularly because PCA models can be:

- Easily validated and interpreted
- Investigated using a wide range of graphical and diagnostic tools

The general PCA equation is as follows:

$$X = TP' + E$$

where

X is the original data to be analyzed

T is a matrix of sample structure information known as scores (Section 65.3.3.2)

P is a matrix of variable structure information known as loadings (Section 65.3.3.3)

E is a matrix of residuals that cannot be explained by the PCA model

Therefore, returning to the definition provided in Section 65.2,

X is the data component to be analyzed after centering and application-dependent scaling of the variables.

TP' is the information part of the analysis (i.e., what can be explained by the model).

E is the noise part of the analysis (i.e., what cannot be explained by the model).

To redefine the definition in Section 65.2, the PCA model becomes

$$\text{DATA} = \text{MODEL} + \text{ERROR}$$

This is the reason why PCA is known as a dimensionality reduction (or a decomposition method). It is an unsupervised classification method that aims to take large data sets and break them down into smaller yet more informative components (PCs) that are a collection of the most important sources of variability.

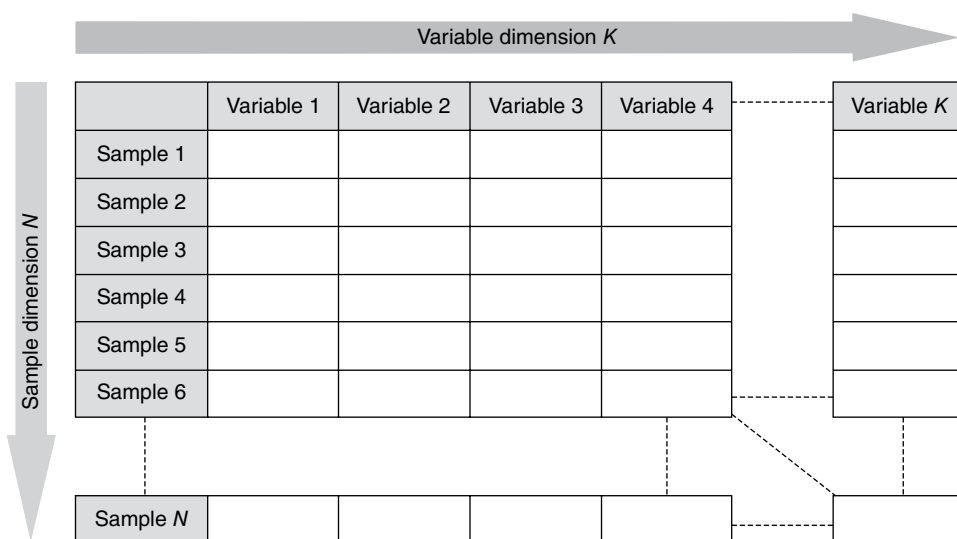


FIGURE 65.6 General table structure for PCA.

Mathematically, each PC is orthogonal to all other PCs and they can be conveniently plotted against each other on a Cartesian coordinate system. This provides PCA with one of the most comprehensive and powerful range of plotting capabilities available to any MVA method.

Before moving on to how PCA works in practice, the next sections define the terminology used when developing a model and how these terms relate to each other.

65.3.3.1 The PCA Problem and the Dual Nature of Data Samples and variables are not mutually exclusive, that is, a sample is characterized by the variables used to describe it and variables cannot be measured unless a sample is present. A data table suitable for analysis by PCA must be presented as a structured matrix. The general format of the matrix is the rows (N) represent individual samples and the columns (K) represent the variables measured on each table. Figure 65.6 shows the general table structure.

The matrix in Figure 65.6 consisting of N samples and K variables has dimension ($N \times K$). If $K > 3$, then the entire data set cannot be visualized using standard plotting tools since as humans, visualization in greater than 3-dimensions becomes difficult. What is meant by the dual nature of data is that samples can be plotted in variable space so that sample groupings and other sample relationships can be investigated. Conversely, variables can be plotted in sample space to understand the relationship between variables. The biggest challenge of the aforementioned approach is that it is equivalent to a univariate analysis of multivariate data when N and $K \gg 3$.

When samples and variables are plotted in the ways described previously, an analyst can get a small insight into the correlation structure of the data. PCA extends on this

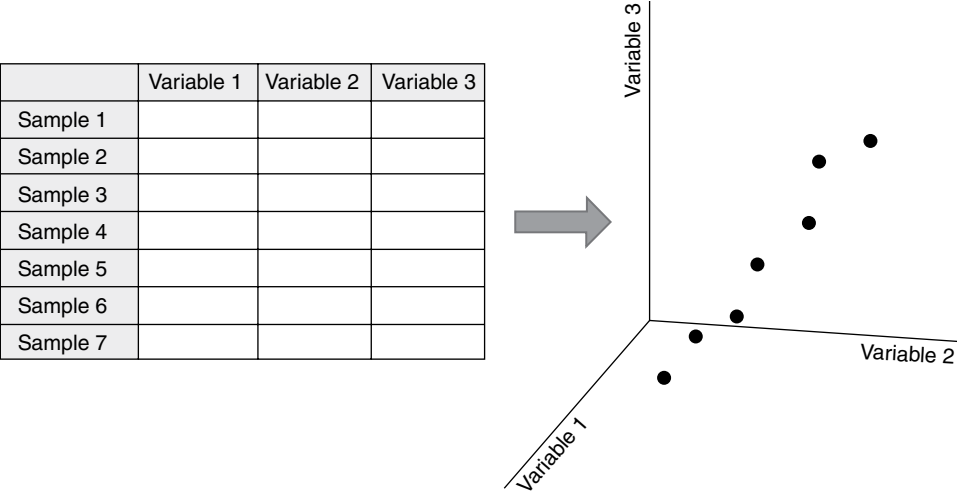


FIGURE 65.7 Plotting samples in variable space to understand the latent structure.

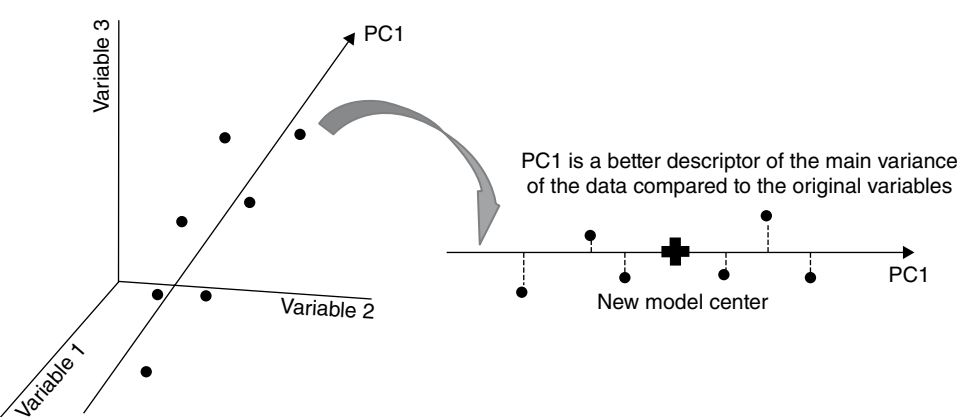


FIGURE 65.8 Fitting the first principal component to a data set.

principle by finding the main sample and variable relationships in multivariate data and reducing them down into simpler and more interpretable PCs. Consider a matrix with $N > 3$ and $K = 3$. This means that the entire table can be plotted in a 3D scatter plot. Consider the hypothetical data set shown in Figure 65.7.

PCs are sometimes referred to as latent variables. By definition, latent means hidden; therefore, PCA aims to find the hidden structure within a data set that may not be obvious from a simple univariate analysis. Note the points plotted in Figure 65.7 are not random but have a distinct structure in 3D space. By fitting the least squares line through the greatest direction of variability shows that there is a better direction in space that describes the data than the original three variables. This is shown in Figure 65.8.

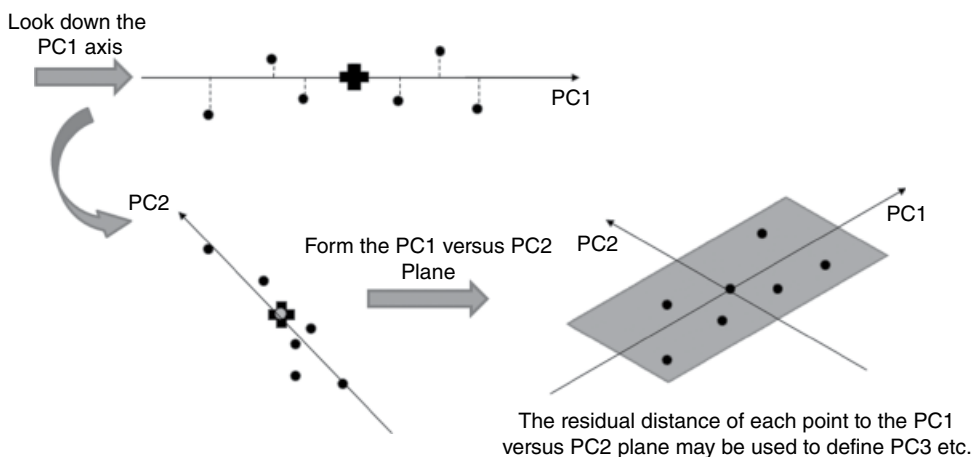


FIGURE 65.9 Fitting the second principal component to a data set and forming the PC1 versus PC2 plane.

PCA aims to describe the greatest sources of systematic variation in a data set. In the 3D data in Figure 65.8, this describes the direction of greatest elongation in the data. This direction is known as the first PC. When displayed in typical software packages, the first PC is plotted as the new x -axis of the data. This is also shown in Figure 65.8.

To find the next greatest source of variability with the constraint that the next PC is orthogonal to the first, the next PC is estimated from the remaining variance. To visualize this, the data must be conceptually viewed down the axis of the first PC. This is shown in Figure 65.9.

The data are now distributed around the first PC axis and a least squares line is fitted through the next greatest source of variability. The combination of PC1 and PC2 now forms a new plane in 2-dimensions. This is also shown in Figure 65.9.

Since the conceptual data set being analyzed has an original dimension of three, in order for PCA to be effective, one or two PCs maximum should describe the data with minimal error, that is, in a 2-PC model, all points should fit closely to the surface of the PC1 versus PC2 plane. If the data still do not fit the model, then the information in all three variables is contributing to the data set. This is called a full rank problem. If, however, two PCs describe the majority of the data and only random fluctuations occur around the PC1 versus PC2 plane, then the rank of the data set is 2 (compared to the dimension of the data $K=3$).

Plotting the sample information in PC space leads to the so-called scores plot of the data. When the variable information is plotted in PC space, this leads to the so-called loadings plot. Scores are discussed in more detail in Section 65.3.3.2 and loadings are discussed further in Section 65.3.3.3.

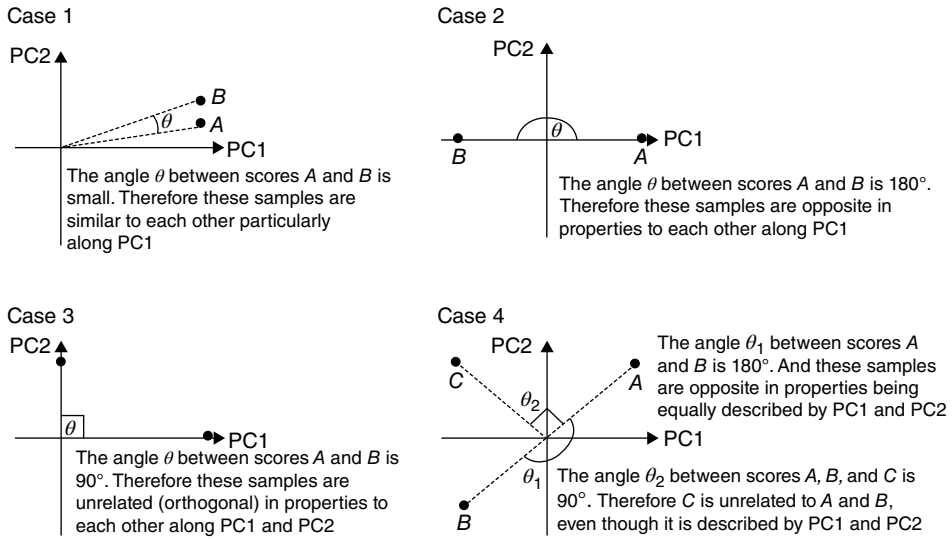


FIGURE 65.10 Sample relationships along a selected PC axis.

65.3.3.2 PCA Scores The PCA method was first introduced in the early 1900s by Pearson [16] and was later adapted for the analysis of psychology data and the terminology has remained to this date. The term scores relates to the samples' importance in describing the variability of a data set. Consider again the conceptual data set introduced in Section 65.3.3.1. If the first PC is plotted in the original variable space, the samples distribute themselves along this direction as shown in Figure 65.8.

A score (t) is the orthogonal projection of each sample onto the selected PC axis. The length of the distribution of the samples along the axis is proportional to how much information is contained in the PC. When the first PC is fitted to the data, a new model center (or origin) is established. This is the point of minimum variability in the data set, that is, samples at or close to the origin are not well described by the model. PC space covers both positive and negative regions along the PC axis. Along any given PC axis, the following holds. These properties are presented graphically in Figure 65.10 with a short explanation of each case:

1. Samples that group close to each other are similar (i.e., share the same pattern for the variables, case 1 in Figure 65.10).
2. Samples that lie close to the origin show the least variability compared to samples at the extremes of the PC axis (i.e., they are average in characteristics).
3. Samples that lie at one extreme of the PC axis are systematically different for the variables that contribute to this PC compared to samples lying at the other extreme (case 2 in Fig. 65.10).

When the next (and subsequent) PC is added to a model, higher-dimensional scores plots are possible. In the case of the two-dimensional (2D) scores plot, the following relationships between samples hold:

1. Along any selected PC, the rules as aforementioned hold.
2. Any samples lying exactly along one PC (but not at the origin) lie perpendicular to any samples that are exactly on the second PC. Based on simple geometry, the directional cosine between such samples is zero; therefore, it can be concluded that the sample types are independent in character from each other (case 3 in Fig. 65.10).
3. Samples that occupy spaces in between the PC axes are influenced by variables that are important on both axes. Although the PC axes describe independent sources of variability, samples may be influenced by more than one PC. There is one score for each sample as for the individual variables and the scores can be described as “super variables.” To understand which variables contribute to individual PCs, interpretation of the loadings is required (case 4 in Fig. 65.10).

Overall, a scores plot is a map of sample relationships plotted in either 1-, 2-, or 3-dimensions. These plots are an excellent way to study sample relationships; however, this is where the dual nature of data becomes important. The samples can only group the way they do, based on the variables measured to characterize the samples. No interpretation of sample grouping is possible without an understanding of the variables and their relationships contributing to describing the samples. This is where the loadings plot described in Section 65.3.3.3 is required.

65.3.3.3 PCA Loadings Loadings (dimension P (where $P \leq K$)) for each PC, relate to the weighting placed on each variable for describing a particular PC. PCA loadings can be described as the individual contributions of the input variables for describing a sample set, and the following equation can be used to describe a particular PC:

$$PC_A = z_1x_1 + z_2x_2 + \cdots + z_px_p$$

where

PC_A is A th PC being investigated

z_p is the loading or variable contribution to variable x_p

x_p are the original variables used in the analysis

PCA is an empirical modeling method, which means that the quality of the model is limited by the quality and scope of the data used to construct it. As is the case of any statistical method, PCA is used to describe a smaller sample of a larger population. Consequently, the variable contributions may be completely different for a smaller

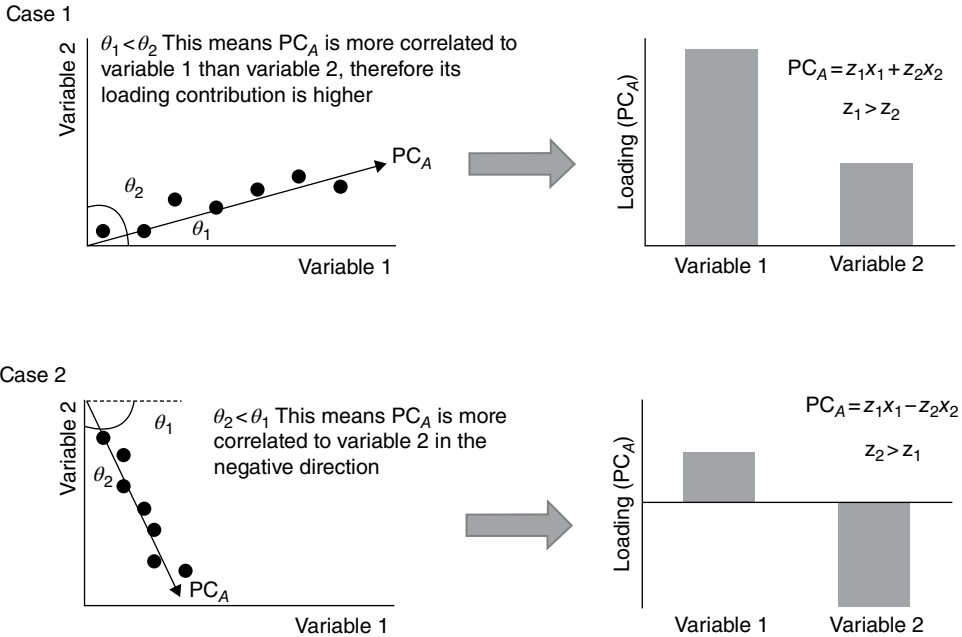


FIGURE 65.11 Definition of PC loadings in variable space.

data set compared to a larger data set (or the entire population). What does this mean then for the loadings? To interpret the variable contributions to a PC, the magnitude of each loading must be taken into account. Loadings are scaled between $[-1, 1]$ and the squared sum of the loadings along any PC sum to 1 in most implementations of PCA. This allows a direct comparison of the contribution of each variable to each other. Figure 65.11 shows the conceptual data plotted in the original variable space to describe how PC loadings are calculated.

Using a geometrical interpretation, a PC loading is defined as the correlation (up to a scaling factor) of the PC axis with each of the individual variables used to characterize the data. If, for example, the PC direction is perfectly parallel to variable x_1 , then it is also perfectly correlated to this variable and its directional cosine is either -1 or 1 . This means (for a three-variable situation) that the PC equation can be represented as

$$PC_1 = 1 \times x_1 + 0 \times x_2 + 0 \times x_3$$

Or in other words, the entire system is described by the variable x_1 only and that the other variables do not contribute to this PC at all. If, for example, the weightings in subsequent PCs were all zero, the dimensionality of the problem has been reduced from 3 to 1. This is a special case and there are other ways of reducing the problem to a one-dimensional problem as shown in the following:

$$PC_1 = 0.333 \times x_1 + 0.333 \times x_2 + 0.333 \times x_3$$

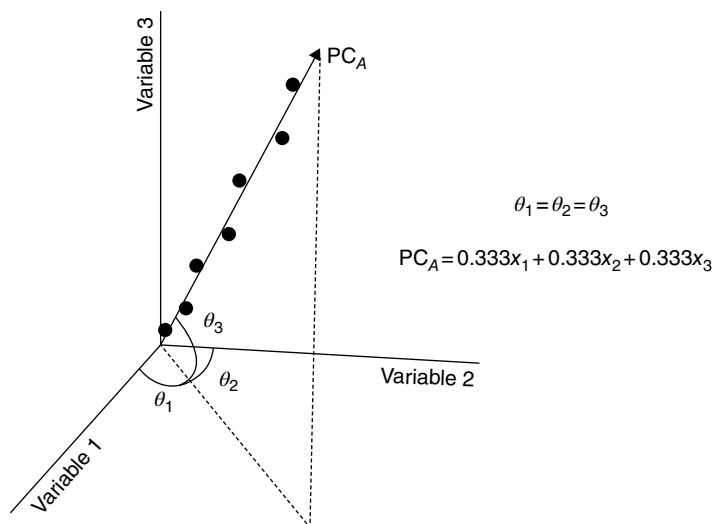


FIGURE 65.12 Equally contributing variables in the $K=3$ variable situation.

In this case, all variables contribute equally to describing the PC. This situation is shown graphically in Figure 65.12 where the PC intersects all three variable axes equally.

Loadings can be plotted in a number of ways and some of these are defined as follows:

1. Loadings line plots: Commonly used with spectroscopic or chromatographic data to show the importance of spectral bands or peak elutions for interpreting such data.
2. Loadings scatter plots: Commonly used in the interpretation of process variables or sensory data to understand discrete variable relationships.

Figure 65.13 provides some examples of the loadings plots used in describing PCA models.

In section, “Singular Value Decomposition” (SVD) algorithm is defined where it will be seen that although the scores vectors are also orthogonal, they are weighted by their importance in the model, that is, the variance explained by each component. This means that a direct comparison of scores and loadings on a single plot, known as the biplot, may not be entirely reliable for interpretation and it is suggested here that the interested analyst use caution when comparing scores and loadings in this manner. However, like scores, loadings with values close to zero (or close to the origin) contribute little to describing the samples in the PCs under investigation.

Since the loadings are scaled, the most important variables will have the highest absolute values and these are the ones that should be interpreted. Section 65.3.3.4 provides a brief introduction to variable scaling and its importance when developing a PCA model, particularly when taking into consideration the magnitude of the loadings.

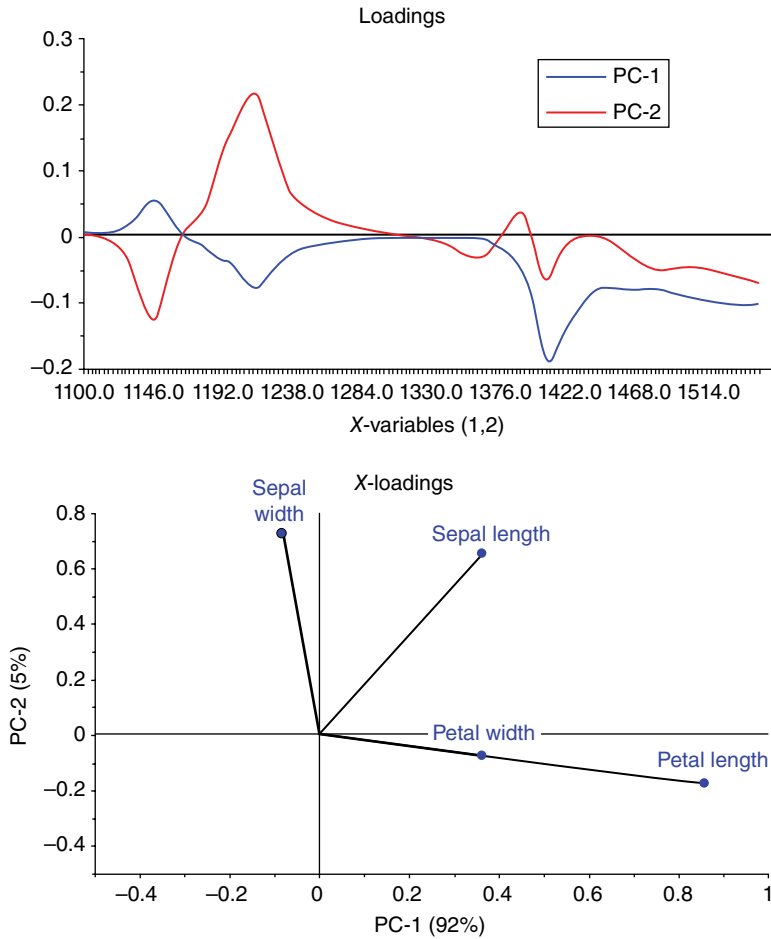


FIGURE 65.13 Examples of loadings plots used for different kinds of analysis situations.

In some cases, when the scaling is performed correctly, even though some variables do not contribute to the model, they may still be highly correlated to a particular PC. To view such correlations, the loadings are calculated to be scale invariant and this allows the generation of the correlation loadings plot. Figure 65.14 provides an example of a loadings plot and its corresponding correlation loadings plot showing how unimportant variables can be highly correlated to a PC.

The essential features of the correlation loadings plot are as follows:

1. The outer ellipse shows the points where variables are perfectly correlated to one or more PCs.
2. The inner ellipse shows the points where variables are, in sum, 50% explained by one or more PCs that are plotted. Since variance is equal to correlation, the explained variance can be calculated directly from the correlation loadings.

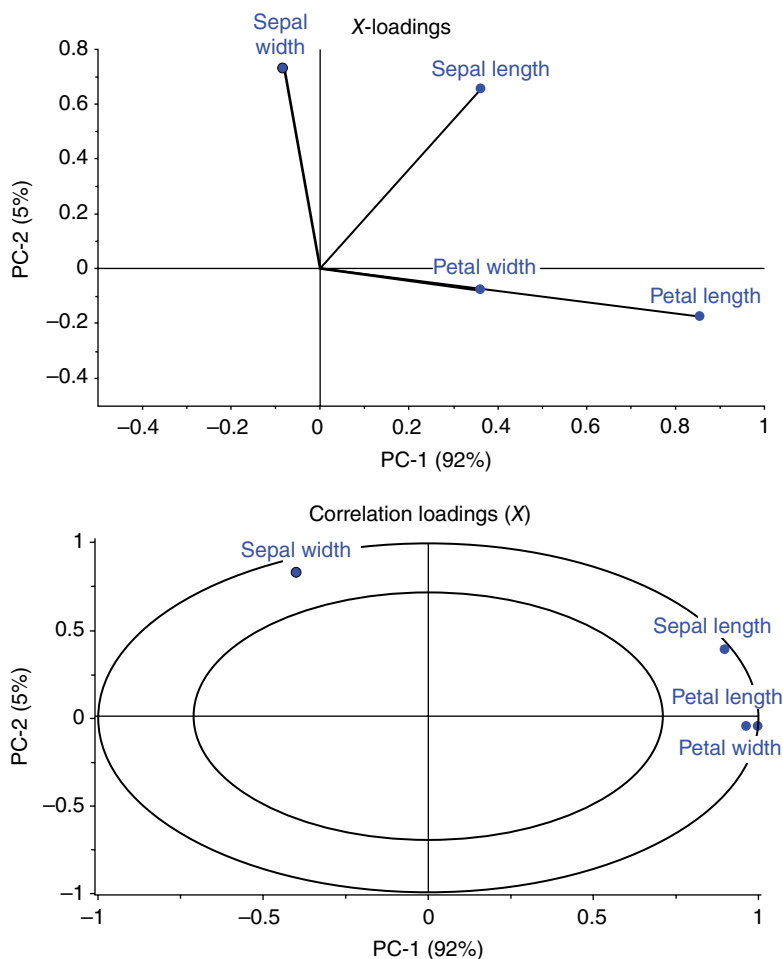


FIGURE 65.14 Example loadings plot and its corresponding correlation loadings plot.

Generally speaking, if a loading value is high and its correlation loading is also high, then the variable contributes highly to describing the PC. If the loading value is low for a variable and its correlation loading is high, then the variable is not contributing much to describing the PC, but its correlation structure should be noted for interpretation purposes. Finally, if a variable correlation loading lies within the 50% ellipse and the origin, for example, with a correlation loadings of 0.5 (25%), then most likely, this variable does not contribute at all to describing the PC. This rule of thumb, as a cutoff value in percentage, is highly dependent on the type of data. Spectroscopic data may hold important information in PCs that explain below 1% of the total variance, whereas for process data this will, in general, not be the case.

Overall, like scores, PC loadings provide a map of the variable relationships based on the characteristics of the samples measured. Loadings cannot be interpreted without

scores and vice versa. The process of interpreting scores and loadings will be provided by example later in this chapter.

65.3.3.4 A Short Introduction to Variable Scaling and Preprocessing The two common types of data analyzed using PCA are data generated by spectrometers/chromatograms or discrete process variables; however, many other data types exist. When dealing with these data types, variable scaling or preprocessing is an important first step prior to data analysis. Typically there are two separate approaches used to scale spectra and process data. This section is not meant to be an exhaustive description of data preprocessing, and there are many excellent subject-specific references in the literature [11, 12, 15–18]; it is meant only to be used as a guide for preprocessing data and interpreting it.

Preprocessing of Spectral Data The data typically generated by spectrometers (or chromatographs) is usually scaled to a set axis, that is, in absorption spectroscopy, the scale for each variable measured is between 0 and 5. It is usually not the intent in the analysis of such data to weight individual variable for two main reasons (although exceptions can and do occur):

1. A typical spectrometer/chromatograph can generate many points and individual variable scaling can be highly tedious and unnecessary.
2. Since the variables are all measured on a common scale, it is usually the intent to use PCA to find regions of the spectra that contribute to the variability in the data that are distinct from other regions.

In the analysis of spectral data, an analyst is trying to extract some chemical or biological information from the data. If the data contain unwanted physical effects, and these are the dominant source of variation in the data, then the PCA model will focus on these effects in the first one or first few PCs, rendering the chemical/biological information less important. Preprocessing is used to minimize such physical effects before analysis such that the desired information can be found in the first few PCs.

Depending on the type of spectroscopic or chromatographic method used, the most common preprocessing methods are presented in Table 65.3 with a brief listing of common methods and applications.

It is noted here that each preprocessing method has its own purpose for a particular effect. In some cases, correction of multiple effects can be performed by using combinations of preprocessing techniques; however, the overuse of preprocessing can distort the data leading to potentially false data interpretations after chemometric analysis.

Scaling of Discrete Data Sources Unlike spectroscopic or chromatographic data, process data usually comes from multiple sources with each source having its own variable scale range. As discussed in Section 65.3.3, the purpose of PCA is to find the

TABLE 65.3 Some Common Preprocessing Techniques Used for Spectroscopic and Chromatographic Data

Physical Effect	Preprocessing Methods	Common Applications
Baseline offset	Derivatives, baseline correction, detrending	Near- and mid-infrared, Raman, UV–visible
Scatter effects	Standard normal variate (SNV), multiplicative scatter correction (MSC)	Near-infrared, diffuse reflectance mid-infrared
Scale shift	Correlation optimization warping (COW)	Chromatography, nuclear magnetic resonance (NMR) spectroscopy

greatest sources of variability in a data set. Consider, for instance, a chemical reactor being measured by three discrete variable sources:

1. pH on a scale from 1 to 14
2. Temperature on a scale from 100 to 200°C
3. Pressure on a scale from 1000 to 2000 psi

This data was deliberately chosen with an order of magnitude step between the variables. Based on a straight univariate statistical analysis, the variability in the highest magnitude variable (i.e., pressure) may far exceed even the entire scale of the lowest magnitude variable (i.e., pH). It may be the case that small changes in pH provide the highest contributions to a quality parameter of the reaction, for example, final yield; however, a straight PCA on unscaled data would indicate that pressure has the greatest influence on data variability. To overcome this issue, the use of variable scaling is required.

Unlike univariate analysis, MVA techniques are typically not concerned with the distributional problems of the individual variables but more concerned with the overall variability and its relation to all variables being analyzed. The most common approach to variable scaling is autoscaling, where the variables are first mean centered and then they are divided by their standard deviations.

Figure 65.15 shows the difference between the original data, the mean centered data, and the autoscaled data.

Importantly to note, when developing a process control model using PCA, the values used to autoscale new data should be typical of the variability commonly experienced in normal process operations. This will ensure that abnormal situations are detected due to better understanding of the system.

The autoscaled data in Figure 65.15 now show that each variable can contribute to a PCA model on an equal scale in order to determine variable importance when assessing the loadings. Table 65.4 provides some examples of common scaling options used for the analysis of discrete source multivariate data.

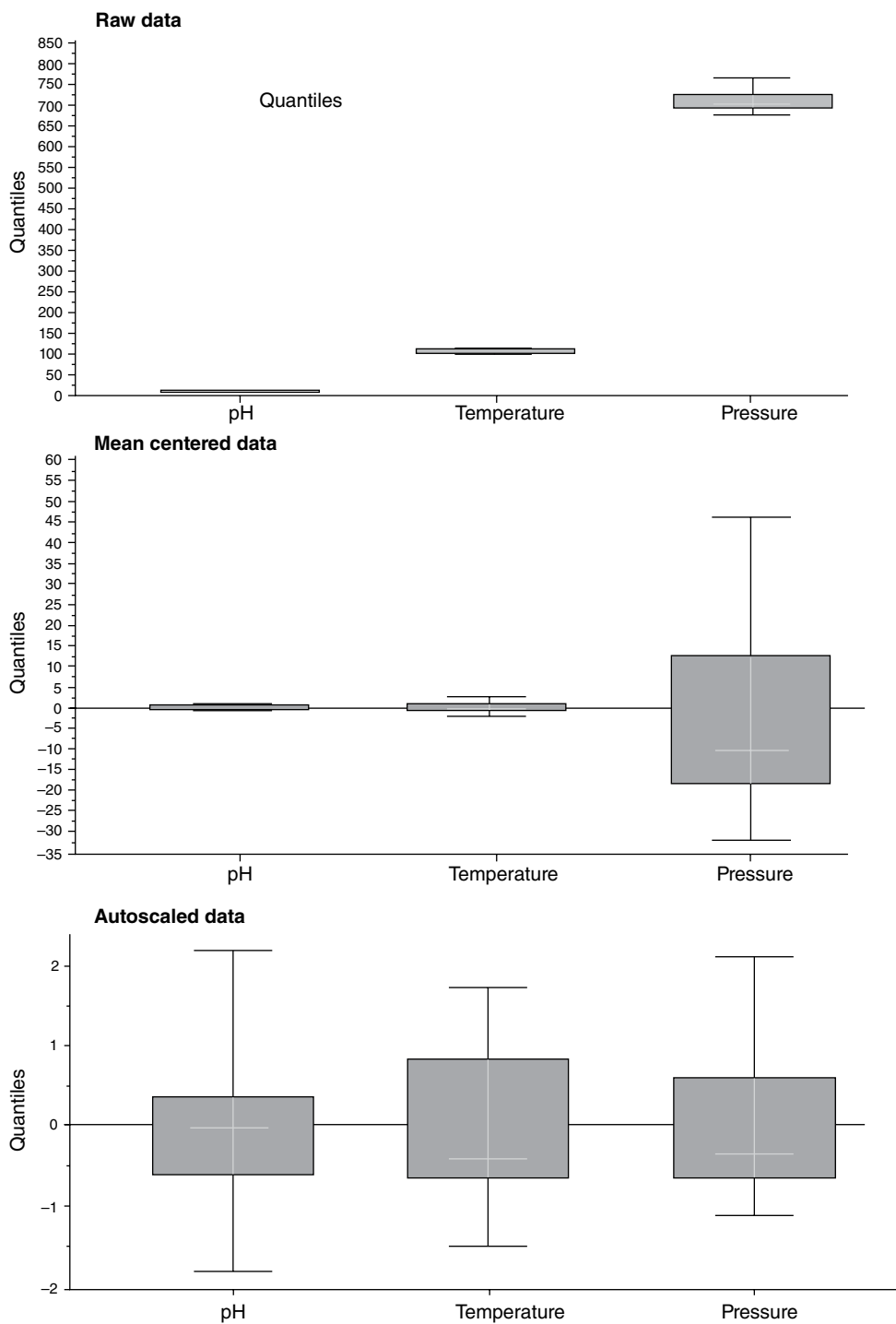
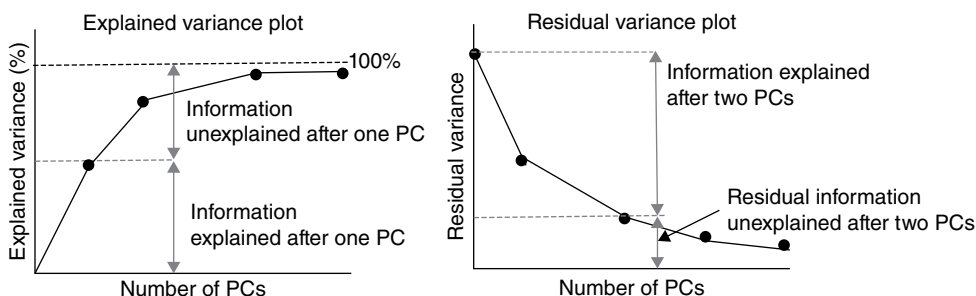


FIGURE 65.15 Examples discrete variable scaling.

TABLE 65.4 Some Common Preprocessing Techniques Used for Discrete Source Process Data

Preprocessing Methods	Common Applications
Autoscaling	Chemical reactors, pharmaceutical processes
Sphering	General multivariate data (of nonspectroscopic origin)
Quantile normalization	Metabolomic and other systems biology data

**FIGURE 65.16** Explained and residual variance plots for well-behaved models.

As was the case with spectral/chromatographic preprocessing, subject matter expertise is the best way of deciding how to scale a data set. It is important to keep in mind that scaling is used to enhance information extraction during chemometric analysis and for no other purpose. Only use a transformation where it is warranted; otherwise, the model can become overcomplex for the wrong reasons.

65.3.3.5 Explained/Residual Variance There are a number of steps involved in the validation and interpretation of a PCA model. Once a data set has been loaded into a suitable chemometric software package, preprocessed, or scaled, the appropriate validation method has been selected (refer to Section 65.6), and the analysis completed, the first step is to look at the complexity of the model.

The principle of parsimony is very important in chemometric modeling, that is, the simplest models are usually the most robust and easiest to interpret. Most chemometric programs provide a user with the explained or residual variance plots. These plot how much information is captured by each PC, either as:

1. **Explained variance:** This has a maximum value of 100% and plots how much each PC contributes to explaining the information in the data set.
2. **Residual variance:** This has a minimum value of zero and plots the residual variance in the original data after each PC has been added to the model.

Figure 65.16 provides examples of both plots for well-behaved data.

The faster these plots converge toward a plateau (i.e., toward 100% for explained variance or zero for the residual variance), the more systematic information is contained

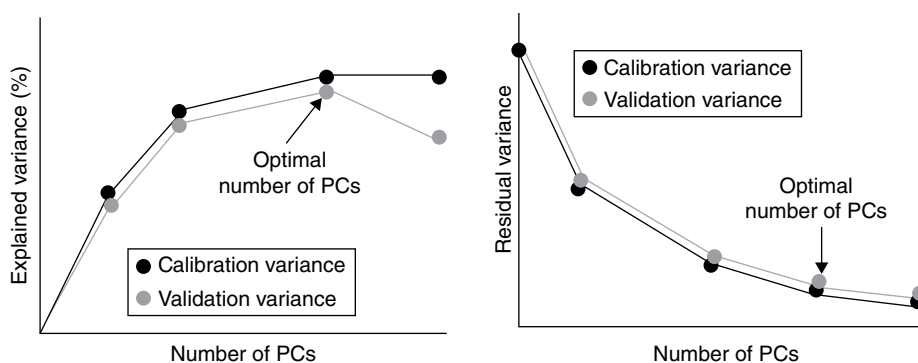


FIGURE 65.17 Explained and residual variance plots calibration and validation data.

in the model. Before model interpretation is attempted, an analyst should have some basic idea as to the sources of variation present in the data. For example, if the explained variance plot converges to 100% after only two PCs, it can safely be assumed that there are only two major sources of variation in the data to be interpreted.

Good chemometric software packages allow a user to view the calibration and validation variance of a model in one plot. Calibration variance is calculated by adding a component to a model and using the model to predict the calibration samples. Although this may seem invalid statistically, the calibration variance is used to determine the point where all information has been captured by the model and therefore provides a baseline for determining model complexity. Validation variance is used to assess the model performance using some form of validation (either cross validation or independent test set). Since the noise characteristics of the validation set may differ from the calibration set, comparison of the two curves allows an analyst to determine how many components to safely interpret. Figure 65.17 provides two examples of explained and residual variance curves for calibration and validation data.

In the case of the explained variance curve, the calibration curve continues to converge toward 100%, while the validation curve diverges at the third PC. In the case of the residual variance plot, the calibration and validation curves follow each other closely and converge together with no signs of diverging. In this case, the point where the two curves plateau is the best place to start interpreting the model.

These variance plots are also useful diagnostic tools for detecting outliers in the data set. This is discussed in more detail in Section 65.3.3.5. Once the number of PCs to interpret in a model has been established, a detailed assessment of scores and loadings can be performed.

65.3.3.6 Residuals and Diagnostic Tools in PCA One of the most powerful aspects of PCA is its comprehensive graphical and diagnostic toolkits for validating and interpreting a model. This section provides a brief description of some of the methods used to detect outliers in a model and also effectively interpret it for use in practical applications.

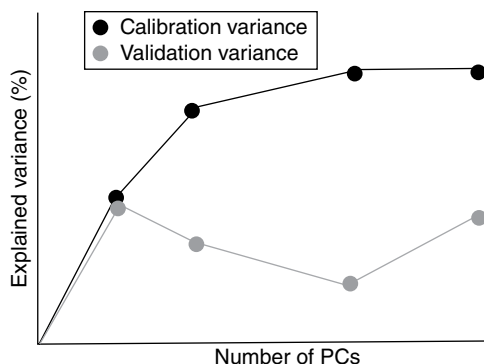


FIGURE 65.18 Outlier detection in the explained variance plot.

Outliers in PCA The explained/residual variance plots introduced in Section 65.3.3.5 are useful in determining model complexity, but they can also be used as a first check for the presence of outliers. By definition, an outlier can be classified into one of the following situations:

1. Measurement error
2. Wrong labeling
3. Deviating sample
4. Noise
5. Extreme/interesting sample

From the aforementioned definition, an outlier does not necessarily have to be a bad measurement. For example, if a group of 10 samples were measured and 1 had different natural characteristics than the other 9, this is not a measurement error or noise, but an interesting sample (or possibly the discovery of a new sample class).

A gross outlier can usually be detected in the explained/residual variance plot by an abnormal structure in the validation curve. Figure 65.18 provides an example of such a curve and the erratic behavior of the validation curve profile.

When such behavior is observed, an analyst must then look at the scores or influence plots (section “Influence Plots”) to look for the samples causing the deviation.

Hotelling’s T^2 The Hotelling’s T^2 statistic [19] is a multivariate generalization of the Student t -test. It is similar to the Mahalanobis distance [20] reported by texts on chemometrics but has the advantage that statistical limits can be placed on the distances. The form of the Hotelling’s T^2 statistic is as follows:

$$T^2 = (X - \bar{X})W^{-1}(X - \bar{X})$$

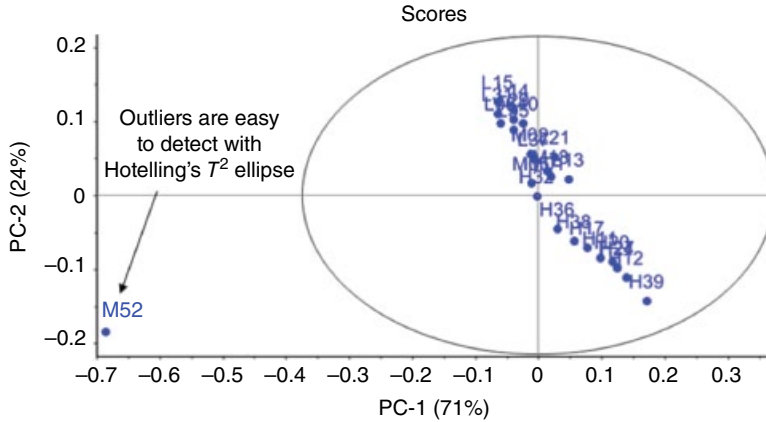


FIGURE 65.19 Using Hotelling's T^2 ellipse in scores space to detect outliers.

where

X is the original data matrix

\bar{X} is the mean of the data set

W is the covariance matrix of X

The Hotelling's T^2 statistic is approximately F -distributed as follows:

$$F_{p,n,\alpha} \sim T^2 \frac{(n-a)}{a(n-1)}$$

Any sample that has a calculated F -value that exceeds the critical F -value can be considered for investigation as an outlier. In chemometrics the scores rather than the original variables are used in the calculation of Hotelling's T^2 . A convenient way to display the Hotelling's T^2 statistic is by displaying it on a scores plot, where it can be shown at a number of levels of significance. Figure 65.19 shows how the T^2 ellipse can be used to detect an outlier in scores space.

Hotelling's T^2 is one of the most important statistics in MSPC. This topic is discussed in more detail in Section 65.7 where uses of the T^2 chart will be provided.

Leverage A closely related diagnostic tool to Hotelling's T^2 for detecting outliers is leverage (h_i) and is defined as the potential of a sample to be influential [15]. Leverage is calculated as follows:

$$h_i = \frac{1}{N} + x_i'(X'X)^{-1}x_i$$

where

N is the number of samples in the model

x_i is the centered x -vector for sample i

This equation defines an ellipse and samples with equal leverage lie at equal distances around the center of the model. There are no strict rules for setting up when a sample is a leverage outlier and a general rule of thumb is to investigate a sample whose leverage is 2–3 times larger than $(1 + A)/N$, with A being the number of PCs used in the model. There is a one-to-one relationship between Hotelling's T^2 and leverage where leverage typically is used as outlier criteria when the number of samples is limited.

X-Residuals An X -residual is defined as that part of the data that has not been modeled. Expanding on the principle that data is equal to information plus noise, the X -residual is a measure of the noise after a PCs have been taken into account. Expanding on the form of the PCA model, the following equation can be derived:

$$X = TP' + E = \sum_{a=1}^a t_a p'_a + E$$

This equation can be rearranged as follows:

$$E = X - \sum_{a=1}^a t_a p'_a$$

where the residual E is what remains after a PCs have been subtracted from the original data set. This can be simplified down to individual sample residuals as follows:

$$e'_i = x'_i - t_a p'_a$$

A desirable property of residuals is that they should be randomly distributed without any systematic structure. If a residual has any form of interpretable structure, there are two main causes:

1. Not enough PCs have been added to the model to account for the systematic variation still remaining in the data.
2. The sample is an outlier and its structure is not well described by the model.

X -residuals in PCA find most use for the analysis of spectral outliers; this is because regions where the spectra are not being adequately modeled can be visualized and interpreted. Figure 65.20 provides an example of a single outlier and the diagnostics available to detect the outlier. Note the structure remaining in the X -residuals.

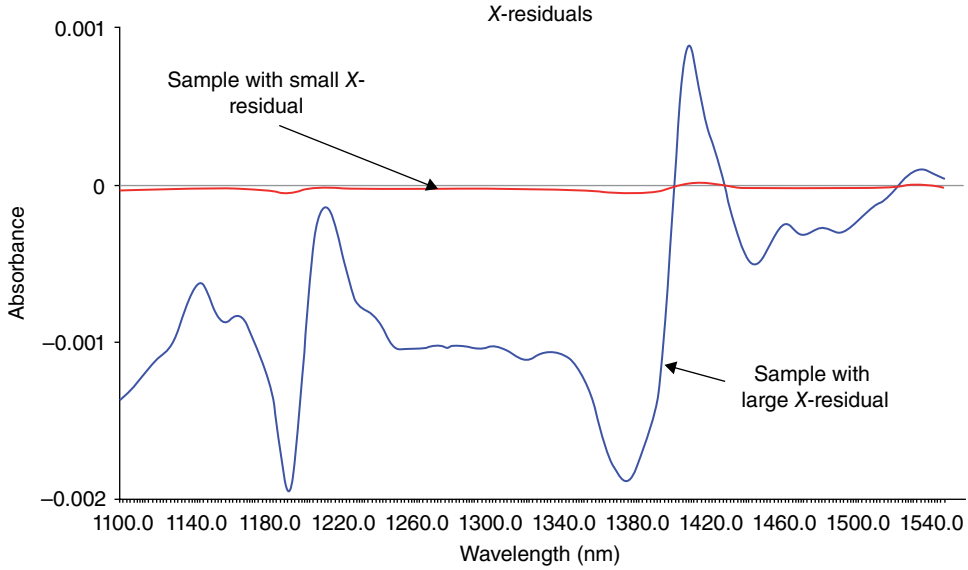


FIGURE 65.20 Using X -residuals to detect spectral outliers.

X -residuals alone can be subjective when using them as a visualization tool. Taking the sum of squared residual values can reduce the X -residual to a single point for each sample. To be able to assess these residuals objectively, the use of Q -residuals and F -residuals is discussed in the next sections.

Q -Residuals Q -residuals were first introduced by Jackson and Mudholkar [21] to detect the so-called type B outliers. These outliers can be the result of:

1. Too few PCs used to adequately describe the original data X .
2. The samples are truly outliers from the model.

The Q -residual is calculated from the regular X -residual as a squared sum:

$$Q_i = e_i' e_i = (x_i - t_a p_a')' (x_i - t_a p_a')$$

The critical value for Q can be obtained from the following formula:

$$Q_\alpha = \theta_1 \left[\frac{c_\alpha \sqrt{2\theta_2 h_0^2}}{\theta_1} + \frac{\theta_2 h_0 (h_0 - 1)}{\theta_1^2} + 1 \right]^{\frac{1}{h_0}}$$

where Q_α is the critical value for the Q -distribution and the remainder of the terms can be found in the text by Jackson [22]. The use of the Q -residual in MSPC applications is very important for detecting the onset of failure before it becomes a critical issue. Refer to Section 65.7 on MSPC for more details.

F-Residuals F -residuals are calculated from Q -residuals and are considered more conservative by many with respect to Q -residuals. The F -residuals are calculated as follows:

$$F_i = \frac{Q_i}{(K - A)}$$

where

F_i is the F -residual

Q_i is the Q -residual

K is the number of variables

A is the number of PCs used in the model

The F -residual is compared for significance against the standard F -test hypothesis:

$$F_{\text{crit}} \sim (\alpha, 1, i)$$

where i is the number of samples in the model. The major advantages of the F -residual over Q -residuals include the following: the F -test is a more established test compared to the Q -test, the calculation of F -residuals is computationally faster, and, most importantly, F -residual analysis can be applied to both calibration and validation residuals, where Q -residuals are only applicable to calibration residuals.

Influence Plots A convenient way to visualize the presence or absence of outliers is to plot leverage on the x -axis and X -residual on the y -axis. This is known as an influence plot. The idea behind the influence plot is shown in Figure 65.21.

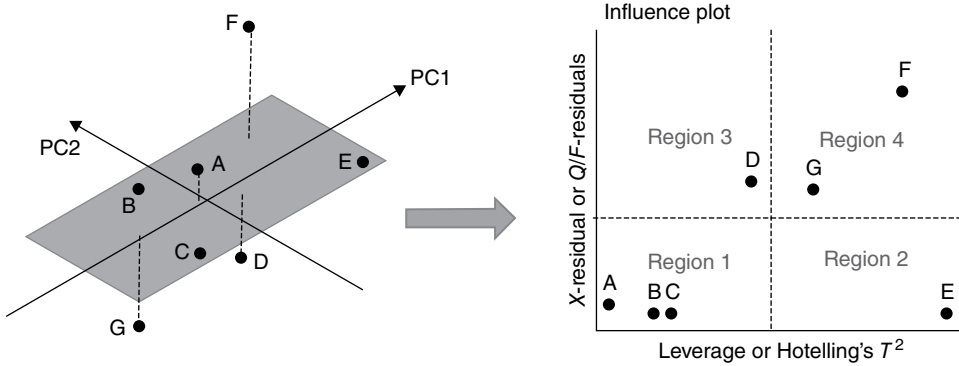
Consider the plane formed by two PCs in Figure 65.21. Samples that lie close to the plane but are extreme along the PC axes are leverage samples, that is, they can influence the orientation of the plane if they become too extreme with respect to the center of the model.

Samples that do not lie on the plane in an orientation perpendicular to the plane show some form of X -residual. As samples lie further away from the plane, they do not fit the model well and are therefore X -residual outliers. When a sample is extreme in both X -residual and leverage, it can in most cases be concluded that the sample is a true outlier.

There are a number of variants of the influence plot; the most common forms are:

1. X -residual versus leverage
2. Q -/ F -residual versus Hotelling's T^2

Any combination of the aforementioned variants is possible. The advantage of using Q -/ F -residuals and Hotelling's T^2 in influence plots is such plots allow the placement



Region 1: Samples similar to the majority of the calibration population.
 Region 2: Samples fit model but are extreme in properties.
 Region 3: Samples differ from the average model population.
 Region 4: Samples are different and extreme (most likely outliers).

FIGURE 65.21 The idea behind the influence plot.

of statistical limits on the plots, thus providing objective evidence for the presence or absence of outliers. Figure 65.21 shows an influence plot with statistical limits. Such a plot forms the basis of MSPC control charting introduced in Section 65.7.

65.3.3.7 Application of PCA to Fisher's Iris Data To demonstrate the graphical power and diagnostic capability of the PCA method, the iris data set introduced earlier will be used. Preliminary investigation of the data shows that only four variables were measured, so a maximum of three PCs should be interpreted. If less PCs are needed, the more similar is the information contained in the original variables. Figure 65.22 shows the PCA overview for the iris data analysis.

Explained Variance The plot of explained variance shows that PC1 describes approximately 92% of the total data variability and PC2 a further 5% for a total of 97% explained in two PCs. PC3 only contributes slightly and can be excluded from the analysis. The calibration and validation curves follow each other closely; therefore, a 2-PC model can be validated and possibly interpreted.

Scores Plot It is now justified to interpret a 2-PC model only. The scores plot in Figure 65.22 has been grouped based on iris type. It can be interpreted from this plot that:

1. Setosa can be uniquely distinguished from versicolor and virginica.
2. Versicolor and virginica are slightly distinct from each other; however, the potential for overlap can be seen, therefore leading to ambiguous interpretation of the data.

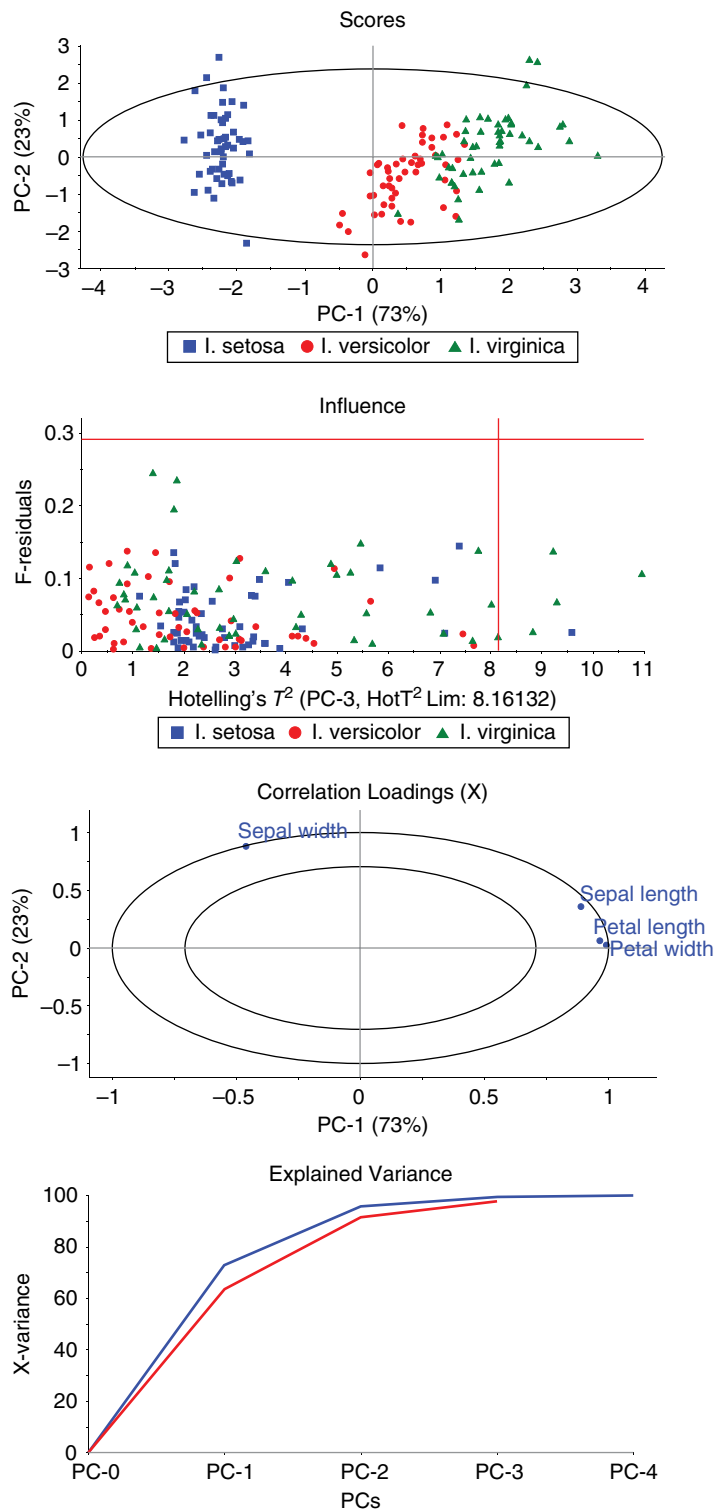


FIGURE 65.22 PCA overview of Fisher's iris data.

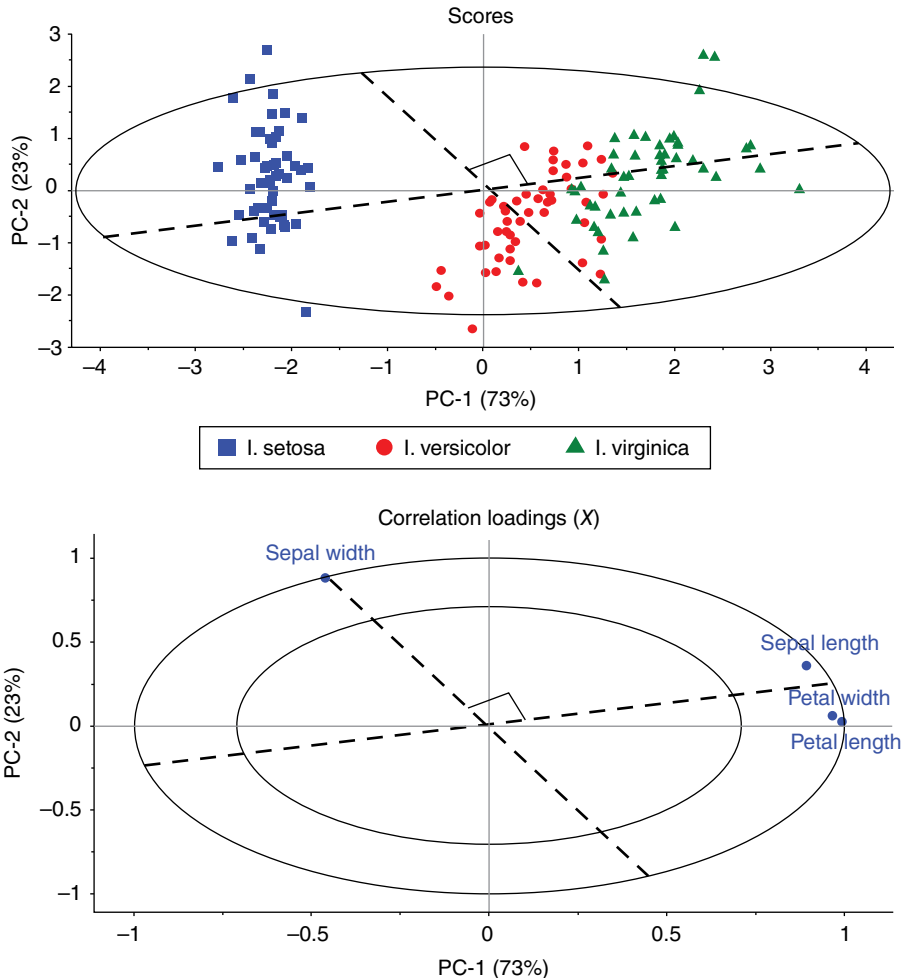


FIGURE 65.23 Geometrical interpretation of PCA scores and loadings for Fisher's iris data.

To understand why the samples group and spread the way they do, an interpretation of the loadings plot is required.

Loadings Plot For the purposes of clarity, the correlation loadings plot is shown in Figure 65.22. Along PC1, petal length and petal width almost lie exactly on the positive PC1 axis. This indicates that these two variables contribute most to PC1 and since they lie on top of each other, it can be assumed that they are highly correlated.

Sepal length also lies in the positive PC1 direction, but since it also occupies some of the PC2 space, it contributes to the spread of the samples along the PC2 direction in the scores plot. The largest contributor to the spread in the PC2 direction is sepal width. Figure 65.23 shows the scores and loadings plot for the iris data, this time showing the geometrical interpretation of the data.

Starting with the loadings, the direction that captures the joint petal length and width and the sepal length describes why *setosa* is different from *versicolor* and *virginica*; in particular, *versicolor* and *virginica* are distinguished from *setosa* based on larger values of petal length and width and sepal length.

Two new lines were plotted in the scores and loadings plot, and it can be seen that they are nearly orthogonal to each other. This indicates that the effect of sepal width is independent of the effect of the other variables. As PC2 only describes approximately 5% of the variability, sepal width is the minor contributor to distinguishing between the three classes of iris. Removal of this variable actually has no effect on the discrimination power of the PCA model. The recalculated PCA without sepal width is shown in Figure 65.24.

This ability to remove variables and recalculate a model based on observation and the interpretation of diagnostics is what makes PCA one of the most powerful data analysis methods available.

Outlier Analysis Figure 65.22 shows the F -residuals versus Hotelling's T^2 influence plot for the iris data. The plot shows that there are no observable samples grouping plot and no X -residual outliers, and therefore the data set is representative and indicates that all samples fit the model well. There are, however, six potential leverage outliers, primarily from *virginica* and one from *setosa* but one from *versicolor*. Figure 65.25 shows the scores plot with the 95% Hotelling's T^2 ellipse drawn and the influence plot with leverage outliers marked. The leverage samples are those that exceed the Hotelling's T^2 boundaries in PC1 and PC2. These samples are extreme within the data set, but they are close to the boundaries. These samples were not considered to highly influence the model.

PCA Summary for Fisher's Iris Data Overall, the objective of the analysis was achieved; the original four-dimensional data set (i.e., four variables) was reduced to a 2D problem. The information in petal length and width was found to be highly correlated, so there is no advantage to having both measures in the analysis.

It was also found that sepal width did not contribute to the separation of the classes and after elimination, there was no change in the data structure. The ability to justifiably remove unimportant variables is a powerful tool of PCA. PC1 was found to be the main contributor to distinguishing between *setosa*, *versicolor*, and *virginica*. The variables most responsible for this distinction were petal length and width and sepal length. Sepal length was responsible for the spread of the sample data in PC2.

An outlier analysis detected no X -residual outliers. There were some leverage outliers detected; however, when assessed with the scores plot with Hotelling's T^2 ellipse at 95% confidence, these extreme samples were found not to be too extreme and did not influence the model.

65.3.3.8 Algorithms for PCA There are two main algorithms commonly used in software packages for calculating PCs. These are the SVD algorithm first introduced

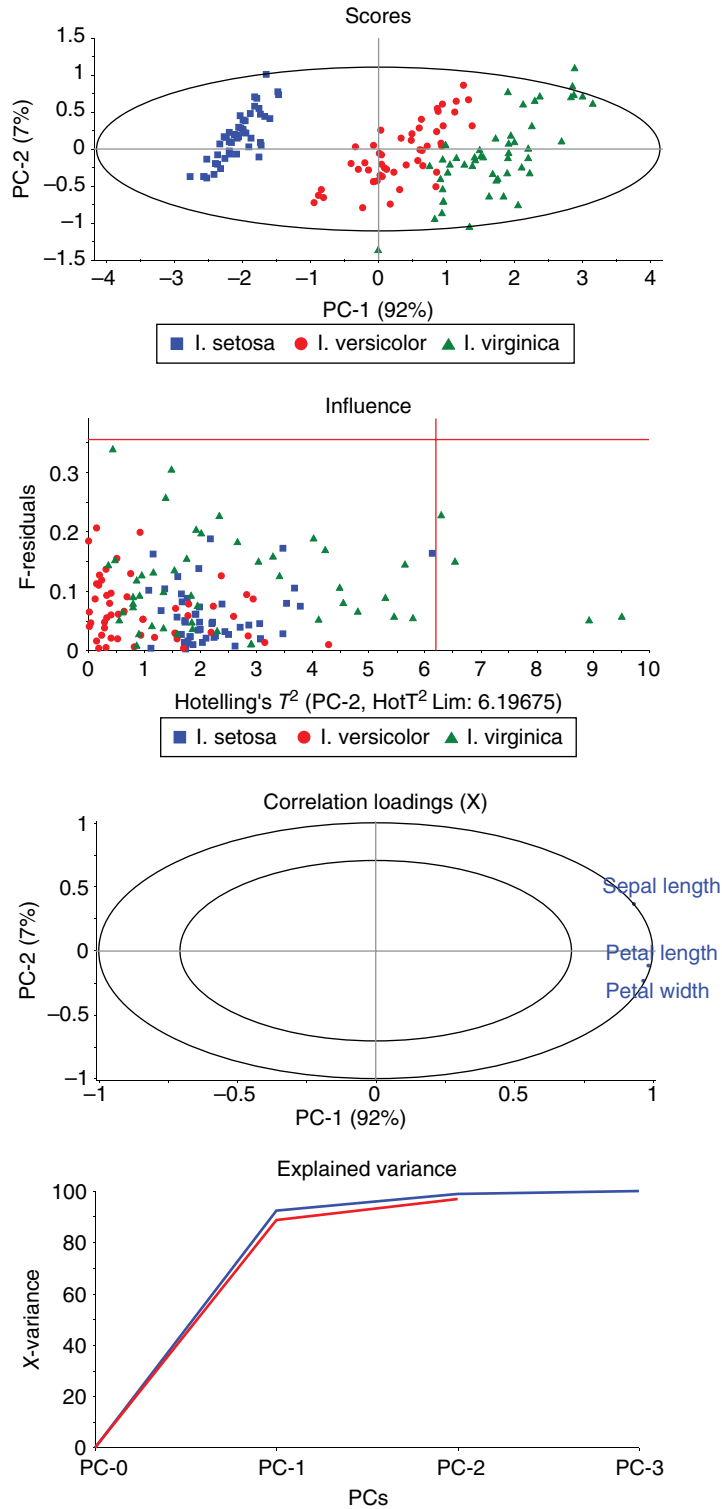


FIGURE 65.24 PCA overview of Fisher's iris data after the removal of an unimportant variable.

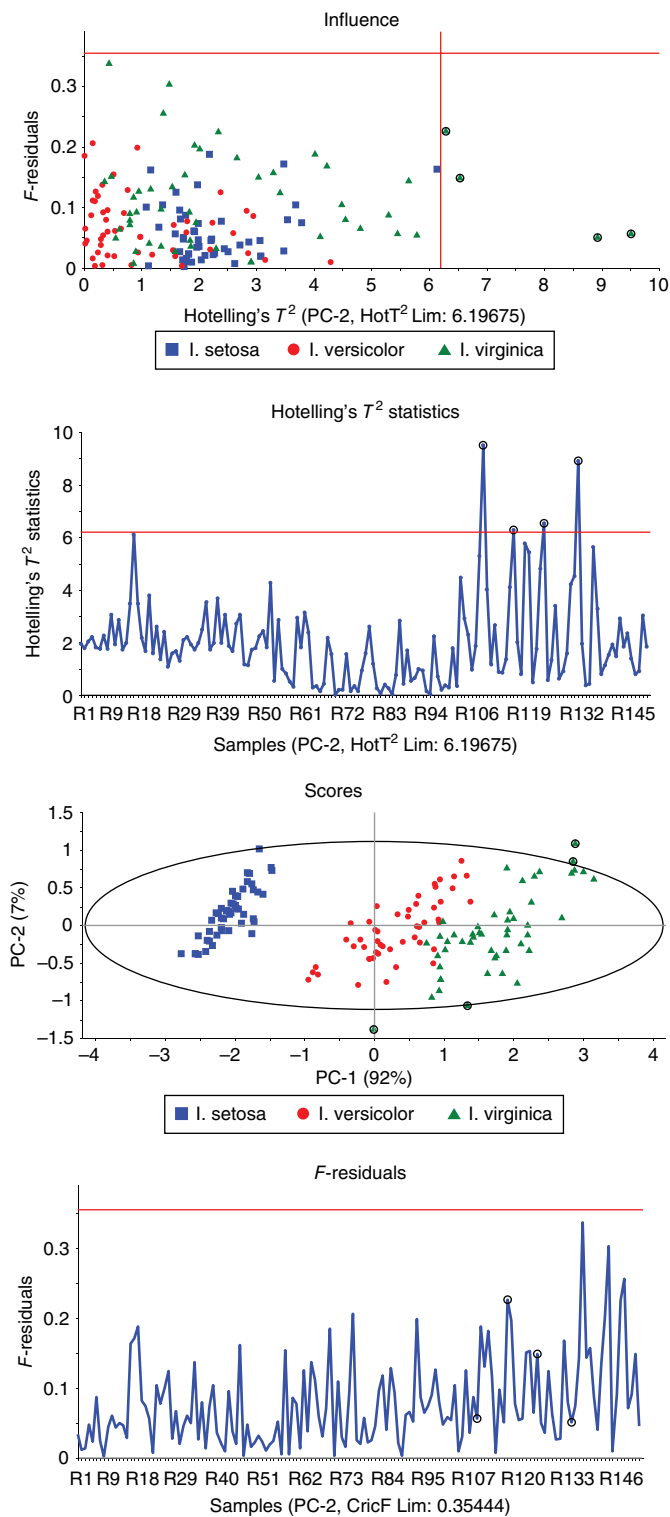


FIGURE 65.25 Outlier analysis of Fisher's iris data.

by [22] and the noniterative partial alternating least squares (NIPALS) algorithm first introduced by Wold [23].

Singular Value Decomposition In SVD, a matrix X is decomposed into a product of the so-called characteristic vectors of $X'X$, the characteristic values of XX' , and a function of their characteristic roots [22]. Now, if all of the PCs are used to describe the data, the original matrix X can be regenerated as follows:

$$X = TP'$$

SVD calculates all of the characteristic vectors of a data set in one operation. The following describes in general how the algorithm works.

The general form of the SVD model is defined as

$$X = U\Lambda P'$$

where

X is the original data to be analyzed

U is an $(N \times K)$ orthogonal matrix containing the so-called left singular vectors (N is the row dimension of U)

Λ is a symmetrical matrix of diagonal elements $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p$ (p being the column dimension of both U and P) (the diagonal elements of Λ are called the singular values and describe the importance of each PC being calculated; singular values are the square roots of the eigenvalues calculated from the covariance matrices of the original data X)

P is a $(p \times p, p \leq K)$ orthogonal matrix containing the so-called right singular vectors

The matrix U is calculated by eigenanalysis from the matrix XX' whose dimension is $(N \times N)$, that is, U describes sample characteristics and correlations. Correspondingly, the matrix P has dimension $(p \times p)$ and describes variable characteristics and correlations.

In order to relate U to P , the singular value matrix Λ describes the importance of each row of U based on the variable contributions in P . It is common practice in PCA to combine the matrix product $U\Lambda$ and define it as the vector T , that is, the scores vector. The common PCA model definition can now be stated as

$$X = TP' + E$$

where the matrix E is the residual that cannot be explained by the model. By combining $T = U\Lambda$ it can now be seen how the explained variance of each PC can be calculated. If X contains 100% unexplained variability, then each singular value λ_a describes a

certain proportion of the original variability, with the proviso that the scores vector t_1 contains the most information followed in order by the remaining scores vectors. The trace of Λ provides an estimate of the total variability described by the PCA model, and its dimension can be used to determine the rank ($p \leq K$) of the original data (i.e., the number of important features in the data).

For a more detailed description of the algorithm details of SVD, the interested reader is referred to the text by Jackson [22] and Hastie [24].

Noniterative Partial Alternating Least Squares The algorithm extracts one component at a time, with each component obtained iteratively by repeated regressions of X onto the scores \hat{t} to obtain improved loading vectors \hat{p} and then repeated regressions of \hat{p} on X to obtain improved \hat{t} [15, 18].

The following algorithm assumes that all preprocessing and centering of the data have been performed prior to analysis:

Step 1: Choose an initial scores vector \hat{t}_a in X as the column with the highest remaining sum of squares.

Step 2: Improve the estimate of the calculated loading vector \hat{p}_a by projecting the matrix X_{a-1} onto \hat{t}_a , that is,

$$\hat{p}'_a = (\hat{t}'_a \hat{t}_a)^{-1} \hat{t}'_a X_{a-1}$$

Step 3: Scale the length of \hat{p}_a to 1.0 to avoid scaling ambiguity,

$$\hat{p}_a = \hat{p}_a (\hat{p}'_a \hat{p}_a)^{-0.5}$$

Step 4: Improve the estimate of score \hat{t}_a for the component by projecting the matrix X_{a-1} onto \hat{p}_a ,

$$\hat{t}_a = X_{a-1} \hat{p}_a (\hat{p}'_a \hat{p}_a)^{-1}$$

Step 5: Improve the estimate of the eigenvector \hat{t}_a ,

$$\hat{t}_a = \hat{t}'_a \hat{t}_a$$

Step 6: Check convergence by subtracting the \hat{t}_a from the previous iteration. If the difference is smaller than some predetermined tolerance, the method has converged; otherwise, perform another iteration (in most software applications, a user will be allowed to specify a maximum number of iterations to converge).

Step 7: Once convergence has been reached for a particular component, subtract this from the starting X -data,

$$X_a = X_{a-1} - \hat{t}_a \hat{p}'_a$$

Step 8: Go to step 1 and repeat to step 7 for the deflated matrix to calculate the next (and successive) components.

It can be seen from the aforementioned algorithm description why it is referred to as an iterative algorithm as it aims to improve the scores and loadings precision through convergence criteria. This is the algorithm used in most software packages as it is relatively fast, but it has the distinct advantage over SVD that it can handle missing values in the original X -matrix. On the other hand, SVD is based on algorithms that give exact results from back-substitution or equivalent procedures and are regarded as numerically superior.

65.3.3.9 Summary of PCA PCA is known as the “workhorse” of MVA methods [18]. As an EDA method, PCA is unrivaled with its ability to provide a highly graphical environment along with many diagnostic tools that can help interpret and improve a model.

PCA separates a data set into sample information (scores) and variable information (loadings). Used in combination, scores and loadings can provide extensive insights into complex data sets. Not only does PCA allow understanding of sample and variable relationships, but it also provides a means to deciding the complexity of the model using explained or residual variance plots. PCA models are validatable using either external data sets or internal segregation of samples for training and validation purposes (refer to Section 65.6 for more details on validation).

The incorporation of statistics into MVA methods allows the definition of multivariate confidence intervals. This provides PCA with the capability of objective outlier detection. Diagnostics such as Hotelling’s T^2 ellipses and intervals and Q - and F -residuals are available at a number of statistical confidence levels. This is highly important for detecting and justifiably removing outliers from a data set. The properties have given rise to MSPC applications further discussed in Section 65.7, where PCA models can supplement traditional SPC models for enhanced understanding of many processes in many industries.

PCA allows visualization of clusters in data sets containing many classes of samples. If separation of the classes can be established visually, individual PCA models can be developed and used in multivariate classification (Section 65.5). Even if complete separation of all classes is not possible, the joining of classes into a single PCA model is possible.

The next section of this chapter focuses on multivariate regression. It will be seen in this section that PCA also forms the basis of many of the methods discussed. Overall, PCA is a recommended first approach to the analysis of any multivariate data set. After the application of an appropriate preprocessing method (Section 65.3.3.4), PCA will provide greater insights into many data sets and avoids the need for the common “one-variable-at-a-time” approach that may lead to false or no conclusions at all.

65.4 MULTIVARIATE REGRESSION

This section provides an overview of some of the most common multivariate regression method currently used in research and industrial practice.

65.4.1 General Principles of Univariate and Multivariate Regression

Regression is a mathematical approach for relating two or more sets of variables to each other [15, 25]. In regression modeling, a representative set of X -variables (where X is multivariate) is used for modeling one or several Y -variables, also known as response variables. Regression methods have played an important role as a tool in analyzing a multitude of samples of various kinds from the pharmaceutical, food and beverage, agricultural, and many other industries.

The ideal situation in a regression context is that the change in the X -variable(s) is 100% related to the change in the Y -variable, thus requiring high sensitivity of at least one X -variable in the set to the response. In many cases it is the simultaneous contribution from several X -variables that enables a multivariate modeling of the property under investigation.

As was the case in PCA (Section 65.3.3), the general regression model has the following form (repeated here for clarity):

$$\text{DATA} = \text{INFORMATION} + \text{NOISE}$$

The main difference between PCA and regression is simple; PCA models the internal structure of the X -variables in one table, where regression models the relationship between X - and Y -variables in two (or more) tables. Mathematically, the regression model is represented by

$$Y = XB + E$$

where

Y is the variable to be predicted (response or dependent variables)

X is the variables measured to predict the response (predictors or independent variables)

B is a matrix of regression coefficients (i.e., the model that relates the X -variables to the Y -variables)

E is a matrix of residuals representing the lack of fit of the model

To introduce some key concepts related to the construction of regression models, the simplest case occurs when a single response has to be predicted based on the measurement of a single independent variable. This case occurs frequently in the

physical and chemical sciences and engineering. In particular, a common application in chemistry is where the concentration of an analyte must be quantified based on the measurement of a univariate physical property or of an individual instrumental signal (e.g., absorbance at λ_{\max} in Beer's law [11] or diffusion limit current in Ilkovič Equation [26]). In mathematical terms, the univariate regression model can be stated as

$$y_i = f(x_i) + e_i = \hat{y}_i + e_i$$

where, for each object, the predicted response value $\hat{y}_i = f(x_i)$ represents an approximation of the true value y_i , the difference between the two being the residual e_i . Although the general functional relation $f(x)$ is used in this equation, it is assumed that the most common mathematical form of the regression models is a linear model. Therefore, under the assumption of a linear relationship between the dependent and the independent variables, the model becomes

$$y_i = \hat{y}_i + e_i = b_1 x_1 + b_0 + e_i$$

where b_1 is the slope and b_0 is the intercept of the model (or simply put, the straight line model of common form $y = mx + b$). The generation of a regression model requires finding the optimal values of the model parameters b_1 and b_0 . In this context, the criterion that is normally used is the so-called least squares criterion [27, 28], which is based on a sum of squared error loss function. In particular, the concept behind approximating the functional relation between X and Y via a least squares model is to look for the value of the parameters that allows fitting the data with the minimum possible error [29]:

$$\min_{b_0, b_1} \sum_{i=1}^r e_i^2 = \min_{b_0, b_1} \sum_{i=1}^r (y_i - \hat{y}_i)^2 = \min_{b_0, b_1} \sum_{i=1}^r (y_i - b_1 x_1 - b_0)^2$$

r being the number of X -/ Y -pairs used to build the model (training set) and the rest of the symbols as previously defined. The optimal value of the parameters is the one that minimizes the sum of squares of the model residuals, known as least squares fitting. The interested reader is referred to the literature for the derivation of the normal equations used to solve the least squares problem [25].

The simple univariate model described previously is easily generalized to the multivariate situation. The general form of the multivariate regression model is as follows:

$$y_i = b_1 x_1 + b_2 x_2 + \cdots + b_p x_p + b_0 + e_i = \hat{y}_i + e_i$$

This is the general form of the multiple linear regression (MLR) model discussed in Section 65.4.2.

65.4.2 Multiple Linear Regression

With reference to the simple univariate least squares approach discussed in Section 65.4.1, MLR is relatively simple and straightforward. However, the mathematical structure of the model can severely limit the possibility of its practical application to many real cases, where a large number of variables are measured on a relatively small number of samples. The general model for finding the regression coefficients in the MLR model is shown as follows:

$$\mathbf{b}_{\text{MLR}} = X^+ y = (X'X)^{-1} X'y$$

where the symbol $+$ indicates Moore–Penrose pseudoinverse [30, 31].

Estimation of the optimal value of the regression coefficients relies on the inversion of the matrix $(X'X)$, and for many experimental data this inverse doesn't exist or it is ill-conditioned (i.e., it is unstable when the model is applied in practice).

In particular, the conditions that have to be satisfied in order for $(X'X)^{-1}$ to be estimated in a reliable way are that the columns of X are linearly independent (meaning that the predictors are uncorrelated) and that the number of training samples r is greater than the number of independent variables p . From a practical standpoint, the latter condition could be, at least in principle, met by increasing the samples to variables ratio either by measuring more samples or by variable selection. However, the former is rarely satisfied, especially when signals coming from modern instrumentation are involved, as the variables are quite often correlated by nature or by sampling. The direct consequence of the matrix X being ill-conditioned is that the coefficients are not stable and are characterized by high variance, since the solution is mostly affected by the noise part of the data [15]. Dependent on the underlying structure of X , this may also give higher prediction error. This must be investigated by proper validation. It should be mentioned that this instability may also lead to false interpretation of the coefficients as well as the interpretation of results from ANOVA (p -values) and that these pitfalls occur long before there are numerical problems in inverting $(X'X)$. For this reason many implementations of MLR check the so-called condition number to give warnings about the rank of X . However, to take out some variables because they happen to be correlated to others due to the fact that one is not applying a suitable method to handle this situation is scientifically unsatisfactory.

To deal with these drawbacks, different methods have been proposed in the literature, most of them based on the concept of bilinear modeling, already introduced in Section 65.3.3. Indeed, when the description of the data set using the experimentally measured variables is substituted by a more parsimonious one, relying on the concept of latent variables, then it is often possible to capture the essential structure of the data with a very limited number of descriptors. It is then evident that these two characteristics (low number of mutually orthogonal predictors) allow overcoming all the limitations described previously and make multivariate calibration applicable to a wider host

of real-world problems. In this framework, the most commonly used latent variable-based methods are principal component regression (PCR) and partial least squares regression (PLSR), which will be described in Sections 65.4.3 and 65.4.4, respectively.

65.4.3 Principal Component Regression

As the name suggests, PCR [15, 32] is based on the use of PCA [22] (see Section 65.3.3) to produce a parsimonious description of the independent matrix X .

Indeed, as projection of the samples onto the first PCs constitutes the best low-dimensional approximation of the original data matrix, the natural extension is the use of PCA scores as the independent variables in the MLR problem, which may overcome the limitations of the method when dealing with ill-conditioned experimental matrices. Therefore, PCR modeling is a two-step process firstly involving PCA decomposition of the X -variables and successively the generation of an MLR model on the scores [32].

In matrix notation, the independent matrix X is described by the bilinear model already discussed in Section 65.3.3 and repeated for clarity here:

$$X = TP' + E$$

where T and P are as previously defined as the scores and loadings matrices, respectively, while E are the X -residuals of the model. Based on this decomposition, the PCR method proceeds by building an MLR model on the scores computed in the PCA step. Accordingly, the regression model for PCR is

$$Y = \hat{Y} + E_Y = TB + E_Y$$

where the subscript Y was added to the Y -residual matrix to differentiate it from that of the X -block and B is the matrix of regression coefficients for the MLR model relating the dependent variables Y to the PC scores of the independent block T . Extending the definition of the MLR model regression coefficients, the matrix B can be computed as

$$\mathbf{B}_{\text{PCR}} = T^+ Y = (T'T)^{-1} T'Y$$

where the regression coefficient matrix B relates the dependent matrix Y to the X -scores (T).

With respect to MLR, as PCR involves a projection step, where the data are represented on a low-dimensional latent variable space, there is the need of deciding what the complexity of this space should be or, in other terms, how many PCs are needed (refer to Section 65.3.3.5). In general, there is a trade-off in selecting the optimal number of components: including too few components, which could lead to models not able to fit X well and to predict Y accurately, whereas the use of too many

components can result in overfitting Y and X . As a consequence, this may result in unreliable predictions on new samples. Therefore, the choice of model complexity is normally accomplished through some sort of validation procedure (see Section 65.6), in which the optimal number of PCs is selected as the one leading to the lowest prediction error on validation estimates.

One possible drawback of PCR modeling is that it relies on using the PCs as predictors for the responses, but PCs do not necessarily correlate with Y . Indeed, the main characteristic of PCA is to extract features that capture as much as possible the variation in X ; however, in cases where many sources of uninformative variation and/or a high level of noise are present, they can be poorly related to the Y (and, hence, not predictive). To overcome this problem, some authors suggest to only choose those latent variables (PCs), correlating maximally with the responses [33].

65.4.4 Partial Least Squares Regression

As discussed in Section 65.4.3, PCR is a two-step process, in which the projection stage is separated and independent from the regression one and this can lead to the drawback that the components that are extracted in the decomposition step, based only on the information about the X -matrix, can be poorly predictive for the Y block. Starting from these considerations, an alternative method was proposed called PLSR [15, 34, 35], in which information in Y is actively used for the definition of the latent variable space. PLSR extracts components (known as PLS factors), which compromise between explaining the variation in the X -variables and predicting the responses in Y . This corresponds to a bilinear model, whose mathematical structure is summarized as

$$\begin{aligned} X &= TP' + E \\ Y &= \hat{Y} + E_y = UQ' + F \end{aligned}$$

This definition is formally identical to that for PCR, although the calculated components and the model coefficients are not the same, as the two projections are governed by different criteria. In particular, a major difference revolves around the way scores T are defined for PLSR, that is, they are defined in a way such that they are relevant both for interpretation and prediction, through the statistical concept of covariance. Accordingly, PLSR is a sequential algorithm: the PLS latent variables are computed so that the first PLS component is the direction of maximum covariance with the dependent variables, the second PLS component is orthogonal to the first and has maximal residual covariance, and so on.

65.4.4.1 PLS Scores As was the case in PCA (Section 65.3.3.2), the typical scores plot as a 2D scatter plot provides a map of the objects where similarity between objects and groups of objects can be interpreted. In the case of PCR, the scores are computed from PCA on the X -data.

The major difference between PCA/PCR and PLS scores is that in PLSR the scores are estimated from X and the loading weights and are thus based on the covariance between X and Y . This means that PLS scores capture the part of the structure in X , which is most predictive for Y . Therefore, although PCA and PLS scores can be visualized and interpreted in a similar way, they are not the same. PCA scores model variations in X only, whereas PLS scores model variations in X most related to Y . This is an important concept that required this further elaboration.

65.4.4.2 X -Loadings in PLSR Assume the model $X = TP^T + E$ and then the loadings reflect the importance of all X -variables for each component/factor. For spectral data a plot of the loadings as a line plot may indicate if the factor carries information: if the vector looks like random numbers, it should not to be included as structure that can be modeled.

As the loadings in PCR are normally scaled to unit variance, there can be no ad hoc rule set if a loading value above a certain value is important. As an alternative the correlation loadings can be plotted, which are simply the correlation between the original variable and the scores vectors (refer to Section 65.3.3.3). For PLSR the loadings do not exactly have length=1.0, but correlation loadings are still valuable for interpretation about how the variance in X is modeled from the scores.

65.4.4.3 Y -Loadings in PLSR The Y -loadings express the importance of the individual Y -variables for the factors $1:F$. The following equation depicts how Y is decomposed in the U scores and Y -loadings Q :

$$Y = UQ' + F$$

Refer to Section 65.4.4.8 regarding more details on the algorithmic details of the PLSR method.

In the version of the PLSR algorithm where the vectors w_a are scaled to unity, the inner relation coefficients are 1.0 and thus

$$Y = TQ' + F$$

65.4.4.4 Loading Weights Loading weights are specific to PLSR (they have no equivalent in PCR) and express how the information in each x -variable relates to the variation in Y summarized by the u -scores. They are called loading weights because they also express, in the PLSR algorithm, how the T -scores are to be computed from the X -matrix to obtain an orthogonal decomposition. The loading weights are typically normalized to 1.0. Variables with large loading weight values are important for the prediction of Y . The first loading weight vector in case of only one response variable is the covariance (or correlation if the variables are scaled to unit variance) between the individual x -variables and y .

X-loadings and Y-loadings or loading weights and Y-loadings are often shown together in 2D plots and interpreted similarly to loadings from PCA. Figure 65.26 provides an excellent example of the comparison of loadings and loading weights using a set of NIR spectra of gasoline samples with some samples containing an additive and the majority containing no additive.

In the top left plot of Figure 65.26, the NIR spectra of the samples are shown. Two classes of samples are present, those with an additive (yes) and those without an additive (no). The spectra are grouped based on this classification, and it can be seen that above 1380 nm, there are spectral differences between the two sample types.

The scores plot for the PLSR is shown in the bottom left quadrant of Figure 65.26. It can be seen that the samples with an additive separate from those without the additive, primarily along score factor 1. The top right-hand plot is the PLS loadings for the analysis. Since the loadings represent the structure contained in X , the spectral features greater than 1380 nm are found to be important. However, when the loading weights for these samples (shown in the bottom right quadrant), this region is weighted less, as they do not contribute to predicting Y . This example provides an excellent overview of the difference between loadings and loading weights, and this type of analysis must be performed as part of any PLSR model interpretation.

The scores and loadings may also be visualized together in a biplot [36], but there is no “truth” when it comes to scaling of the axes in a biplot and caution should be used in the interpretation of the relative position of the objects and variables [37].

65.4.4.5 Regression Coefficients Regardless of the type of regression, the model can be represented in terms of regression coefficients (B). The ideal situation occurs when the individual elements in the regression vector are directly interpretable as to provide the true model of the system. This is however only the case if the x -variables are orthogonal as in a factorial design [38]. When some of the x -variables are correlated, the situation of indeterminacy due to collinearity is encountered. As already discussed in this chapter for the latent variable methods, they handle collinearity from a numerical point of view but not necessarily from an interpretational point of view. Let a model for body weight be a function of two x -variables: height and shoe size. In the case of MLR, the model will use all the variance in X and the coefficient for shoe size may be negative although it clearly is a positive correlation between shoe size and body weight. In this case the underlying dimensionality is one and the regression vector from PLS regression will in this case reflect the true relationships between X and Y in the first PLS factor. Nevertheless, with latent variable models and a correct assessment of the dimensionality of the model from proper model validation, the regression vector may give valuable information about the underlying phenomena, be they physical, chemical, or biological in nature.

Assume that in a system there is one response variable Y of interest and other sources of systematic variance (e.g., other constituents, properties) that give rise to signals in X . Under noise-free conditions the regression vector estimated by PLSR is, up to

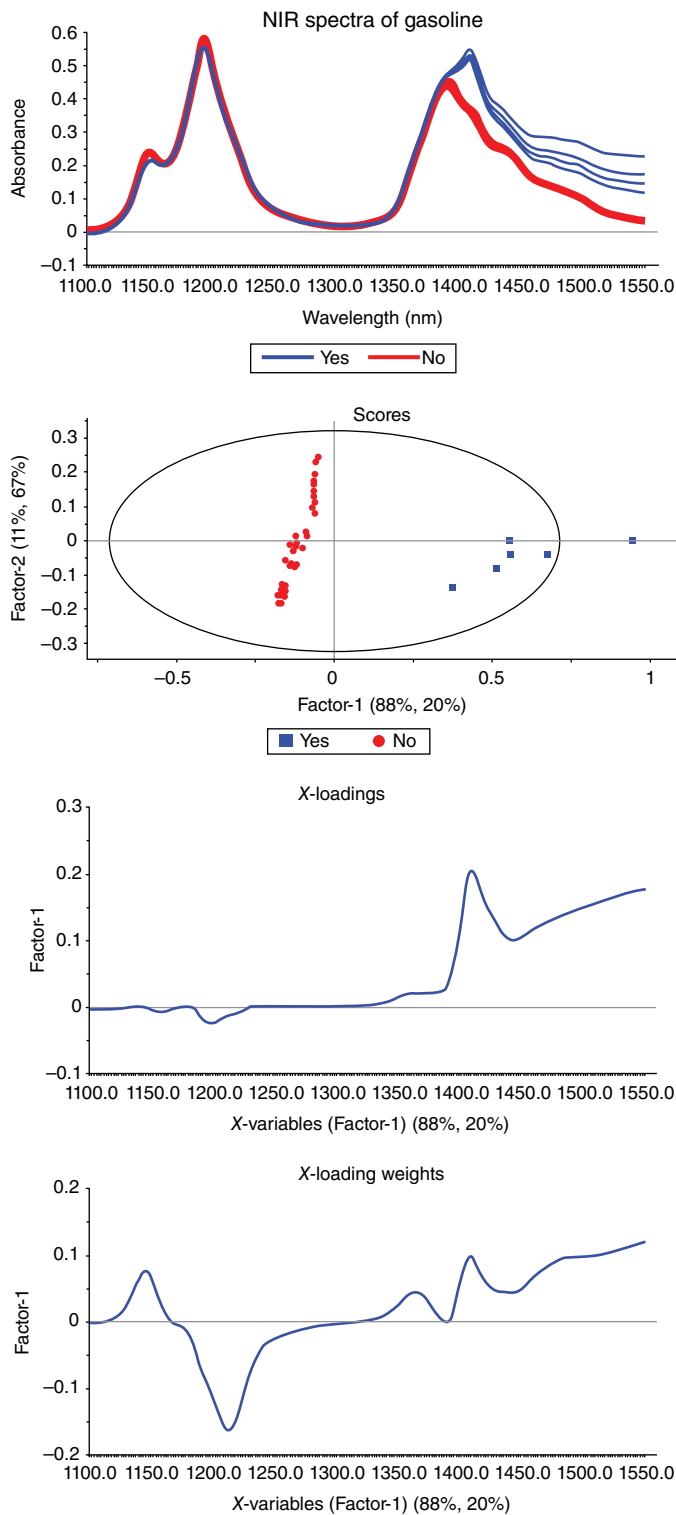


FIGURE 65.26 Comparison of PLS loadings and loading weights for the NIR spectra of gasoline samples.

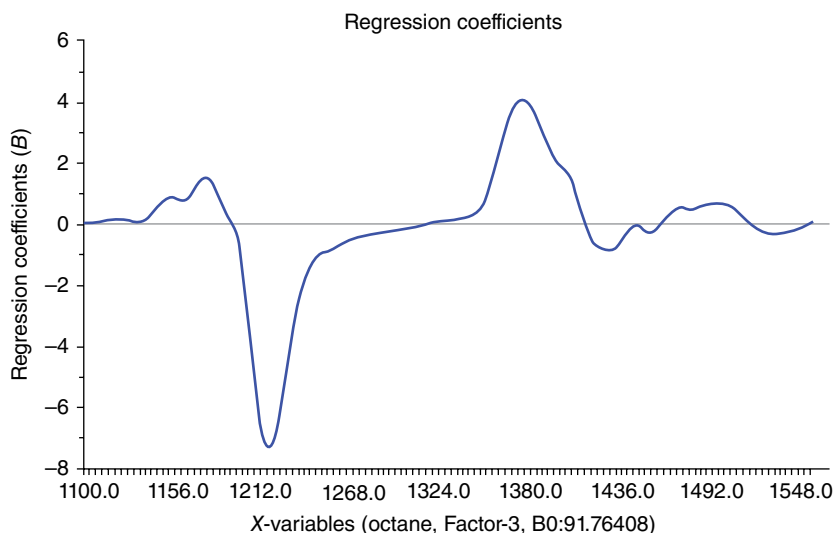


FIGURE 65.27 Regression coefficients for the prediction of octane number in gasoline samples using PLSR.

normalization, the net analyte signal. This vector is defined as the part of the response Y that is orthogonal to the response vectors of all other constituents/properties. In the case of unstructured noise, PLSR computes a final regression vector that is not in general purely proportional to the net analyte signal vector but has the important property of being optimal under a mean squared error of prediction criterion [39].

For the gasoline example introduced previously, the regression coefficients for the prediction of octane number are shown in Figure 65.27. It is noted in this plot that three PLS factors were used to generate the model (based on validation diagnostics). It can be seen in Figure 65.27 that the region above 1380 nm does not contribute to modeling octane number. The important regions can be found between 1120–1220 nm (interpreted as aromatics and straight chain hydrocarbon content) and 1350–1380 nm (hydrocarbon content).

65.4.4.6 Predicted versus Reference Plot The predicted versus reference plot should show a straight line relationship between predicted and reference values, ideally with a slope of 1 and a correlation close to 1. The predicted versus reference plot for the gasoline example is shown in Figure 65.28 for a three-PLS factor model.

From a philosophical point of view, in some software packages and in the chemometric literature, this predicted versus reference plot is sometimes called the predicted versus actual plot. The use of the term “actual” gives the indication that the reference value is actually the absolute truth, where in fact it has some form of error associated with it. The term “reference” indicates that the value was generated from a reference method and therefore carries the connotation that it has error associated with it.

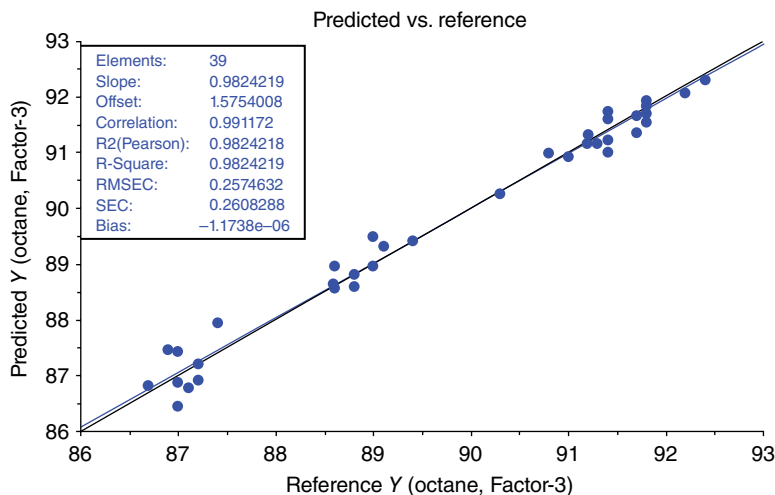


FIGURE 65.28 Predicted versus reference plot for octane number in gasoline analysis.

65.4.4.7 Residuals and Diagnostic Tools in PLSR

X- and Y-Residuals All the bilinear models described in this chapter operate by fitting a portion of the variation in the *X*- and the *Y*-blocks and can be summarized as

$$X = \hat{X} + E_x$$

$$Y = \hat{Y} + E_y$$

where the aim of the modeling phase is normally to find estimates that fit as well as possible the corresponding block matrices. However, inspection of the residual matrices E_x and E_y can provide useful information about the model quality. In this respect, residuals can be investigated at different levels, and different information can be obtained depending on the level considered. Indeed, residual analysis can be carried out for the detection of outliers, for the identification of systematic variation that was not accounted for by the model (especially when variables are homogeneous, such as in a spectrum or a chromatogram), for the detection of drifts or trends in the data, or in general to define a distance to the model. From a practical standpoint, each of these tasks is better accomplished by adopting a proper representation of the information contained in the residuals. In this framework, the first way of looking at the residuals is to consider the distribution of $e_{x_{i,j}}$ and $e_{y_{i,j}}$, which are the elements of the matrices E_x and E_y , respectively. Many models assume random Gaussian noise or, in general, symmetrically distributed residuals. Therefore, plotting the residuals or verifying whether the distributional assumptions are met (e.g., by means of normality tests) can provide a good diagnostics of the model. As an example, in Figure 65.29a the distribution of the *X*-residuals for a situation where no anomalies are present is shown: the histogram shows almost perfect symmetry and assumes a Gaussian-like

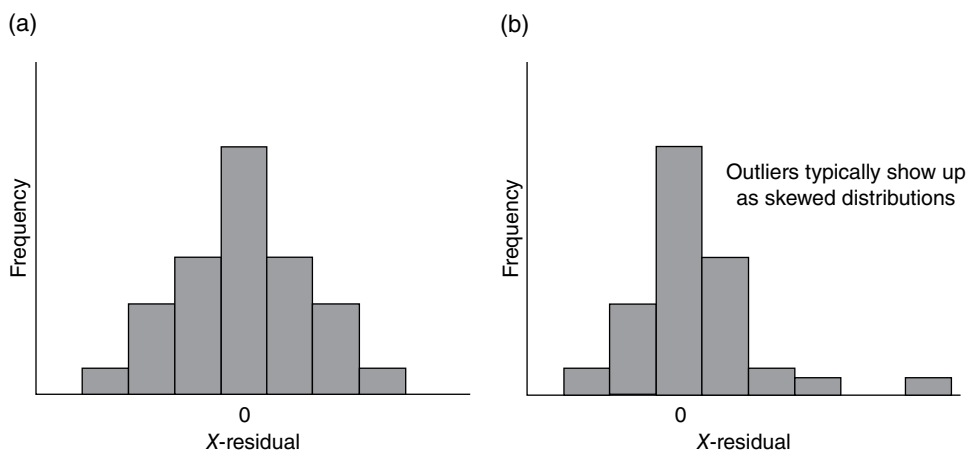


FIGURE 65.29 Distribution of X -residuals for a well-behaved model (a) and a model with an outlier (b).

shape, as expected. On the other hand, the distribution of the X -residuals for a case where outlying observations are present in the data set is reported in Figure 65.29b. It is evident from this figure that the histogram is no longer symmetric and that there is an increased probability associated to high values of the residuals, indicating that some anomalies are present in the data.

A second set of diagnostic measures can then be inspected when considering that the residual matrices E_X and E_Y have the same dimensions as the fitted matrices \hat{X} and \hat{Y} and therefore it is possible to extract rows and columns to investigate the residual variances associated with one particular sample or variable compared to the rest of the model. It is customary to summarize the variation in one direction or the other by calculating the sum of squares of the vectors corresponding to the individual samples (or variables). In particular, the sum of squares residuals of the i^{th} sample can be expressed, for the X - and the Y -blocks, as

$$e_{X,i}^2 = \|e_{X,i}\|^2 = \sum_{j=1}^v e_{X,ij}^2$$

$$e_{Y,i}^2 = \|e_{Y,i}\|^2 = \sum_{k=1}^r e_{Y,ik}^2$$

where v is the number of predictors, r is the number of responses, and e_{X_i} and e_{Y_i} are the i^{th} rows of the matrices E_X and E_Y , respectively, while $e_{X_{i,j}}$ and $e_{Y_{i,k}}$ are the corresponding elements. When plotting the values of the sum of squared residuals for the samples, different situations can occur, the most frequent of which are shown in Figure 65.30. In particular, Figure 65.30a shows a random distribution of the summed squared residuals for the different samples, as expected when no outlying observations occur. On the

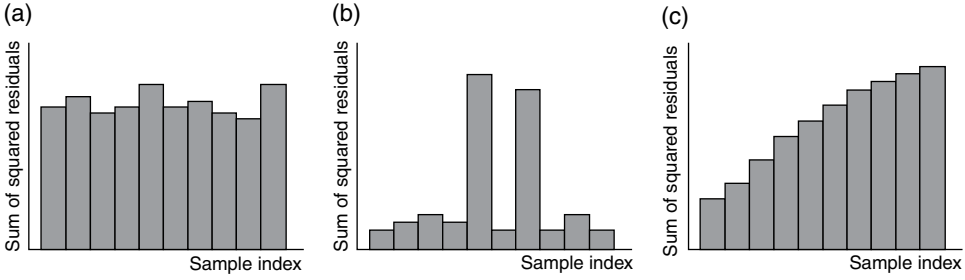


FIGURE 65.30 Example situations of residual patterns in well behaved and models that are not well behaved. (a) Well behaved model, (b) two outliers in the model, and (c) trending in residuals.

other hand, the situation in which two samples are anomalous with respect to the others is plotted in Figure 65.30b; in particular, these two samples are characterized by containing an additional interferent, which is not present in the rest of the objects, and this, in turn, results in a significantly higher value of the residuals. Lastly, the situation where there is a trend along the objects, which is not captured by the model, is depicted in Figure 65.30c.

Besides plotting the sum of squares for each sample, sometimes it can be more useful also to plot, sample-wise, the whole vector of residuals, in order to evidence the presence of unmodeled systematic structure (especially when the variables are homogeneous), or to identify blocking effects. Indeed, when there are sources of systematic variation that are not explained by the model, the residuals for that particular sample are no longer randomly distributed and present a structured shape.

Analogously, the sum of squared residuals for the individual predictors or response variables can be obtained by summing over the analyzed samples, according to

$$e_{X,j}^2 = \|e_{X,j}\|^2 = \sum_{i=1}^m e_{X,ij}^2$$

$$e_{Y,k}^2 = \|e_{Y,k}\|^2 = \sum_{i=1}^m e_{Y,ik}^2$$

where m is the number of samples and $e_{X,j}$ and $e_{Y,k}$ are the j th column of E_X and the k th column of E_Y , respectively, while $e_{X,ij}$ and $e_{Y,ik}$ have the same meaning as previously defined.

Error Measures The simplest and most efficient measure of the uncertainty on future predictions is the root mean square error (RMSE). This value (one for each response) is a measure of the average uncertainty that can be expected when predicting Y -values for new samples, expressed in the same units as the Y -variable. The results

of future predictions can then be presented as “predicted values $\pm 2 \times \text{RMSE}$ ” (at $\sim 95\%$ confidence).

This measure is valid provided that the new samples are similar to the ones used for the development of the calibration model; otherwise, the prediction error might be much higher. For an MLR calibration model, the RMSE from calibration (i.e., RMSEC) is expressed by

$$\text{RMSEC} = \sqrt{\frac{\sum_{i=1}^N (y_i - \hat{y}_i)^2}{N - df}}$$

where N is the number of samples and df is the number of variables -1 . For test set validation for any regression method, the formula for the root mean square error of prediction (RMSEP) is

$$\text{RMSEP} = \sqrt{\frac{\sum_{i=1}^N (y_i - \hat{y}_i)^2}{N}}$$

For cross validation the formula is as for test set validation but should formally be reported as RMSECV. Validation residual, explained variances and RMSEP are also computed in exactly the same way as calibration variances, except that prediction residuals are used instead of calibration residuals.

Plots of RMSE as a function of number of factors (for latent variable methods) are also used to find the optimum number of model components (similar to the residual variance plots discussed in Section 65.3.3.5). When validation residual variance is minimal, RMSEP is also minimized, and the model with an optimal number of components will have the lowest expected prediction error.

RMSEP can and should be compared with the precision of the reference method (refer to Section 65.2). The error of the reference method is sometimes referred to as the standard error of laboratory (SEL). It is of utmost importance to have an estimate of this precision to evaluate to what extent the model has a sufficiently good predictive ability given the actual application. It cannot be expected that RMSEP be any lower than 1.4 times SEL [40]. The RMSEP is a sum of the sampling error, measurement error, model error, and reference method error.

An alternative error measure is the predicted residual sums of squares (PRESS):

$$\text{PRESS} = \sum_{i=1}^N (y_i - \hat{y}_i)^2$$

As PRESS is reported as a square number, it does not directly relate to the values (or range) of the response variable Y .

Some other useful statistics available when assessing regression models are described as follows:

1. Bias

The bias is the average value of the difference between the reference and predicted values:

$$\text{BIAS} = \frac{\sum_{i=1}^N (y_i - \hat{y}_i)^2}{N}$$

2. Standard error of prediction (SEP)

SEP is the standard deviation of the prediction residuals:

$$\text{SEP} = \sqrt{\frac{\sum_{i=1}^N (y_i - \hat{y}_i - \text{BIAS})^2}{N - 1}}$$

3. Ratio of standard error of prediction to sample standard deviation (RPD)

RPD is the ratio of the standard deviation for the response variable Y and the RMSE. There exist some ad hoc rules regarding how RPD relates to a “good,” “fair,” or “bad” model:

$$\text{RPD} = \frac{s_y}{\text{SEP}}$$

0–2.3 very poor

2.4–3.0 poor

3.1–4.9 fair

5.0–6.4 good

6.5–8.0 very good

8.1+ excellent

However, it must be stated that RPD depends on the range used for Y . In this respect RMSE is a more generic error measure. Similarly, this is why the correlation coefficient R^2 does not necessarily provide a good indication about a model’s predictive ability.

The least squares effect also needs to be taken into consideration when discussing the range of Y . For all regression methods with least squares as the objective, overprediction of the low values and underprediction of the high values for Y can occur. Thus, if it is expected that many future samples will lie far from the mean of the model, the model can be corrected to give a bias of 0 and slope of 1. This will now not give the least squares solution but avoid over- and underprediction. This principle is illustrated in Figure 65.31.

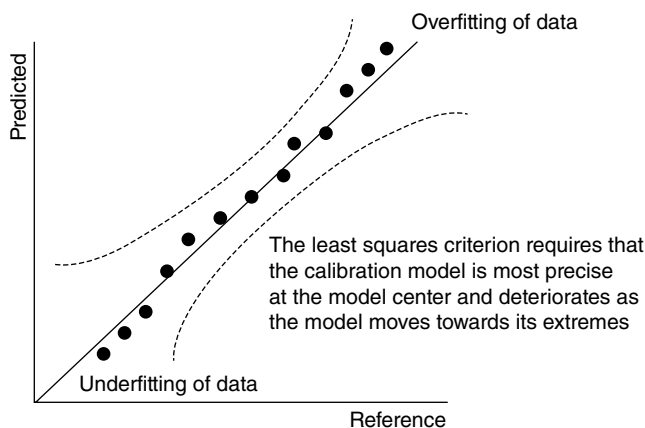


FIGURE 65.31 Error distribution of the least squares fit.

65.4.4.8 Algorithms for PLSR Although there are a number of PLSR alternatives [2, 3, 41], the NIPALS algorithm is still the most commonly used algorithm. The next sections describe the single variable variant (PLS-1) and the multiple response variant (PLS-2) algorithms.

PLS-1 Algorithm

Step 1: Scale and center the X - and Y -variables

$$X_0 = X - 1\bar{x}' \text{ and } Y_0 = Y - 1\bar{y}$$

Step 2: Use the variability remaining in y to find the loading weights w_a where a represents the selected number of PLS factors to calculate,

$$X_{a-1} = y_{a-1}w_a' + E$$

Scale the vector to length 1:

$$\hat{w}_a = cX_{a-1}'y_{a-1}$$

where c is the scaling factor that makes the length of the final \hat{w}_a equal to 1, that is,

$$c = (y_{a-1}'X_{a-1}X_{a-1}'y_{a-1})^{-0.5}$$

Step 3: Estimate the scores \hat{t}_a using the local model

$$X_{a-1} = t_a \hat{w}_a' + E$$

Since $\hat{w}_a' w_a = 1$ the least squares solution is

$$\hat{t}_a = X_{a-1} \hat{w}_a$$

Step 4: Estimate the X -loadings p using the local model

$$X_{a-1} = \hat{t}_a p_a' + E$$

which gives the least squares solution

$$\hat{p}_a = \frac{X_{a-1}' \hat{t}_a}{\hat{t}_a' \hat{t}_a}$$

Step 5: Estimate the Y -loadings q , using the local model

$$y_{a-1} = \hat{t}_a q_a + F$$

where F is the residual term after fitting the model to Y .

This gives the solution

$$\hat{q}_a = \frac{y_{a-1}' \hat{t}_a}{\hat{t}_a' \hat{t}_a}$$

Step 6: Create new X and y residuals by subtracting the new PLS factor as follows,

$$\hat{E} = X_{a-1} - \hat{t}_a \hat{p}_a'$$

$$\hat{F} = y_{a-1} - \hat{t}_a \hat{q}_a$$

Replace the former X_{a-1} and y_{a-1} by the new residuals \hat{E} and \hat{F} and increase a by 1.

Step 7: Determine A the number of valid PLS factors to retain in the calibration model.

Step 8: Compute \hat{b}_0 and \hat{b} for A PLS factors to be used in the prediction model,

$$\hat{b} = \hat{W} (\hat{P}' \hat{W})^{-1} \hat{q}$$

$$b_0 = \bar{y} - \bar{x}' \hat{b}$$

PLS-2 Algorithm The PLS-2 algorithm is almost identical to the PLS-1 algorithm, except the Y vector is replaced by a matrix Y_{ij} where the dimension j represents the number of y -variables to be modeled by the algorithm. In the early days of chemometrics, there was a distinction between PLS-1 and PLS-2 because the PLS-2 is an iterative

algorithm and this increased the computational time. Today there is really no need to make this distinction apart from a conceptual point of view.

Step 1: Define a temporary y-score \hat{u}_a , for example, the column in Y with the largest variance.

Step 2: Use \hat{u}_a that summarizes the remaining variability in Y to find the loading weights \hat{w}_a by least squares fitting using the local model

$$X_{a-1} = \hat{u}_a w'_a + E$$

Scale the vector to length 1. The least squares solution is

$$\hat{w}_a = c X'_{a-1} \hat{u}_a$$

where c is the scaling factor that makes the length of the final \hat{w}_a equal to 1, that is,

$$c = \left(\hat{u}'_a X_{a-1} X'_{a-1} \hat{u}_a \right)^{-0.5}$$

In the first iteration \hat{u}_a is given a starting value, typically the column in Y with the largest sum of squares.

Step 3: Estimate the scores \hat{t}_a using the local model

$$X_{a-1} = \hat{t}_a \hat{w}'_a + E$$

Since $\hat{w}'_a \hat{w}_a = 1$ the least squares solution is

$$\hat{t}_a = X_{a-1} \hat{w}_a$$

Step 4: Estimate the X-loadings p using the local model

$$X_{a-1} = \hat{t}_a p'_a + E$$

which gives the least squares solution

$$\hat{p}_a = \frac{X'_{a-1} \hat{t}_a}{\hat{t}'_a \hat{t}_a}$$

Step 5: Estimate the Y-loadings q , using the local model

$$\hat{u}_a = \hat{t}_a q_a + F$$

where F is the residual term after fitting the model to Y .

This gives the solution

$$\hat{q}_a = \frac{\hat{u}_a \hat{t}_a}{\hat{t}_a' \hat{t}_a}$$

Step 6: Test for convergence to see that all elements no longer change significantly from the last iteration.

Step 7: If convergence has not been achieved, estimate temporary factor scores u_a using the model,

$$Y_{a-1} = u_a \hat{q}_a' + F$$

Giving the least squares solution

$$\hat{u}_a = Y_{a-1} \hat{q}_a (\hat{q}_a' \hat{q}_a)^{-1}$$

Return to step 2.

Step 8: If convergence has been reached, create new X and y residuals by subtracting the new PLS factor as follows,

$$\hat{E} = X_{a-1} - \hat{t}_a \hat{p}_a'$$

$$\hat{F} = y_{a-1} - \hat{t}_a \hat{q}_a$$

Replace the former X_{a-1} and y_{a-1} by the new residuals \hat{E} and \hat{F} and increase a by 1.

Step 9: Determine A the number of valid PLS factors to retain in the calibration model.

Step 10: Compute \hat{B} and b_0' for A PLS factors to be used in the prediction model,

$$\hat{b} = \hat{W} (\hat{P}' \hat{W})^{-1} \hat{Q}'$$

$$b_0' = \bar{y}' - \bar{x}' \hat{B}$$

For more details on the algorithms for PLS-1 and PLS-2, the interested reader is referred to the literature [15, 18].

65.5 MULTIVARIATE CLASSIFICATION

In Section 65.3 the topic of EDA was discussed in great detail. The methods of cluster analysis discussed were defined as unsupervised methods, that is, they looked for the natural patterns in the data without being guided by an external classification rule.

The method of PCA discussed in Section 65.3.3 provides a highly graphical environment for detecting clusters within data set of samples. By making a separate class model for each cluster, a library of clusters can be developed and thus a classification rule can be established in order to group new samples into known classes. This is known as supervised classification, or pattern recognition, and this topic is an important area known generally as multivariate classification.

Multivariate classification methods take full advantage of the multivariate nature of the data. Although there are many multivariate classification methods available, this section will only look at four commonly used methods:

1. Linear discriminant analysis (LDA)
2. Soft independent modeling of class analogy (SIMCA)
3. Partial least squares discriminant analysis (PLS-DA)
4. Support vector machine classification (SVMC)

The aforementioned classification methods will be discussed in order and comparisons will be made to their regression counterparts already discussed in the previous sections.

65.5.1 Linear Discriminant Analysis

By definition, discriminant analysis aims to find discriminating features that separate samples into different data classes. In the case of the classical LDA, also known as Fisher's LDA [42], the algorithm aims to find the discriminating axes, that is, linear combinations of the original p variables that optimally separate two or more classes.

In order to separate samples into classes, a training set of two or more known samples must be available to develop the classification rule. There are a number of distance measures that can be used to optimally separate classes including:

1. Linear separators
2. Quadratic separators
3. Mahalanobis distance

Figure 65.32 shows diagrammatically how the linear and quadratic methods separate samples into classes.

When separating the two-class problem, the axes used to discriminate between the classes may be viewed as a projection onto $A = 1$ dimensions and the discrimination axis could be viewed as a component vector separating the two classes (refer to Fig. 65.32).

The regression equivalent of LDA is MLR (Section 65.4.2). LDA suffers from the same collinearity effects that MLR does and also requires more samples than variables being measured. To overcome these issues, a version of LDA, known as PCA-LDA, is

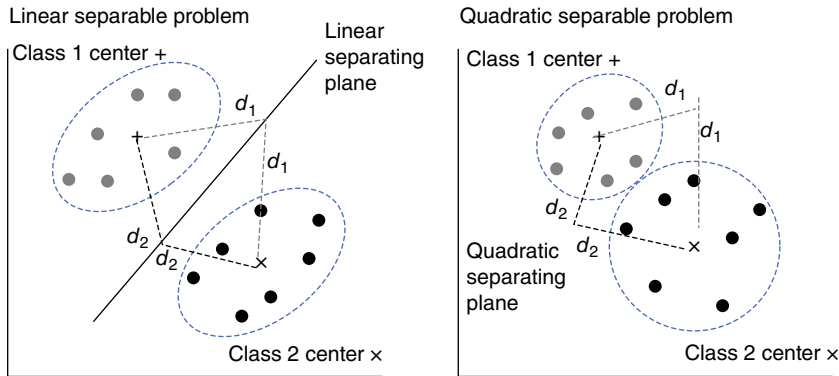


FIGURE 65.32 Linear and quadratic separation in LDA.

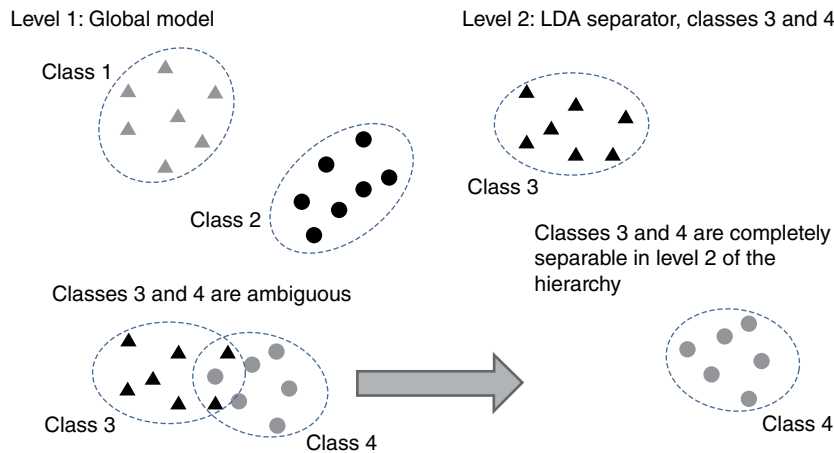


FIGURE 65.33 Resolving classification ambiguities using hierarchical models.

available to overcome collinearity issues when higher-dimensional data (particularly spectral data) are being analyzed.

Two other major limitations of LDA are that the assumption of a common covariance structure in the classes exists, which is very rare in practice, and LDA is usually only suited to the two-class separation problem. The latter point is further exacerbated by the limitation that if LDA does not uniquely classify a sample into a unique class, it puts the sample into the class that has the closest distance to the class center. This limitation can have serious ramifications, particularly in applications such as pharmaceutical raw material identification or antiterrorist hazardous material applications.

For these limitations alone, LDA is best suited when used in hierarchical classification schemes. In many instances, when a classification model contains many classes, ambiguities may arise due to limitations in the measurement system(s) used to separate the classes. Figure 65.33 shows graphically the concept of ambiguity and how a hierarchical model can be used to resolve the two-class problem.

The main model is commonly known as the global model. Here, a new sample is assessed by each class library in the model and the highest match is found based on some predefined statistical limits. If a sample is classified into two or more classes, the sample is said to be ambiguous with the global model, that is, a unique separation is not possible. To overcome this problem for the two-class ambiguity, LDA is ideal for the following reasons:

1. It is ideal for separating the two-class problem based on its simplicity.
2. Since the hierarchical model has already determined that the sample is either one of two classes, the use of LDA is more reliable since it will classify the sample into one of the two classes without the risk of the sample being mistaken for something not in the global model.

Due to the reasons provided earlier, a classification method that can detect a null class, that is, detect that a new sample is a complete unknown, is much more useful than LDA. This method known as SIMCA is discussed in the next section.

65.5.2 Soft Independent Modeling of Class Analogy

PCA allows a user to develop class models based on their clustering in PC scores space. Once this library of PCA models is developed, a rule must be established that can direct new samples to the class(es) they belong to. The method of SIMCA provides such a rule.

SIMCA was first introduced by Wold in 1976 [43] and it allows new samples to display their uniqueness as well as their common patterns, which provides the advantage of SIMCA over LDA to be able to reject samples as not belonging to any class, rather than put the sample into the closest class.

65.5.2.1 Practical Steps for Building a SIMCA Model As with any empirical modeling strategy, a training phase is required. The term soft in SIMCA relates to the model being empirical and therefore a representative set of samples must be modeled to understand the natural class variability. It is important to note here that since PCA (or PLSR) is used to develop class models in SIMCA, if all samples in the set have identical variability, then no PCA/PLSR model is possible. This is one of the potential drawbacks of the SIMCA methodology.

By representative (refer to Section 65.2) each class must consist of samples that will represent future sample variability. This variability is considered to be the natural variability to be expected for this sample type. Once the data are collected, a PCA/PLSR model is calculated for all classes to understand the variability within and between classes. This represents the ANOVA problem defined in Section 65.2, and in general, PCA is a multivariate version of ANOVA and can be described as a visual ANOVA. Figure 65.34 shows a PCA scores plot with a number of different classes identified.

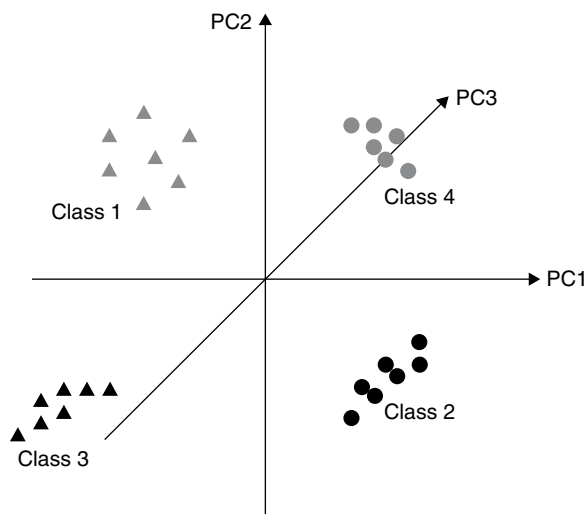


FIGURE 65.34 PCA scores plot showing class separation of data.

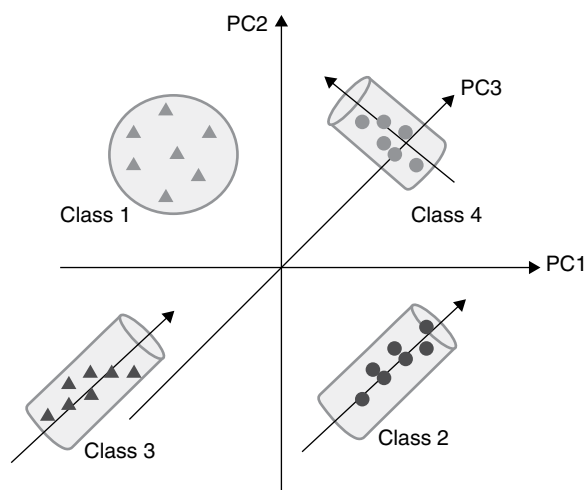


FIGURE 65.35 PCA scores plot showing class separation of data with class limits.

To place objective limits around each class, the diagnostic tools associated with PCA/PLSR are used to determine confidence intervals. In particular, the statistics, sections “X-Residuals” and “Leverage” are used. Figure 65.35 shows the data in Figure 65.34, this time with statistical limits around the classes.

To develop a SIMCA training model, each class must be saved as an individual PCA model and validated, preferably with a test set (see Section 65.6.1). The development of individual PCA models results in what is known as a disjoint class model, since the individual libraries retain their uniqueness and must be joined using an external rule.

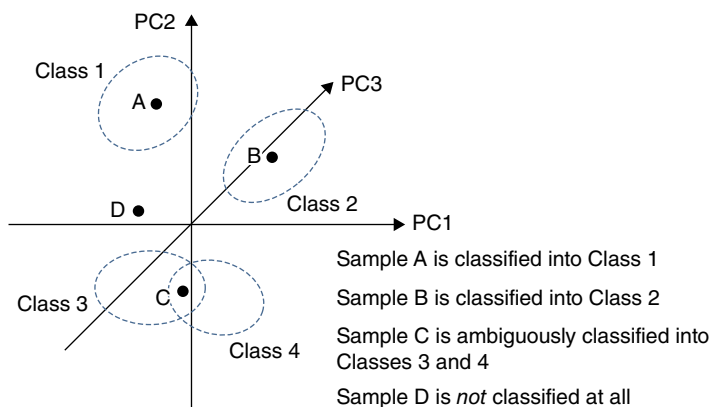


FIGURE 65.36 Three cases of classification outcome when using SIMCA.

It is important to note here that even though SIMCA uses individual PCA models as class rules, there must be complete commonality of the PCA models to each other in terms of number of variables and data preprocessing, that is, each PCA library model must have identical variables and must be preprocessed exactly the same way. This is to ensure complete statistical representation of the results. It is also a potential drawback to the SIMCA method; however, these issues can be overcome in the case of ambiguities by using a hierarchical model. Also, the trick of weighting down variables for different ranges of the total number of variables in the individual models may be used.

The next step in SIMCA library development is to select the validated PCA class models and enter them into a scheme that allows new samples to be assessed by each model for class assignment. In SIMCA, there are three possible outcomes:

1. Unique classification: The new sample resides in one class only.
2. Ambiguous classification: The sample could be a member of two or more classes simultaneously.
3. No classification: The new sample is not a member of any of the class models present in the SIMCA scheme.

It is point three that provides SIMCA with the most versatility compared to most other classification methods. Figure 65.36 provides a diagram of all three cases described earlier.

When validating a SIMCA model for practical usage, there are two approaches that should be used:

1. Use the model to predict the training set: Although this is highly biased, it should be the first step to provide assurance that the library is capable of predicting itself. During this stage, the confidence intervals for the model can be fine-tuned.

Sample ID	Class 1	Class 2	Class 3	Class 4
A	X			
B		X		
C			X	X
D				

FIGURE 65.37 Classification table showing all three classification scenarios for the data shown in Figure 65.34.

2. Use the model to classify a separate test set of samples to understand how the library will perform on new samples in the future.

The next sections describe the diagnostic and graphical tools available to assess the performance of SIMCA models.

65.5.2.2 Diagnostic Tools for SIMCA The SIMCA approach generates a multitude of statistics relevant for validating the performance of the classification model. Effectively, the class models are wrapped into an envelope that defines the class membership boundaries. The major tools and statistics to be described for SIMCA are listed as follows:

1. The classification table
2. The Coomans plot
3. The distance versus leverage plot
4. Model distance
5. Variable discrimination power
6. Modeling power

The Classification Table To provide an overview of the classification results, the classification table is the first diagnostic tool to check. It is a standard table with sample ID's listed as the rows and class model names as the columns. Each time a sample is classified into a class, it is marked in the table by some form of symbol. If the sample simultaneously belongs to two or more classes, there will be a symbol shown for each class the sample belongs to. In the event of no classification, the table entry for the particular sample will remain blank. Figure 65.37 shows an example classification table for the example shown in Figure 65.36, that is, with the three possible classification results shown.

Coomans Plot Named after the Belgian chemometrician Prof. Danny Coomans [44], this plot shows the sample to model distances for the training and validation sets two

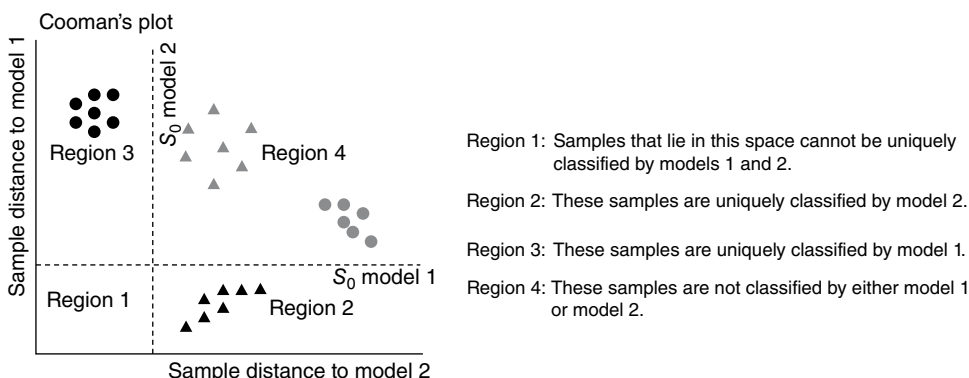


FIGURE 65.38 The Coomans plot and its interpretation.

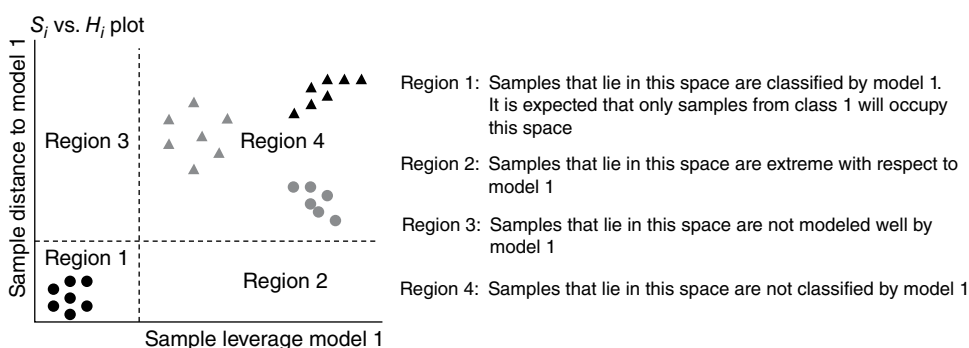


FIGURE 65.39 Construction of the S_i versus H_i plot and how it is interpreted.

models at a time, that is, it shows the orthogonal distances between models. Two sets of limits (S_0) are shown on the horizontal and vertical scales of the plot showing where the statistical limits are located for the two models under investigation.

Since the Coomans plot is a pairwise model comparison tool, it is used to assess the degree of model overlap in the case of ambiguity. To interpret the plot, if any samples lie in the boundary between the origin and where the two limits cross, then the individual library models are not capable of uniquely separating the classes. If the samples from both models lie orthogonally separate in their own classes, then the models are capable of uniquely separating the two classes under investigation. Figure 65.38 provides an overview of the Coomans plot and its interpretation.

The Distance versus Leverage (S_i vs. H_i Plot) This plot is also called the membership plot [18] because it shows the statistical limits used to envelop the class model. The statistic S_i is similar to X -residual discussed in section “ X -Residuals” and is the residual standard deviation of the sample to the model. H_i is the leverage and shows the distance to model center for each sample. Figure 65.39 shows diagrammatically how S_i and H_i are represented in scores space and how this translates to the S_i versus H_i plot.

Interpretation of the S_i versus H_i plot is simple. It compares a selected class model to all other class models in the library. If the model is able to uniquely classify samples of its own class, the samples will all lie between the S_i and H_i limits for the class at a specified confidence interval. The leverage limit is determined by the number of PCs used in the model. This plot is very similar to the influence plot discussed in section “Influence Plots.” There is from the conceptual point of view nothing wrong by applying the Hotelling’s T^2 statistics as the limit of distance within the model. As the number of samples is often limited for the individual classes, the leverage-based rule of $3\times$ the average leverage was historically assumed to be better than using a distribution-based metric (Hotelling’s T^2 is based on an F-distribution).

Samples close to the origin of the S_i versus H_i plot are considered to be real members of that class. As the sample moves toward the leverage limit, the sample is considered to be extreme in characteristics for that class. If it exceeds the boundary but is within the S_i boundary of the model, then the sample is too extreme to be part of the model. An analogous situation is described as follows. Consider the origin representing an espresso coffee and as the leverage moves toward the limits, the coffee is becoming more latte in characteristics, then the sample exceeds the limit, and it is more latte in character than espresso. In general, leverage samples represent the same class but only extremes within that class.

Continuing with the same analogy, as a sample moves along the S_i axis, these samples are not as well modeled as those closer to the origin. Therefore, the espresso changes from coffee to being more tealike in character. If the sample lies outside of the S_i axis, it is now characteristic of tea rather than coffee. A sample that exceeds both boundaries simultaneously is not a member of that class and in terms of the analogy is now a milky tea rather than an espresso.

Another variant of the S_i versus H_i plot is the S_i/S_0 versus H_i plot. The main difference is that the value S_i is now scaled to the average distance to the model S_0 . This plot is interpreted in the same way as the S_i versus H_i plot.

Overall, the S_i versus H_i plot is a one-stop diagnostic for determining the ability of each model to uniquely classify samples. The diagnostics presented in the next section are used to support the findings made in the classification table, the Coomans plot, and the S_i versus H_i plot.

Model Distance Model distance is usually plotted as a bar chart and shows the relative distances of all models in the library to each other with the proviso that a model’s distance to itself is 1. In general, a large intermodal distance indicates clearly separated models. Model distance is calculated as the pooled residual standard deviations by fitting samples from two different classes to their own models and every other model in the library. Figure 65.40 provides an example of a model distance plot.

A general rule of thumb when interpreting a model distance plot is a model distance greater than three indicates models that are significantly different, although there are

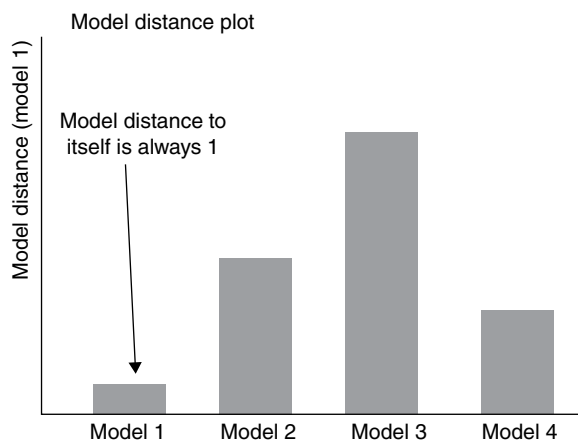


FIGURE 65.40 Example model distance plot.

exceptions to this rule. In most cases, when the model distance is less than three, there is no discrimination power of the models for the classes under investigation.

Variable Discrimination Power The discrimination power of a variable provides information about its ability to discriminate between any two-class models. It is calculated by fitting the samples from one model to all other models in the library and to its own class model.

As it is a pairwise model comparison tool, eliminating variables from one model pair results in removing these variables from all model classes, as based on the rules of SIMCA defined in Section 65.5.2.1.

Therefore, if two models show ambiguities and the removal of specific variable improves the results for the two classes but is detrimental to all other classes, this is where the hierarchical approach becomes of high value. The ideal situation is that by removing noncontributing variables from one model will result in an improvement in all models. Figure 65.41 provides an example of a variable discrimination power plot.

Modeling Power Modeling power quantifies the importance of a particular variable for modeling a particular class. It is a measure of the variables variance that is used to describe the class model. With any variable elimination step in SIMCA, it is important to assess the impact on other models in the library.

Variables with large modeling power have a large influence on the model. As a general rule of thumb, if the modeling power is less than 0.3, this variable could be detrimental to the model. This statistic should also be used in combination with the variable discrimination power to make sure that the discriminating power of the variable is also low before it is considered a candidate for elimination. Figure 65.42 provides an example of modeling power plot.

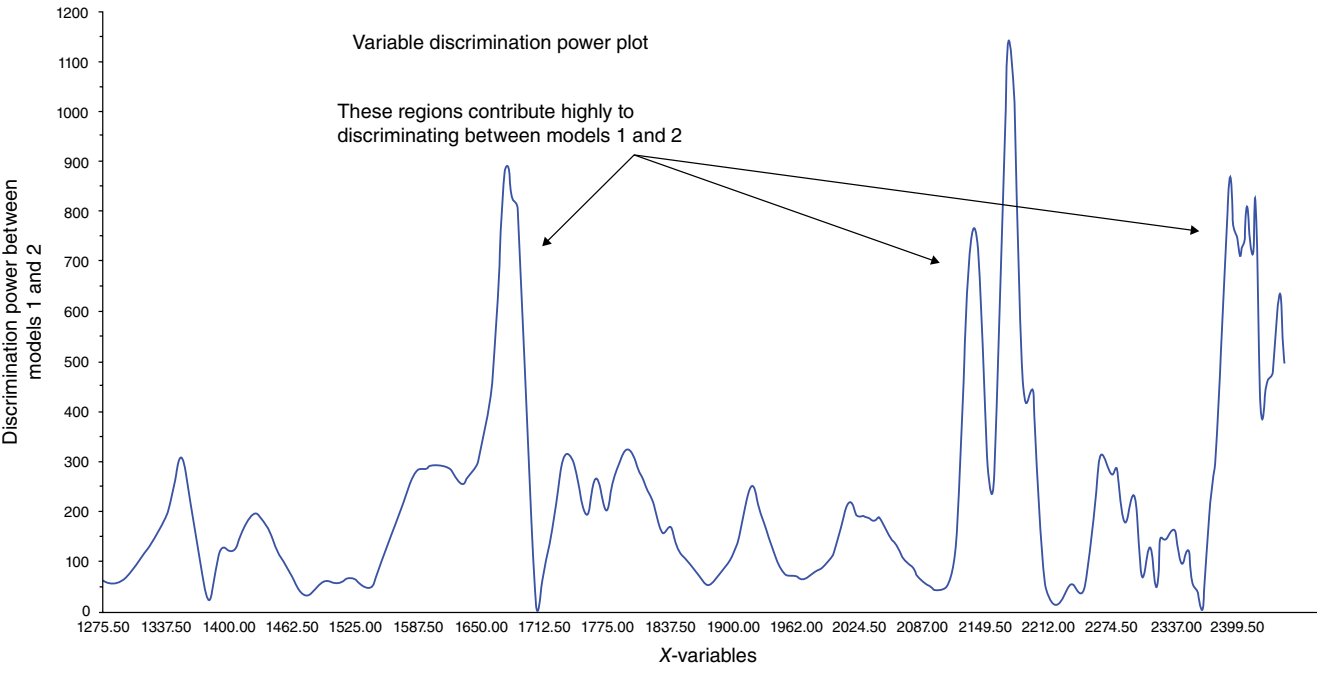


FIGURE 65.41 Example variable discrimination plot.

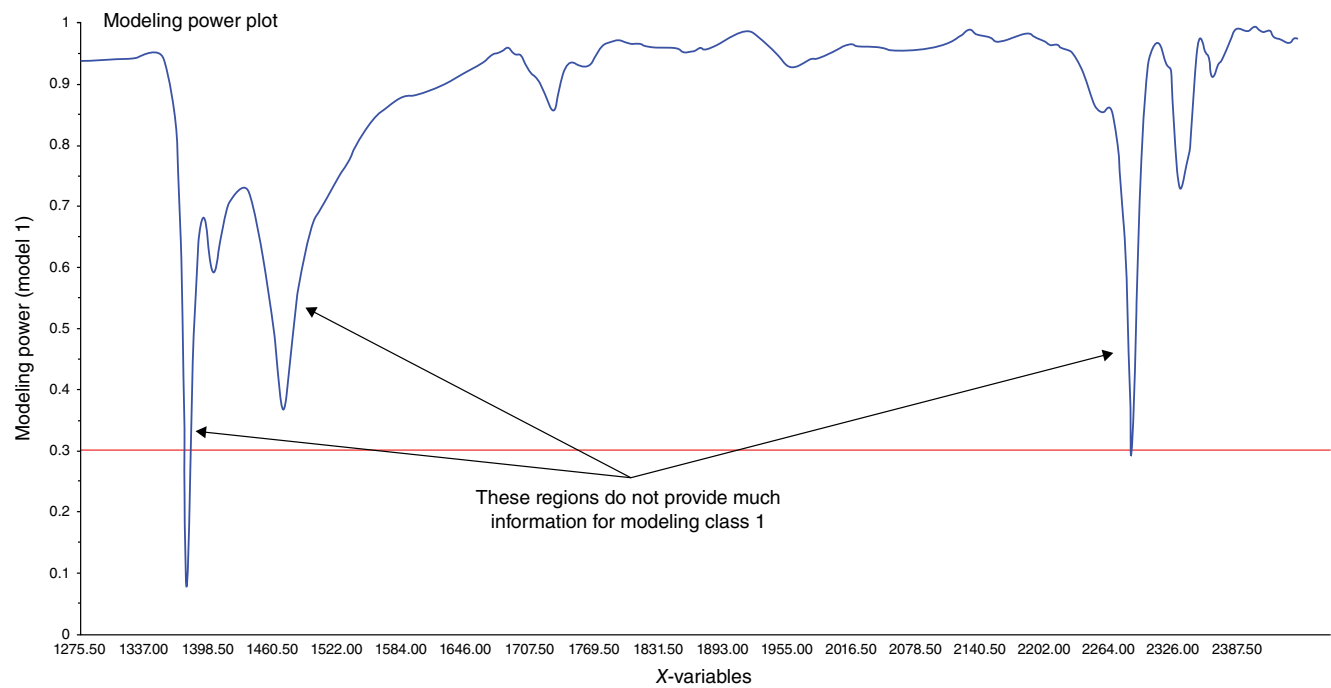


FIGURE 65.42 Example modeling power plot.

65.5.2.3 Summary of SIMCA The discussion presented on SIMCA has shown it to be one of the most versatile and powerful multivariate classification models available. It can not only utilize PCA models, but PLSR models can also be utilized for their scores structure.

The method is based on disjoint modeling, meaning that SIMCA uses individual models to form a library where each individual model defines a classification rule and the SIMCA architecture defines how to apply the library models to each sample. Because of the highly graphical and diagnostic rich nature of PCA and PLSR models, there are a multitude of diagnostic tools available for training and applying SIMCA models to new samples. SIMCA also has the major advantages that it can detect ambiguities and nonclassification of samples in the model.

Some drawbacks to the method include:

1. All PCA/PLSR class models must have the same number of variables and the same preprocessing applied to them.
2. If an unimportant variable is removed from one model, it has to be removed from all library models. This could be beneficial to one model but detrimental to the classification ability of other models in the library.

These disadvantages are only minor though compared to classical LDA, and hierarchical modeling approaches can be used to avoid these issues. As the basis for the classification is individual PCA models, there is no direct objective to separate the classes. Thus, the assumption is that the main components from PCA also are suited for classification. This is also related to the discrimination and modeling power as described previously.

65.5.3 Partial Least Squares Discriminant Analysis

PLS-DA [45] is a method that utilizes the PLSR algorithm for classifying samples into distinct classes by using a binary class separator. In the simplest case, a binary class variable where class *A* is designated 1 and class *B* is designated 0 is defined. The PLSR algorithm is applied to determine if a model can be generated that predicts close to 1 for class *A* and close to 0 for class *B*. This is shown diagrammatically in Figure 65.43.

When there are more than two classes to be separated, the form of the PLSR model becomes the complete PLS-2 algorithm and the *Y*-variable structure must also be changed. This is shown in Figure 65.44.

The structure of the *Y*-variable is now arranged by adding a new column for each class. For the class column, all samples of that class are given a 1 designation while all other classes are given a 0 designation. The PLS-2 algorithm then makes a model for each class setting the class to be modeled as 1 and all other class predictions to 0. In this way, each PLS model is similar to a PCA library in SIMCA, only it is held together in the PLS-2 architecture.

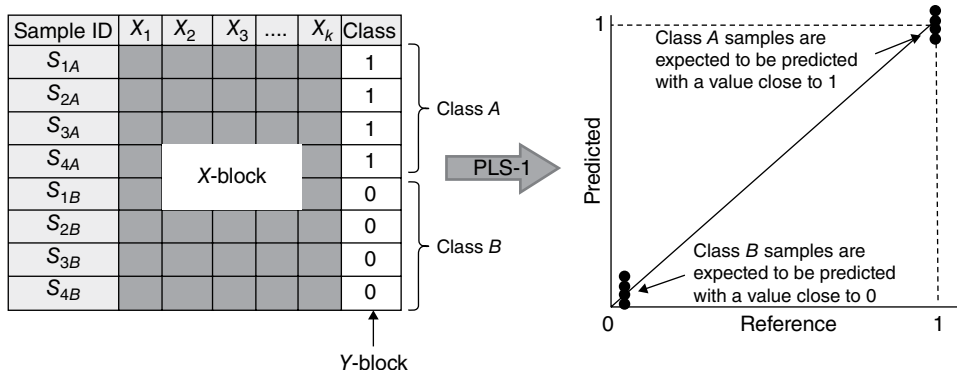


FIGURE 65.43 PLS-DA for the two-class discrimination problem.

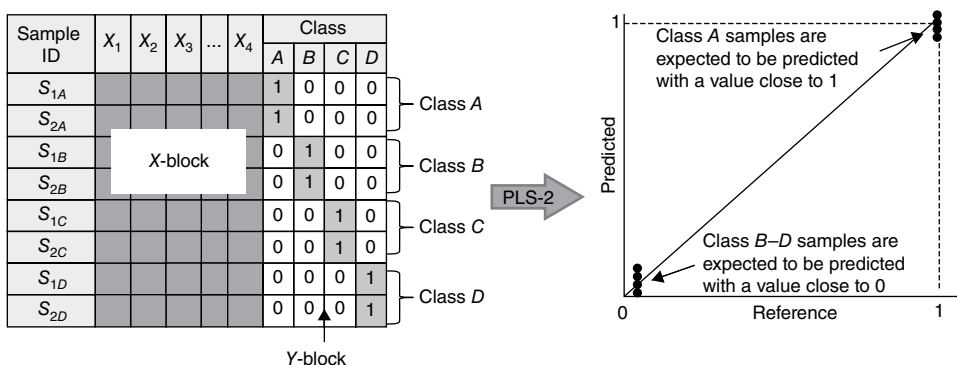


FIGURE 65.44 PLS-DA for the three (and higher) class discrimination problem.

The main advantage of PLS-DA over SIMCA is that the underlying factors as represented by the scores and loading weights/loadings are guided by the Y -variables. This can add more specificity to the training stage. It also allows the full usage of the diagnostic tools available in PLSR. Overall, there are similar advantages and disadvantages as to whether to use PLS-DA over SIMCA; however, like all classification methods, they should be viewed as complementary and again, in a hierarchical modeling sense, the combination of both approaches to solve specific ambiguity situations should be considered.

When interpreting the results of a PLS-DA, the predicted versus reference plot (Section 65.4.4.6) is one of the primary tools for assessing model quality. Based on the principles of least squares modeling, the distribution of results around the regression line is assumed to be normal. This means that for accurate estimations of class membership, a suitable number of samples should be used in the training model to provide an estimate of precision around the predicted values 0 and 1. The predicted results should ideally be distributed as a t -distribution depending on the number of samples used to train each class. From there, statistical limits can be put around 0 and 1 so as to

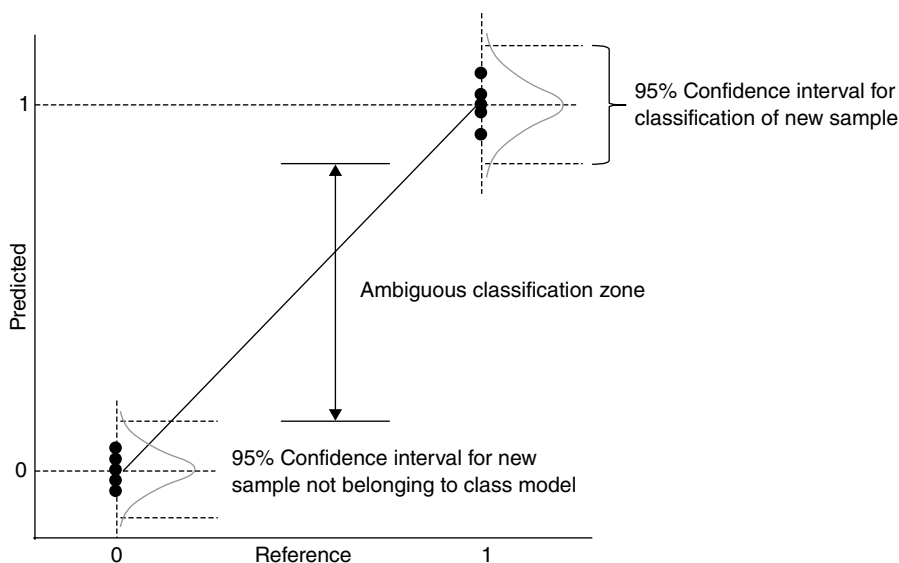


FIGURE 65.45 Predicted versus reference plot for PLS-DA with class membership limits.

determine whether a sample is part of the class or not. This is shown diagrammatically in Figure 65.45.

Overall, PLS-DA is a useful tool for solving multivariate classification problems. It provides the powerful graphical and diagnostic tools associated with the PLSR method and can be used for classification problems where more specificity is required for relating the X -variables to the Y -variables.

65.5.4 Support Vector Machine Classification

SVMC is a pattern recognition method that is used widely in data mining applications and provides a means of supervised classification, as do SIMCA and other linear discriminators. SVM was originally developed for the linear classification of separable data but is applicable to nonlinear data with the use of kernel functions. SVMs are used in machine learning, optimization, statistics, bioinformatics, and other fields that use pattern recognition. In this section a minimum of mathematics will be used to define the SVMC class separation problem. The mathematical principles behind SVM are outside of the scope of this book, and the interested reader is referred to the literature for more information [46].

65.5.4.1 The Idea behind SVMC SVM is a classification method based on statistical learning that fits a hyperplane to a multidimensional data set for optimal separation of classes. As linear functions are not always able to model complex separation problems, in SVM, data are mapped into a new feature space and a dual

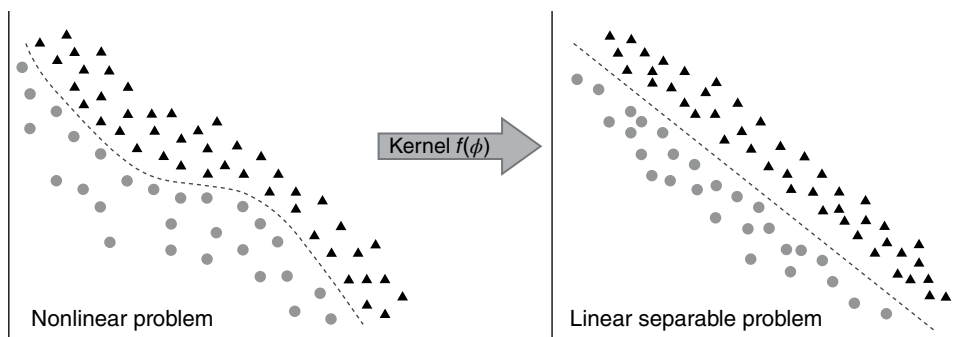


FIGURE 65.46 Using a kernel to map a high-dimensional space to a simpler feature space.

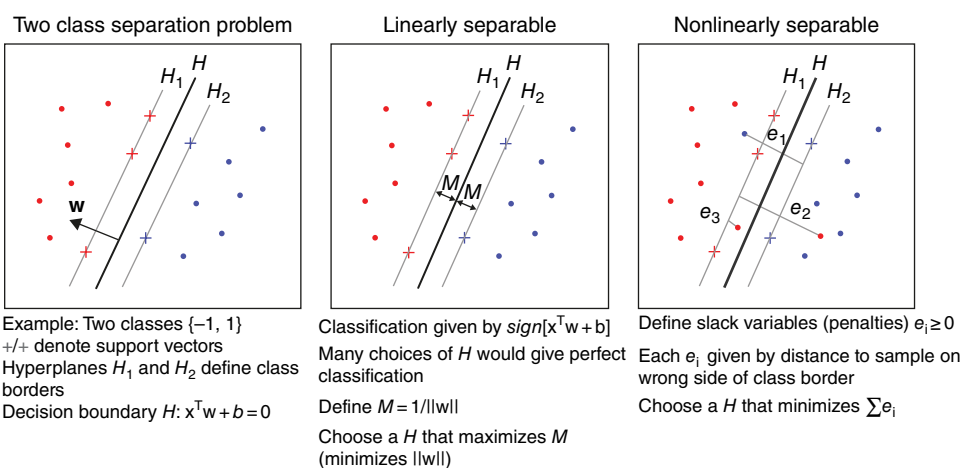


FIGURE 65.47 Samples that define the support vectors.

representation is used with the samples represented by their dot product. A kernel function is used to map from the original space to the feature space and can be of many forms, thus providing the ability to handle nonlinear classification cases. The kernels can be viewed as a mapping of nonlinear data to a higher-dimensional feature space while providing a computation shortcut by allowing linear algorithms to work with higher-dimensional feature space. The support vector is defined as the reduced training data from the kernel. Figure 65.46 illustrates the principle of applying a kernel function to achieve class separability.

In this new space SVM will search for the samples that lie on the borderline between the classes, that is, to find the samples that are ideal for separating the classes; these samples are named support vectors. Figure 65.47 shows this principle where samples marked with $+$ for the two classes are used to generate the rule for classifying new samples.

A situation where SVM performs well is when some classes are inhomogeneous and partly overlapping, that is, where classical methods, such as SIMCA, will result in ambiguities and thus not be effective as a classification rule. SVM will in this case find

a set of the most relevant samples in terms of discriminating between the classes and is invariant to samples far from the discrimination line.

SVM has advantages over classification methods such as neural networks, as its outputs are more transparent, and has less tendency of overfitting when compared to other nonlinear classification methodologies. For finding an optimal model, a grid search is often applied to investigate various combinations of the parameters governed by the type of SVM and kernel. In all cases, model validation is the critical aspect in avoiding overfitting for any method. Not only are SVMs effective for the modeling of nonlinear data, but they are also relatively insensitive to variation in these model parameters. SVM uses an iterative training algorithm to achieve separation of different classes. In the case of similar classification rate for the training data, the model with the most linear parameters and minimum number of support vectors should be chosen.

There are two general SVM classification types that are based on different means of minimizing the error function of the classification:

1. c-SVC: also known as classification SVM type 1
2. nu-SVC: also known as classification SVM type 2

In the c-SVM classification, a capacity factor, C , can be defined. The value of C should be chosen based on knowledge of the noise in the data being modeled. Its value can be optimized through cross validation procedures. When using nu-SVM classification, the nu value must be defined. Nu serves as the upper bound of the fraction of errors and is the lower bound for the fraction of support vectors. Increasing nu will allow more errors while increasing the margin of class separation.

In practice, there are four main kernel types that can be used to separate classes; these are:

1. Linear
2. Polynomial
3. Radial basis function
4. Sigmoid

The detailed explanation of kernel functions and feature spaces is outside of the scope of this chapter, and the interested reader is referred to the article by Luts et al. [47] for an excellent discussion of the various kernel types and their application in chemistry.

65.6 TECHNIQUES FOR VALIDATING CHEMOMETRIC MODELS

Probably the single most important issue of chemometric model development is model validation [48]. The objective of validation is to evaluate the performance of a multivariate model, be this related to modeling and interpretation, discrimination, or

prediction. Whether exploratory or regression methods are being developed, validation is concerned about prediction performance.

As validation has been the topic in many publications, this section is more of a discussion of the principle rather than presenting specific applications.

A distinction between data-driven (internal) and hypothesis-driven (external) validation may be drawn. The latter is focused on confirming the known structure in a system under observation such as to find the true signals of, for example, chemical compounds. Another aspect to consider, when building empirical models based on multi-channel spectroscopic data, is to observed whether the model highlights any known chemical groups based on a-priori assumptions of chemical absorbances in specific regions of the spectrum. This may also be confirmed using existing chemical group information in the literature. This is again related to interpretation, and a good rule is “no prediction without interpretation, no interpretation without evaluating prediction ability.” The following will focus on data-driven validation.

Validation of multivariate (or any model in general) is essential in order to make sure that the model will work in the future for new, similar data sets, and indeed do this in either a qualitative or a quantitative way. In regression models, this can be viewed as prediction error estimation (section “Error Measures”), while in classification models, this can be viewed as misclassification rates. Validation is often also used in order to find the optimal dimensionality of a multivariate model, that is, to avoid either overfitting or underfitting or incorrect interpretation [49].

65.6.1 Test Set Validation

When the objective is to establish a calibration model for predicting quantities such as concentration or determining classification models selectivity and specificity, the most conservative validation is to test the model on a representative, independent test set of sufficient size. This has been discussed in length by Esbensen and Geladi [50].

It may be debated what is meant by a test set given a specific situation. Question such as, can the set be used to test extrapolation of the calibration set, or will changes in sample matrix, with respect to the calibration set, invalidate the model? These sources of variation that are, in principle, unknown for future samples, may be quantified by several approaches.

When a test set is used, the model error will be expressed as the RMSEP (section “Error Measures”). This is the most reliable measure of the models future performance on new samples, provided the sample is similar to the calibration set samples. The diagnostic tools discussed in Section 65.4.4.7 can also be used during prediction to assess the quality of the *X*-data such that the predicted value(s) can be assured of being reliable.

In order to create a representative test set for model evaluation, there are a number of general approaches and two of these approaches will be discussed as follows.

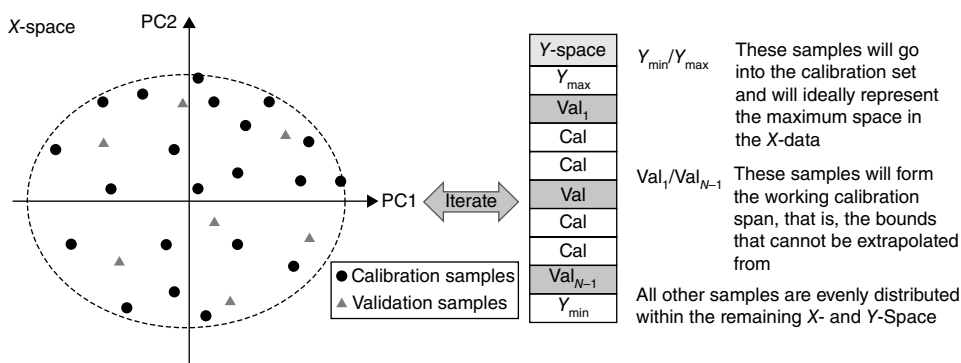


FIGURE 65.48 The process of maximum space sample selection.

65.6.1.1 Maximum Space Sample Selection Maximum space sample selection is a manual process that requires the visualization of the scores space of the entire sample pool and selects samples at the extremes of scores space for all interpretable components/factors in the model. These samples will form the first part of the calibration set. An even set of well-distributed samples can then be selected from scores space to define the most even coverage of X-space. The next step in the process is to order the samples based on their Y -values and first select the highest and lowest value to go into the calibration set. Ideally these samples will match those already selected in X-space.

The calibration to validation sample ratio is usually 2 : 1; therefore, an even 2 : 1 split of samples based on Y -values will yield an even distribution of samples (provided the underlying sample population is not normal or skewed). It is stated here that a boxcar distribution of samples is to be preferred when developing calibration models. Simple row exchange based on the most even distribution of X- and Y-space is performed such that the entire validation set span in X- and Y-space is encapsulated by the calibration sample space. This process is shown in Figure 65.48.

65.6.1.2 The Double Kennard–Stone Method The double Kennard–Stone method [51] starts by first finding the two most different samples, often applied on the scores from PCA rather than the original data. Thereafter the next sample is found from the largest distance to these, and this continues until a given number of samples is selected, eventually until the distance is lower than a preset limit. The Euclidean or Mahalanobis distance is normally used. This ensures an even coverage of the samples in the multivariate space. An even better approach is to for every second sample selected put this in a validation set, the so-called double Kennard–Stone or duplex.

Though the objective is to have enough samples in the available sample pool to put a reasonable number aside as a test set, this is not always possible due, for example, to the cost of samples or reference testing. The best alternative to an independent test set for validation is to apply cross validation when a suitable number of samples is not available.

65.6.2 Cross Validation

With cross validation [52], the same samples are used both for model development and testing. A few samples are left out from the calibration data set (based on some predefined criteria) and the model is calibrated on the remaining objects. Then the values for the left-out objects are predicted and the prediction residuals are computed. The process is repeated with another subset of the calibration set, using a systematic way until every object has been left out once; then all residuals are combined to compute the validation residual variance and root mean square error of cross validation (RMSECV) in prediction or classification rates for classification models. It is of utmost importance that the user is aware of which level of cross validation to use. For example, if one physical sample is measured three times, and the objective is to establish a model across samples, the three replicates must be held out in the same cross validation segment. If the objective is to validate the repeated measurement, keep out one replicate for all samples and generate three cross validation segments. The calibration variance is always the same; it is the validation variance that is the important figure of merit.

Kos et al. [53] make a general comment that for sample sets greater than 50 test set validation is preferred whereas cross validation is best for small to medium data sets. Given a specific stratification of the samples in a data set, the level of validation in cross validation should reflect the objective, for example, is the model to be used for other batches of raw materials?

Some general rules and practical considerations for applying cross validation are provided in the following, given the level of validation:

1. Full cross validation, also known as leave-one-out (LOO), leaves out only one sample at a time. If the number of objects is less than 20, this may be a viable option.
2. Segmented cross validation. There are theoretical and practical results indicating that, for example, 10 random segments give a good estimate of the prediction error. Or in general terms, if the model changes considerably when 10% of the samples are taken out, it means the model is not stable. However, it is only in the case where there is no stratification of the samples based on the underlying sampling strategy or the origin of the samples that a random segment CV is justified.
3. Systematic segmented cross validation leaves out a whole group of samples at a time. A typical example is when there are replicated measurements of one physical sample. Depending on the objective either take out all replicates for each physical sample or replicate n for all samples.
4. Validating across categorical information about the samples. This enables the analyst to validate across the model and evaluate the robustness across season, raw material supplier, location, operator, etc.

The main purpose of establishing a model may not in itself be for predicting or classifying new objects but to understand the inherent structure in the system under

observation. In chemometrics this relates to latent variables that may convey the basic chemical, physical, or biological phenomena. The interpretation of such models is highly dependent on the number of latent variables, and therefore it is vital to assess the correct dimensionality of the model, that is, in more mathematical terms the model rank. It is important to distinguish between numerical rank, statistical rank, and the application-specific rank. Note that even though a representative test set is present, it is nevertheless important to find the correct model rank in the calibration model for predicting the test set.

Both test set validation and cross validation can be applied to any regression model made by either PCA, MLR, PCR, PLS, or other methods. These validation methods are equally applicable to augmented regression models like nonlinear regression, including support vector machine (SVM) models and neural networks, for example, and are perhaps even more important for methods that involve estimates of many parameters as these imply even greater risks of overfitting.

65.7 AN INTRODUCTION TO MSPC

In traditional SPC applications, variables are measured one at a time on control charts where each variable is assumed to be independent of each other. In most process applications, this is not always the case and in many situations, even though individual control charts show the variables to be in control, the overall process is out of control. This is because, as discussed in great detail in this chapter, of the collinear nature of data. Refer to Figure 65.2, case 2, for an excellent example.

Methods such as PCA and PLS are mature and have been used in many applications for quality control using instruments such as spectrometers. Recently, they have gained more interest for modeling the data generated by manufacturing process, and therefore, they form the basis of MSPC methods. Before a detailed discussion of MSPC tools is provided, a short section on multivariate projection will be provided.

65.7.1 Multivariate Projection

The method of PCA can be used effectively as an MSPC tool for monitoring multiple variables simultaneously. In order to be useful as a monitoring tool, it uses the method of projection to achieve this objective. Consider the validated PCA model, where only the informative part of X is retained in terms of scores and loadings:

$$X = TP'$$

This model can be rearranged in terms of the scores T to yield

$$T = XP$$

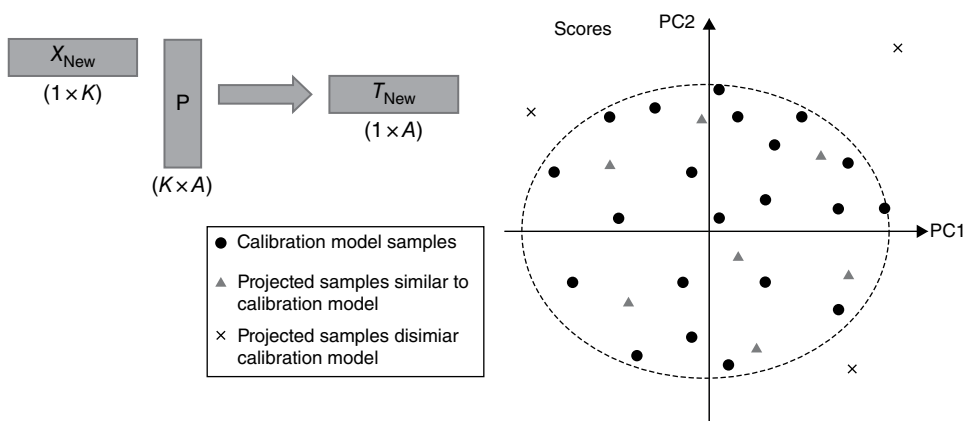


FIGURE 65.49 Some common situations in PCA projection.

This means that, given a set of PCA loadings, new samples X can be projected onto these loadings to provide new scores (\hat{T}_{New}). Given a model with established Hotelling's T^2 limits and Q -/ F -residual limits, the projected sample can be assessed as either belonging to the population the model was developed from or not. This is expressed mathematically as

$$\hat{T}_{\text{New}} = X_{\text{New}}P$$

Some common situations arising from projection are provided in Figure 65.49.

65.7.2 Hotelling's T^2 Control Chart

Monitoring PCA/PLS scores in control charts or as scatter plots is effective when the number of PCs in a model is at maximum 2. Beyond 2, 3D charts, or multiple 2D charts may be required to visualize the entire process. As was discussed in section "Hotelling's T^2 ," Hotelling's T^2 is closely related to leverage and is a measure of the distance of a sample from the center of a model. The center of a model in PCA is analogous to the center line of a control chart; therefore, the further away the sample is from the model center, the more likely it represents an out-of-control situation.

The Hotelling's T^2 chart provides a convenient summary of all variables simultaneously no matter how many PCs/factors are in the model. The minimum value of Hotelling's T^2 is zero; this means that when using this chart in a process monitoring application, only an upper limit is required. Figure 65.50 provides an example of a Hotelling's T^2 chart.

In the event where an outlier is detected in the Hotelling's T^2 chart, in most software applications, a user is able to drill down to the variables contributing to why the sample is an outlier. This plot is called a contribution plot [54] and is a weighted

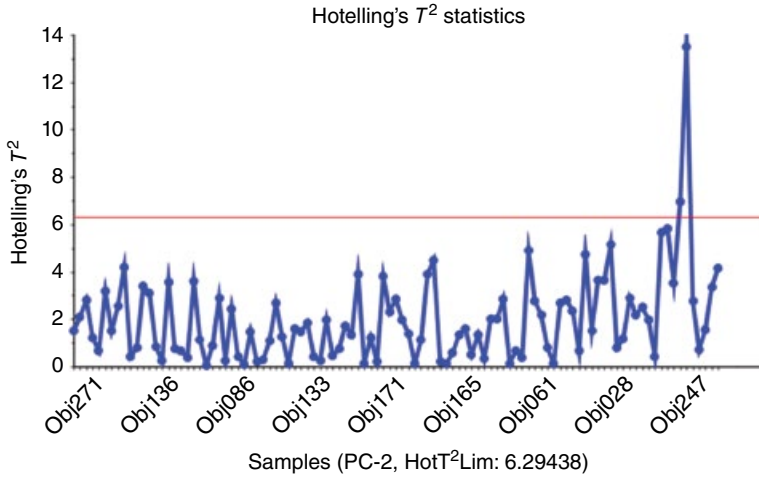


FIGURE 65.50 Example Hotelling's T^2 chart.

loadings plot specific to the sample. The calculation of contribution is provided in the following:

$$c_{nk} = \sum_{a=1}^A S_{aa}^{-1} t_{\text{new},a} x_{\text{new},nk} p_{ak}$$

where S_{aa}^{-1} is a square matrix with the inverse of the eigenvalues on the diagonal.

The contribution plot shows which variables contribute most to the sample being an outlier, with respect to the loadings of the model. The variables that are most weighted can be used to provide feedback to a control system for making control decisions. This is known as advanced process control (APC) [55].

65.7.3 Q -Residuals

Like the Hotelling's T^2 plot, the Q -residuals can be plotted as a line plot to detect those samples that look dissimilar to those used to construct the model. Q -residuals also only have an upper limit with the minimum residual being zero (since Q -residuals are based on a sum of squares of the residuals). Contribution plots can also be generated from a Q -residuals plot to better understand variable contributions to outlying samples.

65.7.4 Influence Plot

The influence plot is fast becoming the standard MSPC plot because it captures all of the diagnostics information for each sample in one plot. The MSPC influence plot is typically constructed as Q -residuals versus Hotelling's T^2 plot with statistical limits typically set at 95%. The space bounded by the Q -residuals and Hotelling's T^2 limits for a validated number of PCs/factors represents the situation where all variables are

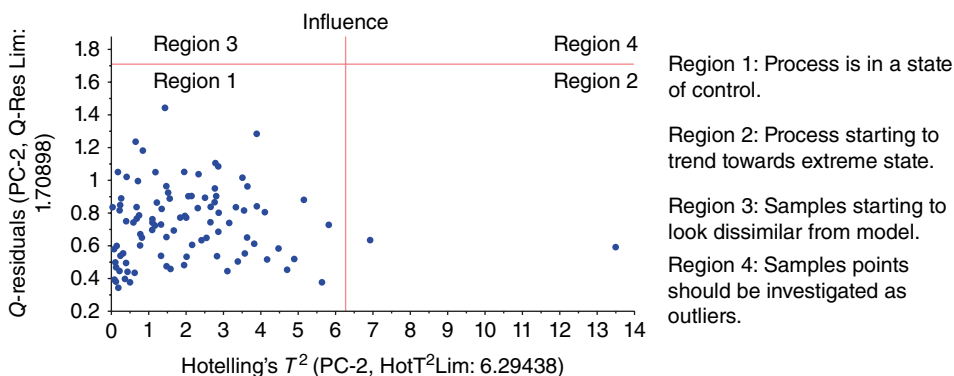


FIGURE 65.51 Example influence plot used for MSPC.

simultaneously in control. Samples that exceed a Q -residual or Hotelling's T^2 boundary only are interpreted as per the charts acting independently. When a sample exceeds both boundaries simultaneously, this situation is indicative of a true outlier and such as situation will require a root cause investigation to resolve the problem. Figure 65.51 provides an example of MSPC influence plot with all of the regions marked on the plot.

65.7.5 Continuous versus Batch Monitoring

In a typical manufacturing environment, there are two main types of process:

1. Continuous manufacturing: This is characterized by processes that have reached a steady state and process models are concerned with detecting out-of-control situations typically based on static models.
2. Batch manufacturing: Typically characterized by processes exhibiting some form of process signature (or trajectory). These processes require a monitoring strategy that assesses whether the process data remain in an envelope defined by the process signature.

65.7.5.1 Continuous Processes Common examples of continuous processes include water quality monitoring, gasoline blending, and many other examples where the process is maintained in a steady state. The variables measured are looking to keep the process in a target range that is static over the entire time the process is being run.

Consequently, the MSPC models used to monitor the process are simpler than those used in batch processes. Section 65.7.6 defines the requirements for variable measurement and alignment systems required for implementing successful process monitoring strategies.

65.7.5.2 Batch Modeling Only a brief discussion on batch modeling will be provided here as a full discussion is outside of the scope of this chapter. In a batch

process, raw materials are combined in a suitable batch vessel before chemical, physical, or biological transformation takes place, resulting in an end product. In many cases the control of the batch process is recipe driven and the operations are not adjusted to accommodate raw material variation, changes in uncontrollable factors, and other changing circumstances. The best possible end product quality is achieved by adapting batch operations according to any detectable changes during processing, thus providing a control mechanism to drive a product toward what is known as its desired state. Optimal run settings and the ability to control them within a design space [56] leads to reduced rework and rejects, and improved end product quality, which has the major benefit of saving industry money and resources and, more importantly, increased consumer trust in the product name.

There are a number of batch modeling approaches and the most common assume equal lengths of batches, that is, the batch is expected to start at the same chemical or biological time t_0 and has the same number of time points for all batches. This leads to problems during model building if the data set has uneven rows and ultimately during monitoring if new batches do not meet these criteria. Numerous approaches to handle uneven batch lengths exist, including replacing time with a maturity index [57], dynamic time warping (DTW) [58], time linear expanding/compressing [44], etc. Complications can occur in all of these methods if the first measurement does not coincide with the true t_0 , that is, the new batches do not start at the same chemical/biological state. The PARAFAC [59] approach models the data as a true three-way model, which has a possible advantage that the time is modeled as a separate dimension and not connected to other samples or variables as in the unfolding case. However, the challenges with unequal batch length and chemical time still need to be addressed. Also, the monitoring phase requires dynamic recalculating of models up to the current point of time [60].

An improved batch modeling approach accommodating uneven batch lengths, unknown true t_0 , phase changes, and uneven residence times has been proposed by Westad et al. [61]. This is achieved by a true multivariate, feature-based approach that does not make any assumptions about the synchronization and duration of batches. Instead the so-called relative time is estimated by the method itself. Relative time is here used in a broad sense for any transient process including nonlinear behavior, and it is often found to correspond with the underlying chemical, biological, or physical changes during the process.

Figure 65.52 provides an example of how a batch modeling MSPC approach looks like and shows how the process trajectory, representing the process signature, can be displayed with an envelope around the trajectory.

The main difference between continuous and batch modeling strategies is that in continuous, the process is considered to have reached steady state; therefore, univariate control charts can be used to monitor individual variables and MSPC charts will be used to ensure that the process remains within the limits of the static model developed for the process. In batch modeling, the limits are dynamic and thus require a different

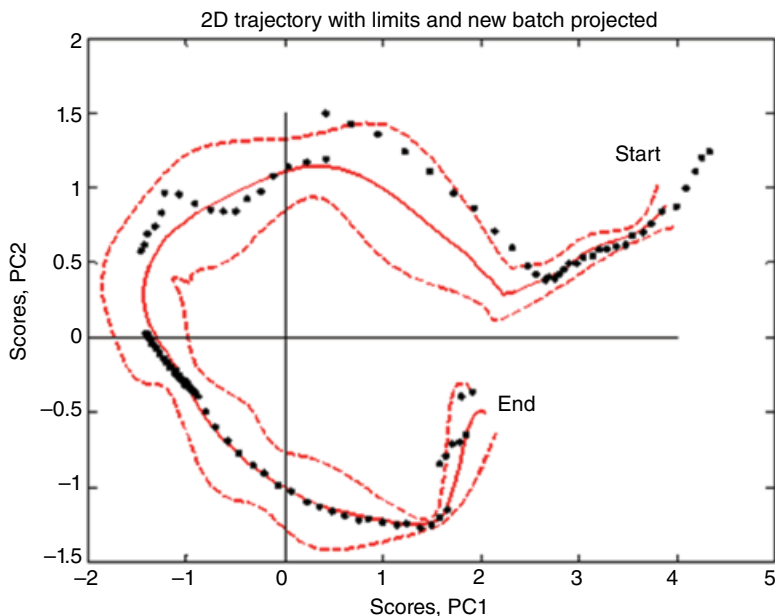


FIGURE 65.52 Example batch process MSPC display.

approach to modeling and monitoring. For more details on batch modeling, the interested reader is referred to the literature cited [55–61].

65.7.6 Implementing MSPC in Practice

The implementation of MSPC requires an intimate knowledge of the process under investigation and a strong data management system. Consider the generic process and data management system shown in Figure 65.53. There are a number of variables being measured and typically these are being measured at different frequencies. This is where subject matter expertise is essential. The time interval for monitoring the process must be set in order to reliably detect a critical change in the process.

If a temperature sensor measures a reading every 100 ms and a critical change in a process is detectable in a time frame of tens of minutes, there is no logic to monitoring the process at such a high sampling rate. Conversely, if a process is rapid and critical changes are picked up in seconds, this would justify a much higher sampling rate.

In today's world of rapid sensors, there is an abundance of information that can be obtained. IBM is quoted as saying that 90% of the world's data has been captured in the past 2 years [62]. With big data, business intelligence, and manufacturing intelligence becoming buzz words in all industries, there will be a much greater expectation on future systems to provide as much data as possible; however, the key question and challenge is, what to do with the data?

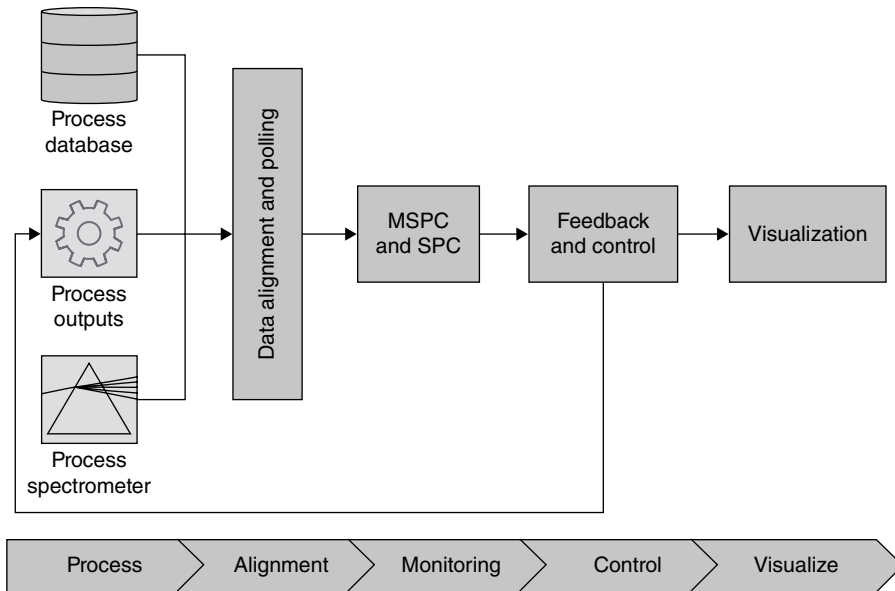


FIGURE 65.53 Generic process and data management system example.

The multivariate methods discussed in this chapter form the basis of what to do with the data, the fundamental challenge if the agglomeration of many data sources into a representative (meaningful) array that can be modeled and analyzed. The first step to a successful process monitoring systems is therefore a data management system capable of importing, in run time, many different formats of data and then compiling them into a representative array. Until the data is time aligned and searchable, the modeling phase becomes a manually driven, painstaking exercise. Figure 65.54 provides a simple situation where four sensors are feeding data at various rates from a process into the data management system.

As the data are being fed at their own natural frequencies (ν_i), a filtering and agglomeration system can be implemented primarily to remove mechanical shock data points from the individual data and to compile the data until they are ready to be polled. Below the agglomeration system is a polling layer. The polling layer is the critical part of representative data collection. It is set, based on subject matter expertise, to poll the data currently sitting in the agglomerator. Whenever the system is polled, the average of the filtered data is used as a single value to build a fused data array that can either be used for multivariate modeling or can be passed onto a multivariate model for evaluation of process state. This data model works for both continuous and batch modeling strategies.

The model shown in Figure 65.54 can be expanded to include vector inputs, such as the data generated by spectrometers, chromatograms, etc. In this case, the whole vector, parts of it, or even the scores obtained by applying multivariate models to the

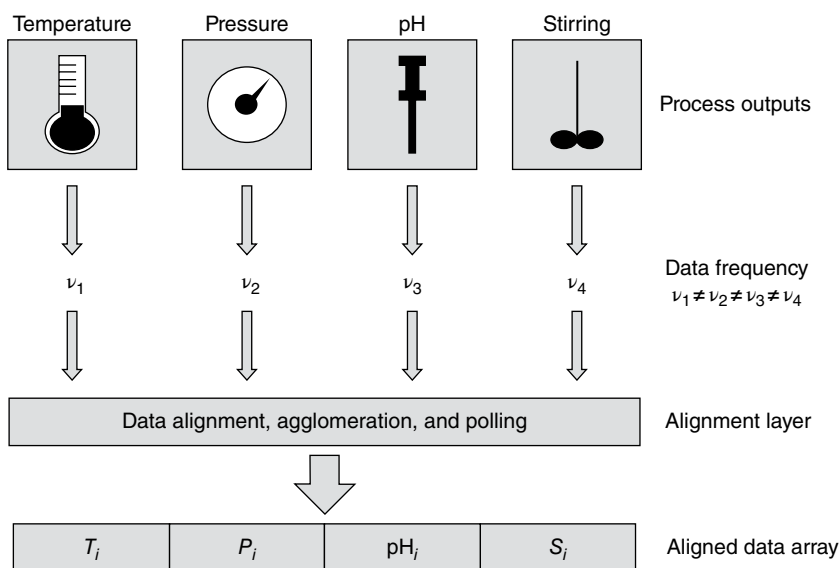


FIGURE 65.54 Process data being fed into a data management system.

data can be used to build a fused data array. The objective of many process control strategies is to achieve feed forward/feedback control loops. By using a modern data management system approach combined with data fusion from process sensors and MVA, the following goals can be achieved:

1. Both SPC and MSPC monitoring approaches can be implemented simultaneously.
2. The alarms generated from SPC/MSPC can be cross-referenced to each other to determine whether an issue is arising from one variable or a combination of variables (refer to Figure 65.2).
3. The polling rate of the process data can be set over multiple operations in order to allow the control system to adapt accordingly to the process.
4. Predictions performed or scores extracted from the fused data can be used as inputs to other parts of the process to provide feed forward/feedback capability to the system. This is particularly important in continuous manufacturing or processes where 100% inspection of units is required.
5. It allows innovation and learning such that the process can be optimized in a shorter time compared to traditional approaches and the predictive ability of the models can allow forward projection (using a time series approach) to warn a system that a process failure may occur.

It is the last point that makes the modern data management system approach most attractive to industry. The ability to predict an event before it occurs allows for a proactive approach, not a reactive approach, to quality, and therefore, this approach should

be considered by any manufacturer looking to improve quality, reduce costs through less scrap and better equipment management, use less energy, and improve brand recognition through better and possibly greener approaches.

65.8 TERMINOLOGY

It must be noted here that the usage of terminology has traditionally been a matter of preference; however, terminology standardization has been a subject of debate in the chemometric community in recent times. Wherever possible, this chapter uses only standardized terminology as used in the wider community and avoids the use of proprietary terms. To add more confusion to the matter, similar methods to chemometrics exist in other disciplines, such as engineering, and even though a chemist and an engineer are talking about the same thing, terminology is the killer of interpretation!

This section aims, as far as possible, to introduce the key terms most commonly used in chemometrics.

Accuracy The closeness of agreement between a predicted result and an accepted reference value.

Bias The arithmetic average difference between the reference values and the values produced by the analytical method under test for a set of samples.

Calibration The stage of data analysis where a model is fitted to the available data set and assessed for quality of fit by a proper validation method.

Category variable This is typically a noncontinuous class variable without any quantitative equivalent. Used to group samples into predefined categories and for cross validation to assess the stability of the model during chemometric modeling.

Classification A systematic approach used to sort a set of objects/samples into a set of distinguishable classes. Rules can then be put in place to direct the classification of new objects into these classes. The first step is called unsupervised classification, where classes are defined, and the second stage is supervised, where the rules are used to classify new samples.

Cluster analysis A group of mathematical methodologies used to find sample patterns in complex data sets with the intent of interpreting the groups based on prior subject knowledge.

Collinearity The linear relationship between variables. Two variables are collinear if the value of one variable can be computed from the other using a linear relation.

Continuous variable Any variable measured on an infinitely divisible scale.

Correlation A unitless measure of the amount of linear relationship between two variables.

Covariance A measure of the linear relationship between two variables of whose scale is dependent on the magnitude of the two variables being analyzed.

Cross validation A simulated version of test set validation that creates independent calibration and validation sets in a predefined manner. The selected validation sets are left out of the model calculations and are used to validate the submodel generated. The validation samples are then returned to the sample pool and new calibration and validation samples are selected and the entire process is repeated until all samples have been used for calibration and validation purposes.

Dependent variables Also known as *Y*-variables or responses are a set of variables collected on a representative data set as a reference set such that they can be modeled by a set of independent variables.

Explained variance The proportion of the total variance in a predefined data set that is accounted for by a model generated from that data.

Independent variables Also known as *X*-variables or predictors are a set of variables collected on a representative data set to be modeled and used to either gain insights into the variable and sample relationships or to be used to predict a dependent variable.

Influence A measure of how much impact a single object/sample (or a single variable) has on a developed model. The influence depends on the leverage and the residuals.

Latent variable A variable that is not directly observed but rather inferred (through a mathematical model) from other variables that are observed and directly measured.

Leverage A measure of how extreme an object/sample or a variable is with respect to the center of the model space. Leverage has a one-to-one relationship to the Hotelling's statistic.

Linear discriminant analysis (LDA) LDA is the simplest of all possible classification methods that are based on Bayes' formula. The objective of LDA is to determine the best fit parameters for classification of samples by a developed model.

Loading weights These are generated in PLSR models and show how much each *X*-variable (predictor) contributes to explaining the *Y*-variable's (responses) variation for each model factor.

Loadings Loadings are a concentration of the main information carried by variables onto a few components. Each variable has a loading along each model component and its magnitude is a representation of how important that variable is. The corresponding correlation loadings simplify the interpretation as it is independent of the number of variables.

Model A mathematical equation summarizing variations in a data set.

Multiple linear regression (MLR) A method for relating the variations in a response variable (*Y*-variable) to the variations of several predictors (*X*-variables). An important assumption for the method is that the *X*-variables are linearly independent, that is, that no linear relationship exists between the *X*-variables. When the *X*-variables carry common information, problems can arise.

Multivariate analysis (MVA) MVA is based on a set of methodologies used for analyzing more than one variable at a time. It encompasses but is not limited to data mining and predictive analytic applications.

Multivariate statistical process control (MSPC) A complementary method to traditional statistical process control (SPC) that overcomes the one-variable-at-a-time issues by utilizing multivariate methods that provides information on not only the main variables but also their interactions.

Orthogonal Two variables are said to be orthogonal if the angle between them is 90° (i.e., at right angles to each other).

Outlier An object/sample or variable that shows abnormal characteristics with respect to the rest of the data set analyzed.

Partial least squares regression (PLSR) A method for relating the variations in one or several response variables (Y -variables) to the variations of several predictors (X -variables). This method performs particularly well when the various X -variables express common information, that is, when there is a large amount of correlation between the variables.

PLS factor The equivalent of a PC in PLSR only the information in the factor is most related to the response.

Precision A measure of the closeness of repeat predictions made on either the same sample (repeatability) or true replicates (reproducibility).

Prediction Estimating response values from predictor values, using a regression model.

Principal component (PC) A PC is a condensation of sample and variable information that describes a particular source of variability in complex data sets. Each PC describes a certain proportion of the total variability of a system, and the first PC (PC1) describes the greatest source of variability each successive PC describing less information than the previous one. Each PC describes an independent source of variability in the data, and therefore, they can be plotted as X - Y scatter plots for increased interpretability. In mathematical terms the direction of the loading vector is the *eigenvector*.

Principal component analysis (PCA) An exploratory data analysis method used to understand complex sample and variable relationships in multivariate data. The aim of PCA is to isolate those variables that most contribute to sample patterns observed. PCA models can be further used for developing classification models or predicting the state of new samples with respect to the original model.

Principal component regression (PCR) PCR is a method for relating the variations in a response variable (Y -variable) to the variations of several predictors (X -variables). This method performs particularly well when the various X -variables express common information, that is, when there is a large amount of correlation. PCR is a two-step method. First, a principal component analysis is carried out on the X -variables. The principal components are then used as predictors in a multiple linear regression.

Projection In multivariate methods such as PCA and PLSR, each object/sample can be considered as a single point in multivariate space. When a sample is projected onto a model, this results in a score, while variable projections are called loadings.

Regression Generic name for all methods relating the variations in one or several response variables (Y -variables) to the variations of several predictors (X -variables). Regression can be used to describe and interpret the relationship between the X -variables and the Y -variables and to predict the Y -values of new samples from the values of the X -variables.

Regression coefficient In a regression model equation, regression coefficients are the numerical coefficients that express the link between variation in the predictors and variation in the response.

Repeated measurement Measurements performed several times on a single sample in a short time period. Repeated measures are used for the estimation of measurement error.

Replicate Measurements carried out several times of different preparations of the same sample. Replicate measures are used for the estimation of experimental error.

Residual A measure of the variation that is not taken into account by a model. The residual for a given sample or a given variable is computed as the difference between observed value and fitted (or predicted) value of the sample.

Residual variance The mean square of all residuals sample- or variable-wise. The complement of residual variance is explained variance.

Root mean square error of calibration (RMSEC) A measurement of the average difference between predicted and reference values for calibration samples only.

Root mean square error of cross validation (RMSECV) A measurement of the average difference between predicted and reference values for validation segments when cross validation is the method used for validating the model.

Root mean square error of prediction (RMSEP) A measurement of the average difference between predicted and reference values for validation samples when the validation method used is test set validation.

Sample Object or unit on which variables are measured and which builds up a row in a data table.

Scores Scores carry information on several variables and are concentrated onto a few underlying variables. Each sample has a score along each model component. The scores show the locations of the samples along each model component and can be used to detect sample patterns, groupings, similarities, or differences.

Standard error of calibration (SEC) Variation in the precision of calibration sample predictions over several samples. SEC is computed as the standard deviation of the prediction residuals and is dependent of the validation method used.

Standard error of cross validation (SECV) Variation in the precision of predictions over several samples. SECV is computed as the standard deviation of the prediction residuals when cross validation is the validation method used.

Standard error of prediction (SEP) Variation in the precision of predictions over several samples. SEP is computed as the standard deviation of the prediction residuals when test set validation is the validation method used.

Support vector The data points that lie closest to the decision surface and are typically the most difficult to classify.

Support vector machine (SVM) SVM is a classification (or regression) method formally defined by a separating hyperplane. In other words, given a known training set of data the algorithm outputs an optimal hyperplane that categorizes or predicts new examples.

Test samples A representative set of samples that are independent of the calibration set used to validate the model using the method of test set validation.

Test set validation A model validation method using a separate test set of samples, providing the most reliable estimate of model performance on future samples.

Validation Validation means checking a model's fitness for purpose. In regression, validation allows for estimation of the quality of future predictions. The validation variance can be used as a way to determine how well a single variable is taken into account in an analysis. A variable with a high explained validation variance is reliably modeled, whereas a variable with a low explained validation variance is not explained very well by the model.

Variable Any measured or controlled parameter that has varying values over a given set of samples.

Variance A measure of a variable's spread around its mean value and is computed as the mean square of deviations from the mean. It is equal to the square of the standard deviation.

65.9 CHAPTER SUMMARY

Chemometrics is the area of MVA applied to chemical data; however, this definition can be extended to the physical and biological worlds while still maintaining the same definition. Chemometrics aims to provide insights into complex and sometimes large data sets. Traditional methods of plotting one variable at a time fail very quickly when the number of variables in a data table becomes large and the tools of chemometrics can be used to overcome these obstacles.

Chemometrics has traditionally been associated with spectroscopic or chromatographic data where typical profiles can consist of hundreds to thousands of points per sample. To investigate the finer or hidden details in the data, chemometrics is used. This is why the methods used in chemometrics are sometimes called latent methods (i.e., they look for the hidden structure in the data). In today's world of process sensors, multivariate data arrays can also be generated from many sensor values fused together. Thus, the data move from being homogeneous (such as a spectrum) to heterogeneous (i.e., temperature, pressure, and pH being simultaneously monitored). For heterogeneous data, weighting strategies must be employed that allow individual variables to contribute to a model on an equal basis.

There are three main approaches to chemometric modeling, these being EDA, regression modeling, and multivariate classification. EDA should be the first method applied to a multivariate data set in order to identify any patterns in the data in an unsupervised manner, that is, the data should be analyzed without any preassumptions applied. From there, an analyst can make certain decisions regarding the homogeneity of the data, whether multiple class models should be developed or whether the natural.

Although there are a number of EDA methods available, the most powerful and the workhorse of chemometric methods is PCA. PCA provides a map of sample relationships (known as the scores plot) and a map of variable correlations (known as the loadings plot). When interpreted together, samples groupings observed in the scores can be interpreted based on how the variables measured correlate to each other. Another powerful property of PCA is the complete range of diagnostic tools available for the interpretation and validation of models. This provides PCA not only with a complete range of data modeling tools but also allows PCA models to be used in real-time process monitoring applications. In the case where a new sample sits outside of the calibration population, it can be investigated and the variables that cause the sample to deviate can be investigated and corrected. This is the basis of what is known as MSPC.

Regression modeling aims to make a model that relates a set of independent (X -variables) to a set of dependent (Y -variables). Multivariate regression methods utilize latent variable approaches to model the structure in X and relate to the structure in Y . Although MLR is not a latent variable approach, if the scores from PCA are used as variables in the MLR model, the method of PCR results. PCR models the X -variables independent from the Y -variables at first via PCA. It then regresses the Y -variables against the scores to create a prediction model. One of the downsides of PCA is that in some cases, the first score in PCA may not be relevant to modeling Y ; therefore, an inflation of variance occurs. As subsequent PCs containing chemical (or other) information are added to the model, the model is better able to describe the variability in Y .

To overcome the limitations of PCR, the PLSR method was developed. Unlike PCR, PLSR models the X - and Y -data simultaneously, finding the factors in X most correlated to Y . This means that the PLSR algorithm will in general converge faster than the PCR algorithm, however, generally yielding the same solution. The PLSR algorithm contains a number of useful diagnostic tools for interpreting and validating the model. From a purist's point of view, PCR is fundamentally simpler and it forces an analyst to better understand their data and the preprocessing used to make a model such that the first component is related to chemical (or other) information. PLSR is the algorithm of choice in most software packages and has therefore become the de facto multivariate regression method. Other regression methods exist, for example, support vector machine regression (SVMR), which are nonlinear approaches to multivariate regression but are outside of the scope of this chapter.

Multivariate classification is the qualitative counterpart of multivariate regression. In multivariate classification, rules are developed, typically during the EDA stage of

analysis, and are used to assign new samples into existing classes. This area is known as supervised methods or pattern recognition. There are a number of multivariate classification approaches available, including LDA and SVMC; however, the most effective methods are again based on latent variable methods. These are SIMCA and PLS-DA.

SIMCA utilizes PCA or PLS models to project new samples onto the PCA or PLS loadings to generate new scores values. These scores are compared to each model in the SIMCA library, and there are three possible outcomes: unique classification, ambiguous classification, and no classification.

A unique classification is based on a predefined statistical confidence interval, thus allowing SIMCA to be validated on a purely statistical basis. An ambiguous classification situation occurs when a sample lies in the same space as two or more classes simultaneously. In order to resolve ambiguities, another model has to be generated and the use of SIMCA in a hierarchical approach can be very useful for the classification of complex systems. SIMCA has many diagnostic tools for the interpretation and validation of a model; however, its greatest strength is its ability to classify a sample into a null class. Methods such as LDA will attempt to put a sample into the nearest class, independent of how far away the sample is from a class. SIMCA will reject such as sample based on the measures of X -residual and leverage used to define a class model.

PLS-DA utilizes the PLSR algorithm but instead of having continuous Y -variables, it uses a binary system to define classes in a data set. For a two-class problem, class A can be designated as a zero (0) and class B designated as a one (1). The PLS algorithm is applied and the predicted versus reference plot can be used to assess the model's ability to classify new samples. When the problem extends to three or more classes, the form of the PLSR model becomes the PLS-2 model and each class will have its own Y -column. Where the class is represented, it has a designation of one (1) in the columns and all other classes are designated zero (0). The same diagnostic and interpretation tools for regular PLSR can be used for PLS-DA.

Until a model is used for a practical application, it is of little value. This is where the area of MSPC is gaining more attention in industry because it allows the combination of traditional SPC and the multivariate methods in unique ways such that process control and fault detection becomes proactive, not reactive. MSPC approaches can be applied to continuous (single or multiple stage) processes or to batch processes. The model development aspects of continuous and batch models vary greatly and both require unique methods to model the system under investigation.

In order for MSPC to be effective, a robust data management system is required to collect data from multiple sources (either scalar sensors or vector sensors such as spectrometers) and fuse the data together in order to generate representative data arrays. These arrays can be used for modeling or for predictive purposes and are sure to find widespread usage in all industries moving forward, particularly because of their ability to be used in closed-loop control systems.

Chemometrics is wide and diverse. It had its infancy over 30 years ago and has now matured into a scientific approach for a number of research and industrial applications. It utilizes as much data as can be analyzed and provides insights into sample patterns, variable relationships, outliers, and the overall importance of variables being used to analyze a system.

REFERENCES

1. Wold, S., "Chemometrics; what do we mean with it, and what do we want from it?". *Chemometrics and Intelligent Laboratory Systems* 30 (1): 109–115 1995.
2. Dayal, B. S. and MacGregor, J. F., "Improved PLS algorithms". *Journal of Chemometrics* 11: 73–85 1997.
3. Rannar, S. Lindgren, F. Geladi, P., and Wold, S., "A PLS kernel algorithm for data sets with many variables and fewer objects, Part 1: theory and algorithm". *Journal of Chemometrics* 8: 111–125 1994.
4. Box, G. E. P. Hunter, J. S., and Hunter, W. G., "*Statistics for Experimenters, An Introduction to Design, Data Analysis, and Model Building*", John Wiley & Sons, Inc., New York 1978.
5. Montgomery, D. C., "*Design and Analysis of Experiments*", 6th Edition, John Wiley & Sons, Inc., Hoboken, NJ 2004.
6. Montgomery, D. C. and Myers, R. H., "*Response Surface Methodology, Process and Product Optimization Using Designed Experiments*", 2nd Edition, John Wiley & Sons, Inc., New York 2002.
7. Whitcomb, P. J. and Anderson, M. J., "*RSM Simplified, Optimizing Processes Using Response Surface Methods for Design of Experiments*", Productivity Press, New York 2005.
8. Montgomery, D. C., "*An Introduction to Statistical Quality Control*", 5th Edition, John Wiley & Sons, Inc., Hoboken, NJ 2005.
9. ICH Harmonized Tripartite Guideline Q2(R1), "Validation of analytical procedures: text and methodology". *Federal Register* 62(96): 27463–7 1997.
10. Swarbrick, B., "*Multivariate Analysis for Dummies*", John Wiley & Sons, Inc., Hoboken, NJ 2012.
11. Miller, J. N. and Miller, J. C., "*Statistics and Chemometrics for Analytical Chemistry*", 5th Edition, Prentice Hall, New York 2005.
12. Adams, M. J., "*Chemometrics in Analytical Spectroscopy*", The Royal Society of Chemistry, Cambridge 1995.
13. Everitt, B. S., Landau, S., and Leese, M., "*Cluster Analysis*", 4th Edition, John Wiley & Sons, Inc., New York 2001.
14. Fisher, R. A., "The use of multiple measurements in taxonomic problems". *Annals of Eugenics* 7 (2): 179–188 1936.
15. Martens, H. and Naes, T., "*Multivariate Calibration*", John Wiley & Sons, Inc., New York 1989.

16. Pearson, K., "On lines and planes of closest fit to systems of points in space". *Philosophical Magazine* 2 (11): 559–572 1901.
17. Naes, T. Issakson, T. Fearn, T., and Davies, T., "A User Friendly Guide to Multivariate Calibration and Classification", NIR Publications, Chichester 2002.
18. Esbensen, K. H., "Multivariate Data Analysis in Practice", 5th Edition, CAMO Software, AS., Oslo, Norway 2012.
19. Hotelling, H., "Analysis of a complex of statistical variables into principal components". *Journal of Educational Psychology* 24: 417–441 1933.
20. Mahalanobis, P. C., "On the generalised distance in statistics". *Proceedings of the National Institute of Sciences of India* 2 (1): 49–55 1936.
21. Jackson, J. E. and Mudholkar, G. S., "Control procedures for residuals associated with principal component analysis". *Technometrics* 21 (3): 341–349 1979.
22. Jackson, J. E., "A User Friendly Guide to Principal Components", John Wiley & Sons, Inc., New York 1991.
23. Wold, S. and Esbensen, K., "Principal component analysis". *Chemometrics and Intelligent Laboratory Systems* 2: 37–52 1987.
24. Hastie, T. Tishbirani, R., and Friedman, J., "The Elements of Statistical Learning, Data Mining, Inference and Prediction", 2nd Edition, Springer Science and Business Media, New York 2009.
25. Draper, N. R. and Smith, H., "Applied Regression Analysis", 3rd Edition, Wiley-Interscience, New York 1998.
26. Daintith, J., "A Dictionary of Chemistry", 6th Edition, Oxford University Press, Oxford 2008.
27. Legendre, A. M., "Sur la Méthode des moindres carrés". In: *Nouvelles méthodes pour la détermination des orbites des comètes*, Firmin Didot, Paris, 72–80 1805.
28. Gauss, C. F., "Theoria Combinationis Observationum Erroribus Minimis Obnoxiae", Henrich Dieterich, Göttingen 1823
29. Sharaf, M. A. Illman, D. L., and Kowalski, B. R., "Chemometrics", John Wiley & Sons, Inc., New York 1986.
30. Moore, E. H., "On the reciprocal of the general algebraic matrix". *Bulletin of the American Mathematical Society* 26: 394–395 1920.
31. Penrose, R., "A generalized inverse for matrices". *Proceedings of the Cambridge Philosophical Society* 51: 406–413 1955.
32. Jolliffe, I. T., "A note on the use of principal components in regression". *Journal of the Royal Statistical Society, Series C* 31: 300–303 1982.
33. Mason, R. L. and Gunst, R. F., "Selecting principal components in regression". *Statistical and Probability Letters* 3: 299–301 1985.
34. Wold, S., Martens, H., and Wold, H., "The multivariate calibration problem in chemistry solved by the PLS methods". In: Ruhe, A. and Kågstöm, B. (eds.), *Matrix Pencils: Proceedings of a Conference Held at Pite Havsbad, Sweden, March 22–24, 1982*, Springer Verlag, Heidelberg, 286–293 1983.
35. Geladi, P. and Kowalski, B. R., "Partial least squares regression: a tutorial". *Analytica Chimica Acta* 185: 1–17 1986.

36. Gower, J., "A general theory of biplots". In: Krzanowski W. J. (ed.), *Recent Advances in Descriptive Multivariate Statistics*. Royal Statistical Society Lecture Notes, 2, Oxford University Press, Oxford, 283–303 1995.
37. Kjeldahl, K. and Bro, R., "Some common misunderstandings in chemometrics". *Journal of Chemometrics* 24: 558–564 2010.
38. Seasholtz, M. B. and Kowalski, B. R., "Qualitative information for multivariate calibration models". *Applied Spectroscopy* 44: 1337–1348 1990.
39. Nadler, B. and Coifman, R. R., "Partial least squares, Beer's law and the net analyte signal: statistical modeling and analysis". *Journal of Chemometrics* 19: 45–54 2005.
40. Broad, N. Graham, P. Hailey, P. Hardy, A. Holland, S. Hughes, S. Lee, D. Prebble, K. Salton, N., and Warren, P. Guidelines for the Development and Validation of Near-Infrared Spectroscopic Methods in the Pharmaceutical Industry", In: Chalmers, J. M. and Griffiths, P. R. (eds.), *Handbook of Vibrational Spectroscopy*, John Wiley & Sons, Inc., New York 2002.
41. Trygg, J. and Wold, S., "Orthogonal projections to latent structures (O-PLS)". *Journal of Chemometrics* 16 (3): 119–128 2002.
42. Fisher, R. A., "The use of multiple measurements in taxonomic problems". *Annals of Eugenics* 7 (2): 179–188 1936.
43. Wold, S., "Pattern recognition by means of disjoint principal components model". *Pattern Recognition* 8: 127–139 1976.
44. Eriksson, L. Johansson, E. Kettaneh-Wold, N. Trygg, J. Wikström, C., and Wold, S., "*Multi- and Megavariate Data Analysis Part I: Basic Principles and Applications*", Umetrics Inc, Umeå, Sweden 2006.
45. Vong, R. Geladi, P. Wold, S., and Esbensen, K., "Source contributions to ambient aerosol calculated by discriminant partial least squares regression (PLS)". *Journal of Chemometrics* 2: 281–296 1988.
46. Cristianini, N. and Shawe-Taylor, J., "*An Introduction to Support Vector Machines and other Kernel-Based Learning Methods*", Cambridge University Press, New York 2000.
47. Luts, J. Ojeda, F. Van de Plas, R. De Moor, B. Van Huffel, S., and Suykens, J. A. K., "A tutorial on support vector machine-based methods for classification problems in chemometrics". *Analytica Chimica Acta* 665: 129–145 2010.
48. Harshman, R. A., "How can I know if it's real?" A catalogue of diagnostics for use with three-mode factor analysis and multidimensional scaling". In: Low, H. G., Snyder, Jr., C. W., Hattie J. and McDonald R. P. (eds.), *Research Methods for Multi-Mode Data Analysis*, Praeger, New York, 566–591 1984.
49. Bro, R. Kjeldahl, K. Smilde, A. K., and Kiers, H. A. L., "Cross-validation of component models: a critical look at current methods". *Analytical and Bioanalytical Chemistry* 390: 1241–1251 2008.
50. Esbensen, K. H. and Geladi, P., "Principles of proper validation: use and abuse of re-sampling for validation". *Journal of Chemometrics* 24: 168–187 2010.
51. Kennard, R. W. and Stone, L. A., "Computer aided design of experiments". *Technometrics* 11 (1): 137–148 1969.

52. Stone, M., "Cross-validators choice and assessment of statistical predictions". *Journal of the Royal Statistics Society* 36: 111–147 1974.
53. Kos, G. Lohniger, H., and Krska, R., "Validation of chemometric models for the determination of deoxynivalenol on maize by mid-infrared spectroscopy". *Micotoxin Research* 19: 149–153 2003.
54. MacGregor, J. and Kourti, T., "Statistical process control of multivariate processes". *Control Engineering in Practice* 3 (3): 403–414 1995.
55. Zhao, C. Zhao, Y. Su, H., and Huang, B., "Economic performance assessment of advanced process control with LQG benchmarking". *Journal of Process Control* 19 (4): 557–569 2009.
56. ICH, "Pharmaceutical development". ICH Harmonized Tripartite Guideline Q8(R2), Federal Register, Vol. 71 (98) 2009.
57. Nomikos, P. and MacGregor, J. F., "Monitoring of batch processes using multi-way principal component analysis". *AIChE Journal* 40: 1361–1375 1994.
58. Kassidas, A. MacGregor, J., and Taylor, P., "Synchronization of batch trajectories using dynamic time warping". *AIChE Journal* 44: 864–875 1998.
59. Smilde, A. Bro, R., and Geladi, P. "*Multi-Way Analysis, Applications in the Chemical Sciences*", John Wiley & Sons, Inc., Hoboken, NJ 2004.
60. Meng, X. Morris, A. J., and Martin, E. B., "On-line monitoring of batch processes using a PARAFAC representation". *Journal of Chemometrics* 17: 65–85 2003.
61. Westad, F. Gidskehaug, L. Swarbrick, B., and Flaaten, G.R., "Assumption free modeling and monitoring of batch processes". *Chemometrics and Intelligent Laboratory Systems* 149: 66–72 2015.
62. IBM, "Accelerate delivery of pervasive analytics with a big data platform". <http://www.ibm.com/analytics/in/en/what-is-smarter-analytics/innovate-with-analytics-tools.html> (Accessed December 10, 2015).

LIQUID CHROMATOGRAPHY

ZHAO LI, SANDYA BEERAM, CONG BI, ELLIS KAUFMANN,
RYAN MATSUDA, MARIA PODARIU, ELLIOTT RODRIGUEZ,
XIWEI ZHENG, AND DAVID S. HAGE

Department of Chemistry, University of Nebraska, Lincoln, NE, USA

66.1 INTRODUCTION

Chromatography is a separation method in which chemicals or sample components are separated by their different rates of travel through a system that contains a stationary phase and a mobile phase (see Fig. 66.1) [1–5]. The mobile phase acts to transport the sample components through the system. The stationary phase is used to interact with some or all of these components. The stationary phase is held in place by a solid support and may consist of the surface of this support, a coating on the support, or a bonded layer on the support. If the sample components have different degrees of interactions with the stationary phase, they will spend different amounts of time in the mobile phase and, thus, travel at different rates through the chromatographic system. The result is a separation of these chemicals based on their interactions with the stationary phase and mobile phase [3–6].

Liquid chromatography (LC) is a type of chromatography in which the mobile phase is a liquid [1–6]. The use of LC with a support held in a column, as is shown in Figure 66.1, was first described by the Russian botanist Mikhail Tswett in 1903 who used this technique to separate plant pigments [5, 7]. Since that time, LC has become an important separation component in many analytical assays and chemical purification methods. The applications of this method range from small ions and molecules up to large biological molecules and polymers. LC is also used in fields that span from biomedical research, pharmaceutical science, and clinical chemistry to environmental

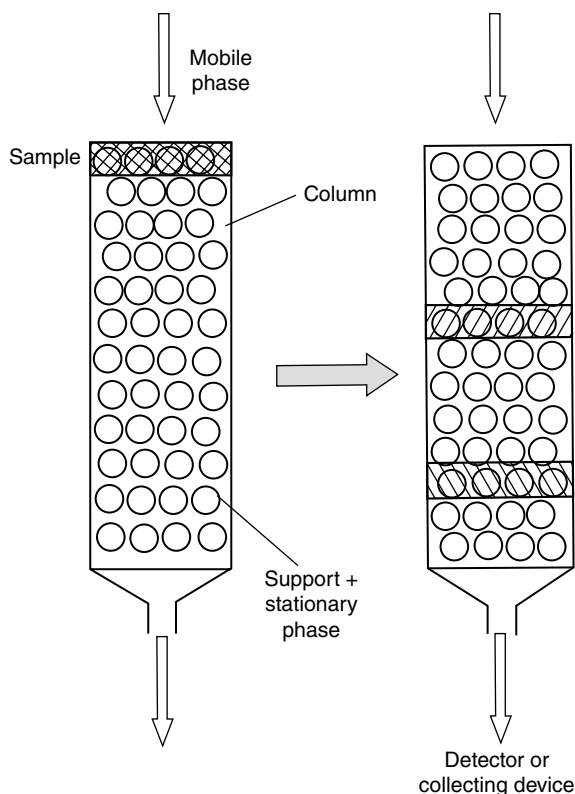


FIGURE 66.1 A typical separation by chromatography, as carried out by using a simple column LC system.

testing, food testing, quality control, and the large-scale purification of chemicals and biochemicals [1–6].

LC can be used in a relatively simple, manual form for chemical isolation and purification, as shown in Figure 66.1 [3–5]. However, instrumental forms of LC are frequently used in many modern applications. A general design of an instrument for conducting modern LC is shown in Figure 66.2. This instrument is known as a *liquid chromatograph* [5, 9]. In this type of system, a pump is used to apply the mobile phase to the column, and samples are injected into the mobile phase stream by means of an injection valve, autoinjector, or similar device. The mobile phase and sample are then passed through a column, which contains the support and stationary phase. A detector is often used to monitor and measure the sample components as they exit the column. Alternatively, a fraction collector may be utilized to collect portions of the eluted components for later detection or use in other methods [5, 9].

A typical separation that is obtained in modern LC is shown in Figure 66.2 [8]. This separation is often represented by a *chromatogram*, which is a plot of the response of the detector or of the measured amount of the eluting compounds as a function of either

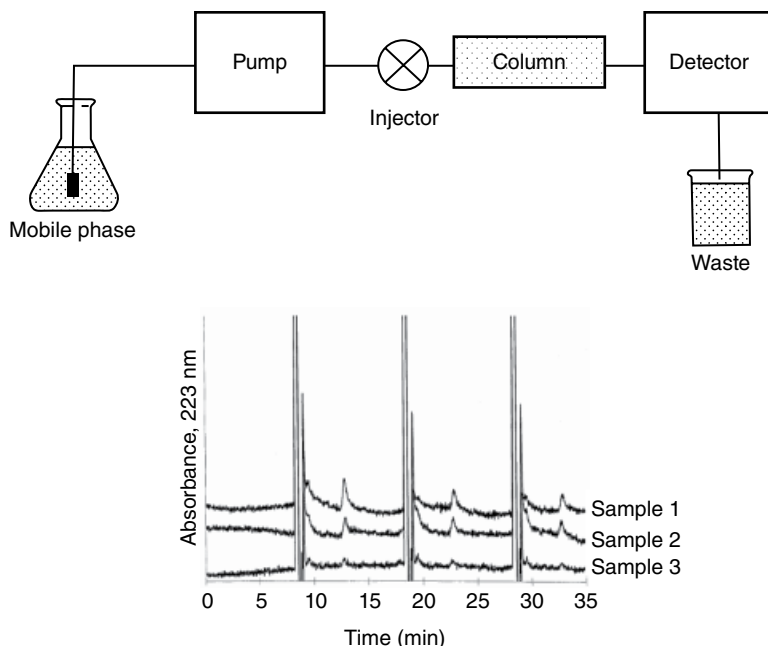


FIGURE 66.2 General design of a modern liquid chromatograph (top) and a set of separations (each done in triplicate) obtained by high-performance liquid chromatography for the analysis of herbicides in three water samples (bottom). The bottom example is adapted with permission from Ref. 8.

the time that has elapsed since the sample was applied to the column (as given by a compound's retention time, t_R) or the volume of mobile phase that has been applied to the column up to this point in time (as given by the compound's retention volume, V_R) [1, 2, 5]. The values of the retention time and retention volume for a chemical on a column are related to a compound's interactions with the stationary phase versus the mobile phase and can be used to identify this substance. The height or area of the peak that is obtained for the same compound can be related to the amount of that substance that was present in the original applied sample [5].

The fact that LC makes use of a liquid as the mobile phase means that the sample and compounds to be separated must be soluble to some extent in this liquid [4–6]. This requirement is needed so that the sample and its components can be applied to and eluted from the column through the use of this mobile phase. The need for solubility in a liquid mobile phase is a much less stringent requirement than what is required in the related technique of gas chromatography (GC), which instead needs sample components to be sufficiently volatile to enter the gas phase for their separation and injection. This difference in requirements, which is due to the use of a liquid instead of a gas as the mobile phase, is a major advantage for LC in terms of its range of applications. For instance, LC can be used with even relatively large biological compounds, such as proteins or DNA, that are difficult to place into the gas phase. The use of a liquid mobile

phase also means that LC is usually operated at a much lower temperature than GC, which makes LC more applicable for work with thermally unstable chemicals [4, 5, 10].

66.2 SUPPORT MATERIALS IN LC

The type of support that is used in LC is often used to characterize this method based on its efficiency or “performance.” For instance, the diameter of a particulate support in LC will affect the plate height (H) and number of theoretical plates (N) of such a system, with smaller diameters leading to systems with higher resolution due to their small values for H and large values for N [4–6].

The first supports that were used in LC were relatively large and nonrigid materials. Examples of these materials included large particles of silica, alumina, agarose, and cellulose [3–6, 11–13]. These supports are inexpensive and have relatively low back pressures, allowing their use with flow based on gravity or a peristaltic pump. However, these same materials tend to have slow mass transfer properties, which lead to poor efficiencies and large plate heights or small plate numbers. This, in turn, often results in separations with broad peaks, only moderate detection limits, and relatively long analysis times. Such a method is sometimes known as “column chromatography” or *low-performance liquid chromatography* [3–6].

High-performance liquid chromatography (HPLC) is a form of LC that utilizes smaller diameter supports (e.g., porous supports with typical diameters of 10 μm or less) or materials with better mass transfer properties when compared to the supports used in low-performance methods [3–6, 9, 11, 12]. These features provide HPLC with a much better efficiency and resolution than traditional LC, which also make it possible to separate a greater number of substances in a shorter period of time and to obtain sharper peaks with lower limits of detection. The use of more efficient support materials in HPLC requires the use of special pumps and instrumental components that can be used at mobile phase pressures ranging from a few hundred to a few thousand psi (e.g., see Fig. 66.2) [3–6]. Higher pressure systems involving even smaller support particles and pressures of 5000–6000 psi or higher are used in an extension of HPLC known as *ultra performance liquid chromatography (UPLC)* [5, 14, 15].

There are many materials that can be used as supports in traditional LC and in HPLC. Particle-based supports have been used in all of these methods, with pellicular supports, perfusion supports and monolithic supports also being of interest in HPLC-related methods [5, 6, 12, 16–22]. Traditional porous particles contain pores throughout their structure, while pellicular supports are made from nonporous particles that are coated with a thin layer of a porous material [12]. The use of small porous particles or pellicular supports helps to provide good efficiency by minimizing the distance that solutes have to travel as they move through the pores or porous layer. Perfusion particles have both small side pores and large through pores that act in a similar manner to improve the efficiency of a column [12]. Monolithic supports consist of a continuous

bed of a porous polymer and also provide improved mass transfer properties compared to traditional particle-based supports [16–18].

Another possible form of LC is *planar chromatography* [2, 3, 5, 23]. This method is carried out by using a stationary phase that is present on a planar support. Examples are *paper chromatography*, which uses paper as the support, and *thin-layer chromatography (TLC)*, which uses some other material as the support (e.g., silica particles coated on a glass or plastic plate) [5, 23]. While column-based forms of LC separate solutes based on the time or volume of the mobile phase that is required to have these solutes travel a given distance (e.g., through the column), planar chromatography separates solutes based on the distances they travel in a given amount of time [3, 5]. In TLC the use of small particles (e.g., 5–8 μm diameter porous silica) that are coated on a surface is a method known as *high-performance thin layer chromatography (HPTLC)* [3, 24, 25]. Recently, the use of even thinner support layers (e.g., based on monoliths or nanomaterials) has led to a method known as *ultrathin-layer chromatography (UTLC)* [26, 27].

66.3 ROLE OF THE MOBILE PHASE IN LC

The retention of a solute in LC can depend on the solute's interactions with both the mobile phase and stationary phase. This means the retention of solutes in LC can also be controlled by varying the type of mobile phase that is applied to a column containing a given stationary phase. Mobile phases in LC can be placed into two categories depending on how similar or different they are in their chemical interactions when compared with the stationary phase. A *strong mobile phase* is a mobile phase that causes a solute to have weak retention on a column, which occurs when the solute tends to spend more time interacting and flowing with the mobile phase than it does interacting with the stationary phase. A *weak mobile phase*, on the other hand, is a mobile phase that produces high retention for a solute on a column. In this case, the interactions of the solute with the stationary phase are favored over the solute's interactions with the mobile phase. In either situation, the type of liquid or solution that constitutes a weak or strong mobile phase will depend on the type of stationary phase that is present in the column [5].

Isocratic elution is a term used to describe a chromatographic separation that uses a mobile phase with a constant composition [2, 5]. A change in the composition of the mobile phase during a chromatographic separation is referred to as *solvent programming* [2, 3, 5]. Solvent programming begins with a weak mobile phase, to allow good retention to be obtained for early eluting solutes, and then moves to a strong mobile phase, to allow later eluting solutes to leave the column in a reasonable amount of time [2–5]. The change in the composition of the mobile phase during solvent programming may be done in a linear fashion, by using a step change or by utilizing a nonlinear change in the mobile phase composition over time [4, 5].

66.4 ADSORPTION CHROMATOGRAPHY

There are five types of LC based on the mechanisms by which they separate solutes. These types are adsorption chromatography, partition chromatography, ion-exchange chromatography (IEC), size-exclusion chromatography (SEC), and affinity chromatography [2, 5]. *Adsorption chromatography* is a type of LC that separates solutes based on their adsorption to the surface of the support. This method is also known as liquid–solid chromatography [1, 2, 4, 5].

The process that gives rise to retention in adsorption chromatography is shown in Equation 66.1. This process involves the binding of solute S to the surface of a support (as represented by the subscript “Support”) in place of n molecules of the mobile phase (M) [3–5]:



The retention of the solute in this type of column will depend on the binding strength of the solute to the support and the surface area of the support [3, 4]. This retention will also depend on the amount of mobile phase that is displaced from the surface by the solute and the strength with which the mobile phase binds to the support. The strength of the mobile phase in binding to a given support is described by a term known as the *eluotropic strength* (ϵ°) [2, 4]. A mobile phase with a large eluotropic strength will bind strongly to the support, which will cause a displaced solute to spend more time in the mobile phase and elute more quickly from the column.

Silica and alumina are the most common stationary phases and supports that are employed in adsorption chromatography. Because these stationary phases are polar in nature, they will retain polar compounds to the greatest degree. Carbon-based materials can be used as nonpolar supports in adsorption chromatography and will retain nonpolar solutes the most. Other supports that have been used in adsorption chromatography are florisil, polyamides, and celite. Increasing the surface area for any of these supports will result in stronger solute retention because this increases the amount of stationary phase versus mobile phase that is present [3–5].

A liquid or solution with a low eluotropic strength on a given support will bind weakly to this material and act as a weak mobile phase for the same support in adsorption chromatography. A liquid or solution with a high eluotropic strength for a support will act as a strong mobile phase for this material. For instance, toluene or heptane are weak mobile phases on a polar support such as silica or alumina but are strong mobile phases on a nonpolar support such as charcoal [3–5].

Adsorption chromatography is a relatively inexpensive and general method for the purification or isolation of organic compounds. For instance, this approach is often used to separate starting materials from products following an organic synthesis [3–5]. This method is especially useful in separating geometrical isomers, such as *cis/trans*-isomers or *ortho/meta/para*-isomers. Adsorption chromatography also forms the basis

of many TLC methods. In this type of application, adsorption chromatography is used for the screening and semiquantitative analysis of chemicals such as drugs of abuse and amino acids [3–5, 23–25].

66.5 PARTITION CHROMATOGRAPHY

Partition chromatography is a type of LC in which solutes are separated based on the degree to which they partition between the mobile phase and a stationary phase that is coated on or bonded to a support. This method is also sometimes known as liquid–liquid chromatography.

The retention of a solute (S) in partition chromatography can be described by the following reaction [3–5]:



The degree of retention for the solute in partition chromatography will depend on the relative solubility of this solute in the stationary phase versus the mobile phase. This retention will also depend on the amount of the stationary phase versus mobile phase that is present in the column [3–5]. Originally in partition chromatography, the support was coated with a liquid stationary phase, which was also immiscible with the mobile phase. However, many modern types of partition chromatography use stationary phases that are chemically bonded to the support [3, 5].

Based on the polarity of the stationary phase, partition chromatography can be divided into two categories: *normal-phase liquid chromatography* (NPLC) and *reversed-phase liquid chromatography* (RPLC) [2–5]. NPLC (or normal-phase chromatography) is a type of partition chromatography that uses a polar stationary phase. Early columns and systems for carrying out NPLC used supports like silica that were coated with liquids such as water, ethylene glycol, dimethyl sulfoxide, or ethylenediamine as the stationary phase. However, these liquid stationary phases could be lost from the column over time in a process known as *column bleed* [1, 5]. To avoid this problem, modern NPLC supports instead often use chemically bonded phases that contain polar functional groups like cyanopropyl, aminopropyl, or diol groups (see Fig. 66.3) [3–5, 23, 28].

The presence of a polar stationary phase in NPLC means that the weak mobile phase is a nonpolar liquid or solution (e.g., hexane, heptane, or octane). The strong mobile phase in NPLC is a more polar liquid or solution (e.g., tetrahydrofuran, ethanol, or 2-propanol) [3–5, 23, 28]. A mixture of a weak mobile phase and a small amount of a miscible strong mobile phase may be used for isocratic elution or a change from a weak mobile phase to a strong mobile phase over time may be used for gradient elution [3–5]. The solutes in NPLC will elute in the order of their polarity. The least polar solutes will have the weakest retention and will elute from the column first,

Normal phase liquid chromatography (NPLC)

Aminopropyl phase	Support-CH ₂ CH ₂ CH ₂ NH ₂
Cyanopropyl phase	Support-CH ₂ CH ₂ CH ₂ CN

Reversed phase liquid chromatography (RPLC)

Octyl phase (C ₈)	Support-(CH ₂) ₇ CH ₃
Octadecyl phase (C ₁₈)	Support-(CH ₂) ₁₇ CH ₃

FIGURE 66.3 Some common stationary phases that are used in normal-phase liquid chromatography and reversed-phase liquid chromatography.

while more polar solutes will be more strongly retained and elute later. The applications of NPLC are similar to those for adsorption chromatography with a polar support in that both methods are general purpose tools for the isolation of organic compounds [3–5, 23, 28].

RPLC (or reversed-phase chromatography) is a type of partition chromatography that uses a nonpolar stationary phase [1–5]. Coatings of nonpolar liquids like heptane, squalene, and hydrocarbon polymers were originally used as the stationary phases in this method. These liquid coatings, however, were subject to the same issues with column bleed that were noted previously for liquid stationary phases in NPLC. Most current stationary phases that are used in RPLC consist of a support such as silica that contains a chemically bonded phase with an *n*-alkane or some other nonpolar group. The most common bonded phases that are used in RPLC are *n*-octyl (C₈) and *n*-octadecyl (C₁₈) groups (see Fig. 66.3). Other bonded phases that are often used are *n*-butyl (C₄) and phenyl groups [3–5, 23, 28].

Solutes in RPLC elute in order of their decreasing polarity. This means the most polar compounds in a sample will elute from the column first followed by more nonpolar solutes. A weak mobile phase in RPLC is polar and is often water or an aqueous buffer. Less polar liquids such as acetonitrile, methanol, and 2-propanol are often used as strong mobile phases. Agents like triethylamine and trifluoroacetic acid may also be added to the mobile phase to prevent interactions of polar silanol groups on a silica support with the injected solutes [3–5].

An important advantage of RPLC is that water acts as a weak mobile phase for this method. This is useful because it makes RPLC compatible with the injection of aqueous-based samples, such as biological samples, food or agricultural samples, and many environmental samples. The fact that RPLC separates solutes based on polarity is another valuable feature. As a result of these combined advantages, RPLC is a popular method in areas such as biochemical research, pharmaceutical analysis, clinical testing, food analysis, and environmental analysis (see Fig. 66.2). Examples of

chemicals and biochemicals that have been separated by this method include drugs, fatty acids, amino acids, peptides, proteins, and nucleic acids, among many others [3–5, 9, 10, 23, 28].

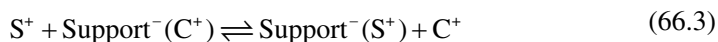
Another type of partition chromatography is *hydrophilic interaction liquid chromatography (HILIC)*. HILIC uses a support with polar functional groups and a mobile phase that often consists of a mixture of water and a miscible organic solvent like acetonitrile [29]. Solutes are separated in this method as they partition between a region near the surface of the support that contains water and an area in the mobile phase that is enriched in the organic solvent [29, 30]. Like NPLC and RPLC, this method separates chemicals based on their polarity. The use of some water in the mobile phase makes this method more convenient to use than NPLC with polar compounds that may have low solubility in a nonpolar mobile phase. Applications of HILIC have included its use in proteomics and in the separation and analysis of polar compounds such as drugs and drug metabolites in clinical samples [30–32].

66.6 ION-EXCHANGE CHROMATOGRAPHY

IEC is a type of LC that separates charged solutes based on their interactions with a stationary phase containing fixed groups with a charge opposite to that of the solutes [1, 2]. IEC can be divided into two categories based on the charge of the stationary phase: *cation-exchange chromatography* and *anion-exchange chromatography* [2–5].

Cation-exchange chromatography uses a stationary phase that has negatively charged groups. This method is used to retain and separate positive ions (cations). The stationary phase that is used in this type of IEC may be the conjugate base of a strong acid (e.g., sulfonate) or it may be a conjugate base of a weak acid (e.g., a carboxylate group). In anion-exchange chromatography, the stationary phase has positively charged groups and is used to separate negative ions (anions). The stationary phase in this second type of IEC can be a conjugate acid of a strong base (e.g., a protonated quaternary amino group) or a conjugate acid of a weak base (e.g., the protonated form of a diethylaminoethyl group) [3–5]. Examples are provided in Figure 66.4.

Retention and elution in IEC can be described by the competition of a solute ion and a competing ion in the mobile phase for stationary phase sites having the opposite charge. This process is illustrated in Equation 66.3 for the retention of a positively charged solute (S^+) on a negatively charged cation-exchange support (Support^-) and in the presence of a positively charged competing ion (C^+):



A similar reaction can be written for the retention of a negatively charged solute in anion-exchange chromatography. The degree to which the solute ion will be retained by the stationary phase in IEC will depend on how strongly the solute ion and competing

<i>Cation-exchange chromatography</i>	
Sulfonic acid	Support-SO ₃ ⁻ H ⁺
Carboxylic acid	Support-COO ⁻ H ⁺
<i>Anion-exchange chromatography</i>	
Quaternary amine	Support-CH ₂ N(CH ₃) ₃ ⁺ Cl ⁻
Diethylaminoethyl (DEAE)	Support-O(CH ₂) ₂ NH ⁺ Cl ⁻ $\begin{array}{c} \text{CH}_2\text{CH}_3 \\ \diagup \\ \text{CH}_2\text{CH}_3 \end{array}$

FIGURE 66.4 Some common stationary phases that are used in anion-exchange chromatography and cation-exchange chromatography.

ion each bind to the fixed charges on the support, the amount of these charged sites that are present, and the concentration of the competing agent [3–5].

Various types of supports have been used in IEC. Silica that has been modified to contain charged groups is one example. Polystyrene supports that have been modified to contain positively or negatively charged groups are often used in IEC to separate small inorganic and organic ions [3, 4]. Carbohydrate-based supports such as agarose, cross-linked dextran, or cellulose have also been modified to contain charged groups and, due to their relatively large pore sizes and low nonspecific binding, are often used to separate charged biological agents such as proteins and nucleic acids [3–5].

A weak mobile phase in IEC will be a solution that has few or no competing ions present. A weak mobile phase in this method should also have a pH that is optimum for creating the charges needed for binding to occur between the applied solutes and the stationary phase. An increase in the concentration of the competing ion is often used to lower the retention of solute ions in IEC and for gradient elution in this chromatographic method. The pH of the mobile phase can also be used to adjust retention if either the solute or the stationary phase is the conjugate acid or base of a weak base or weak acid. The addition of a complexing agent to the mobile phase is another way of altering the retention of some solute ions in IEC. For example, a sample that contains Fe³⁺, which would normally be separated by cation-exchange chromatography, can be combined with Cl⁻ to form FeCl₄⁻, which could then be separated by anion-exchange chromatography [3–5, 9].

One application of IEC is the use of this technique to remove ionic components from samples and solutions. For instance, cation-exchange and anion-exchange supports are used in water purification systems to produce deionized water by replacing the cations in water with H⁺ and the anions with OH⁻. In addition, IEC columns and supports are used in biochemistry to concentrate and purify proteins, peptides, and nucleotides based on their charges at a given pH or based on their isoelectric points (i.e., the pH at which a zwitterionic solute has a net neutral charge). Another application of IEC is its

use to concentrate and analyze organic or inorganic ions in food, environmental samples, and commercial products [3–5, 9, 33, 34].

Ion chromatography is a special type of IEC that is used with a conductivity detector for chemical analysis. A conductivity detector gives a response that is related to the total ionic content of a solution. Many types of traditional IEC can use a fairly high concentration of a competing ion to elute a solute, which will result in a high background signal on this type of detector. In ion chromatography, the concentration of the competing ion that is needed for solute elution is decreased by using a stationary phase that has a small number of charged sites. Furthermore, the IEC column that is used to separate the desired ions in a sample is combined with a suppressor column or membrane separator, which contains groups that have an opposite charge to those present in the first ion-exchange column. This second column/membrane is used to replace the competing ions with other ions that produce a solution with a much lower conductivity. For example, a system for the analysis of cations would use a cation-exchange analytical column and an anion-exchange suppressor column. A similar strategy can be applied to the analysis of anions by using an anion-exchange analytical column followed by a cation-exchange suppressor column [3, 5].

66.7 SIZE-EXCLUSION CHROMATOGRAPHY

SEC is a type of LC in which the separation of solutes is based on their size [1–5]. In this method, a support with a range of pore sizes is used, in which the pores approach the sizes of the solutes to be separated. Solute separation is based on their ability to enter various fractions of these pores. Large solutes will be able to enter none or only a few of these pores and will spend most of their time in the mobile phase that is freely flowing outside of the support and through the column. Smaller solutes will be able to enter most or all of the pores and will take longer to pass through the column. The result is a separation based on the size, shape, and molar mass of these solutes [3–5, 34].

In SEC, the volume of the mobile phase that occupies the region outside of the pores of the support is referred to as the excluded volume, V_E . The total volume of mobile phase that is present in both the excluded volume and within the pores of the support is represented by the void volume of the column, V_M . If no other interactions are present between the injected solutes and the support, the retention volume (V_R) for a solute in SEC should be at or between the values for V_E and V_M . Small solutes will elute with a value for V_R that approaches or is equal to the void volume, and large solutes will have a value for V_R that approaches or is equal to the excluded volume. Solute retention volumes will have values between V_E and V_M [3–5].

The porous support that is used in SEC should have pore diameters that are in the same general range as the sizes of the solutes that are to be separated. This support

should also be inert and not interact directly with the solutes if the separation is to be based only on size. Many of the supports that were discussed in the previous sections can be used in SEC. Biological compounds and aqueous-based samples are usually separated by SEC through the use of agarose, dextrose, and other carbohydrate-based supports. Polystyrene and diol-bonded silica may also be used for work in SEC with samples in organic solvents or aqueous solutions, respectively [3, 33, 34].

Because there are ideally no interactions of the solutes with the support and there is no true stationary phase in SEC, there is also no weak or strong mobile phase in this method. The selection of the mobile phase in SEC is instead made based on the solubility of the chemicals that are to be separated and the stability or properties of the support. Either polar or nonpolar solvents can be utilized as mobile phases in SEC. If the mobile phase is an organic solvent, the term *gel permeation chromatography* is often used to describe the resulting SEC method. *Gel filtration chromatography* is the term used to refer to a SEC technique in which the mobile phase is water or an aqueous solution [1, 2, 5].

SEC has both preparative and analytical applications. As an example, gel filtration chromatography is often used to purify biological samples, such as the removal of small solutes from large biological molecules like proteins. SEC can also be used to characterize the molar mass or distribution in mass for a solute or group of solutes. To determine the molar mass of a solute, standards that are similar to the solute of interest but that have known masses are first injected onto an SEC column. This group of solutes should have a size range that includes some standards that can enter all or most of the pores of the support, some that are completely excluded from the pores, and several that can access intermediate volumes. A plot is then made of the logarithm of the molar mass of each standard versus the solute's measured retention volume, retention time, or some related measure of retention (see Fig. 66.5). This plot can then be used to determine the molar masses of other similar solutes that are injected onto the same column [3, 5].

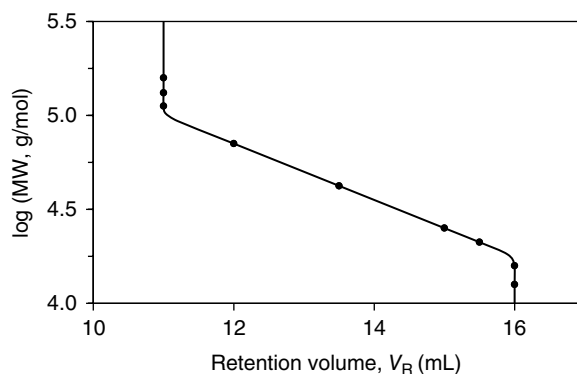


FIGURE 66.5 An example of calibration curve for determination of the molecular weight (MW) of solutes based on size-exclusion chromatography.

66.8 AFFINITY CHROMATOGRAPHY

Affinity chromatography is a type of LC in which solutes are separated based on their binding to a stationary phase that is a biologically related agent [2, 5, 35]. Because of the strong and selective nature of many biological interactions, this method can be a powerful technique for the purification and analysis of solutes that are complementary to the immobilized binding agent. This immobilized binding agent is known as the affinity ligand [35, 36]. The affinity ligand often interacts with its target solute through several interactions, such as dipole–dipole interactions, hydrogen bonding, van der Waals forces, and ionic interactions. The fit of the target at the affinity ligand's binding site may also involve steric effects. The overall result of these various interactions is selective and strong, but usually reversible, binding between the target and the affinity ligand [35, 36].

Figure 66.6 shows a common format for using affinity chromatography [35, 36]. First, a sample containing the target solute is injected or applied in the presence of an application buffer that promotes binding by the target to the immobilized affinity ligand. During this step, the target will be bound by the column while other sample components will tend to be washed away. If the target has strong binding to the affinity ligand, an elution buffer can later be passed through the column to release the bound target. The released target can then be detected or collected for further use. Once the target has eluted, the application buffer can be passed again through the column, and the affinity ligand is allowed to regenerate prior to the next application or injection of a sample [36].

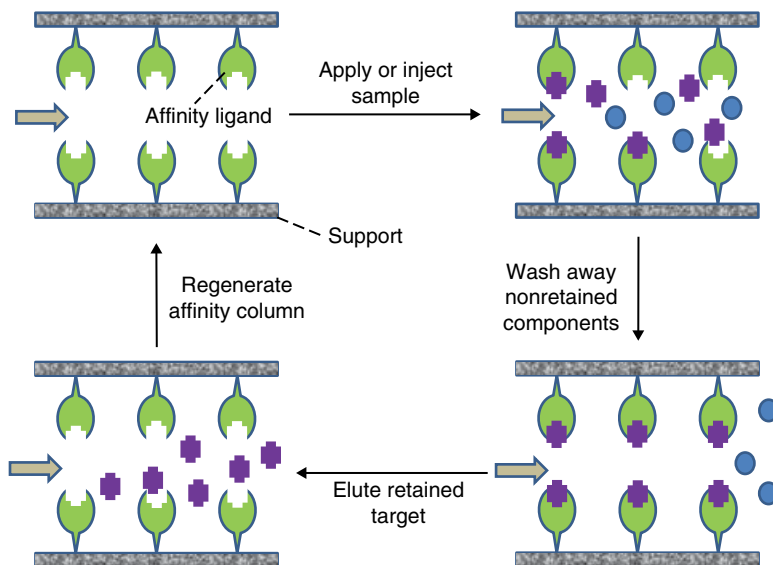


FIGURE 66.6 A typical on/off elution scheme used in affinity chromatography.

The affinity ligand that is used as the stationary phase in affinity chromatography plays a key role in determining which solutes will be retained by the column. This ligand may be a biological agent, a mimic of a biological agent, or even a synthetic compound [35–38]. All of these binding agents can be classified as being either a *high-specificity ligand* or a *general* (or *group specific*) *ligand* [35, 36]. A high-specificity ligand tends to bind to only one target solute or a group of closely related solutes. Common examples of high-specificity ligands are antibodies (which can bind to their corresponding antigens) and enzymes (which can bind to their substrates, cofactors, or inhibitors). A general ligand tends to bind to a set of targets that have a common feature in their structures. Examples of general ligands are immunoglobulin-binding proteins like protein A or protein G, lectins (i.e., nonimmune system proteins that bind sugar residues), boronates, biomimetic dyes, and metal ion chelates [35–38].

An affinity ligand can be immobilized onto the support in various ways. The most common way for immobilizing an affinity ligand is to covalently attach it to the support. This process uses an activated support that can react with functional groups on the affinity ligand, such as amine groups, carboxylic acids, sulfhydryl residues, or aldehyde groups [35–38]. Some affinity ligands can be immobilized through noncovalent adsorption. This approach is often used with antibodies by adsorbing these affinity ligands to immobilized protein A or protein G, and it is used in the adsorption of biotin-tagged affinity ligands on supports that contain immobilized avidin or streptavidin [35, 38]. Another technique for immobilizing large affinity ligands such as liposomes, cells, or proteins is to entrap or encapsulate these binding agents within the support [35–41]. In addition, synthetic affinity ligands against a given target can be generated during the preparation of some polymeric supports in a process known as *molecular imprinting* [35, 42, 43].

The mobile phase will also affect the degree to which a target solute will bind to the affinity ligand. The application buffer is the weak mobile phase in affinity chromatography and is typically a buffer or solution that mimics the natural conditions under which the affinity ligand binds to its target [35, 36]. The strong mobile phase in affinity chromatography is the elution buffer. This buffer may involve a change in the pH, ionic strength, or polarity to weaken the binding of the affinity ligand with its target, giving an approach known as *nonspecific elution*. Another approach for elution is to add an agent to the mobile phase that competes with the target for the affinity ligand or with the affinity ligand for the target. This second approach is known as *biospecific elution* [35, 36]. If the target has sufficiently weak binding to the affinity ligand in the presence of the application buffer, it is possible to even use isocratic conditions to elute the target from the column. This last situation tends to occur with systems that have association equilibrium constants of 10^5 – 10^6 M⁻¹ or less and gives a method known as *weak affinity chromatography* [35].

Probably the most common application for affinity chromatography is the selective purification and isolation of biological compounds [35–38]. This type of application includes the large-scale purification of biopharmaceuticals and enzymes, such as by

using biomimetic dyes or immobilized enzyme inhibitors as the affinity ligands [35, 37, 44]. This also includes the isolation of recombinant histidine-tagged proteins through the use of *immobilized metal ion affinity chromatography (IMAC)* [35, 45, 46]. Other examples are the use of antibodies as ligands for the isolation of various specific targets in an approach known as *immunoaffinity chromatography (IAC)* and the use of lectins to isolate glycoproteins or other carbohydrate-containing targets [35, 38, 47, 48].

There are a variety of analytical applications for affinity chromatography. For instance, many chiral separations are based on the binding of drugs or other chiral solutes to stereoselective affinity ligands such as cyclodextrins, enzymes, or serum transport proteins [35, 49, 50]. Antibodies and immunoaffinity columns have also been used in various formats to analyze specific solutes in a wide range of samples by using chromatographic-based immunoassays [35, 47, 48]. In addition, affinity columns containing ligands such as antibodies and lectins have been combined in both online and offline extraction formats with techniques like RPLC and mass spectrometry for chemical analysis. Columns containing antibodies have further been used to remove possible interfering compounds from samples in a method known as *immunodepletion*. This last approach has been used in proteomics to remove major proteins from samples and to make it easier to detect or measure less abundant proteins [48].

Affinity chromatography can be further used to study biological interactions. This approach is sometimes known as *analytical affinity chromatography* or *biointeraction affinity chromatography* [35, 51, 52]. A variety of data on a biological interaction can be generated through affinity chromatography, such as the strength or rate of the reaction and the number of binding sites that are involved. Information on the location of interaction sites and the binding strength of a target at specific sites on an affinity ligand can also be acquired through this approach [35, 51, 52]. Examples of systems that have been examined by this method are sugar/lectin interactions, protein/protein interactions, and drug/protein interactions [35, 51, 52].

66.9 DETECTORS FOR LIQUID CHROMATOGRAPHY

Many types of detectors can be used to monitor solutes as they elute from modern LC systems. One general detector that can be used is a *refractive index (RI) detector* [5, 6, 9, 34]. This detector responds to the change in RI that occurs in the mobile phase as solutes elute from the column. An RI detector can detect the presence of a solute by comparing the RI of the mobile/solute mixture to a reference stream or portion of the mobile phase with no solutes present. An RI detector has a moderate limit of detection when compared to other detectors for LC, as shown in Table 66.1. However, this type of detector is valuable for work with solutes that have an unknown composition or that may not contain a chromophore or fluorophore (e.g., as is the case for many carbohydrates and lipids). The response of an RI detector can be affected by changes in temperature,

TABLE 66.1 Common LC Detectors^a

Detector Type	Compounds Detected	Detection Limits
Refractive index detector	Universal—all compounds	0.1–1 µg
UV/Vis absorbance detector	Compounds with chromophores	0.1–1 ng
Evaporative light scattering detector	Nonvolatile compounds	10 µg
Conductivity detector	Ionic compounds	0.5–1 ng
Fluorescence detector	Fluorescent compounds	1–10 pg
Electrochemical detector	Electrochemically active compounds	0.01–1 ng
Mass spectrometry	Universal (full scan mode) Selective (selected ion monitoring mode)	0.1–1 ng

^aThis table is based on information provided in Refs. [5, 25] and obtained from manufacturers of these detectors. The concentration limits of detection for these devices can be estimated by dividing the above mass values by 10–100 µl, which is a typical range of sample injection volumes that are used in HPLC.

so it is important to keep this device at a constant temperature. A change in mobile phase composition, as occurs during gradient elution, may also cause a change in the background response of an RI detector if the reference solution does not undergo similar changes in composition [5, 6, 9].

Ultraviolet/visible (UV/Vis) absorbance detectors are also commonly found on modern LC systems [5, 6, 9]. These detectors measure the ability of eluting solutes to absorb either UV or visible light. One device in this category is a *fixed wavelength absorbance detector*, which always monitors a specific wavelength (e.g., 254 nm, where many organic compounds with aromatic groups or unsaturated bonds absorb light). Another device in this group is a *variable wavelength absorbance detector*, which can allow absorbance to be measured at a wavelength that is selected from a relatively broad range (e.g., 190–900 nm). A third type of device is a *photodiode array detector (PDA)*, which uses a detector array to simultaneously measure absorbance at many wavelengths. All of these detectors require that some type of chromophore be present on the solute or that a chromophore first be added to the solute through derivatization. These detectors tend to provide much better limits of detection than an RI detector for such solutes, with detection limits usually in the 10⁻⁸ M range. Absorbance detectors are also easy to use with gradient elution as long as the weak and strong mobile phases do not have significant absorption of light at the wavelength that is being monitored [5, 6, 9, 34].

Another type of general detector for LC is the *evaporative light scattering detector (ELSD)* [5, 9]. This device can be used to monitor solutes that are less volatile than the mobile phase. In this detector, the mobile phase/solute mixture is converted into a spray of small droplets. The solvent is next evaporated away, leaving behind small particles of the less volatile sample components. These particles are detected by examining their ability to scatter a beam of light, where the degree of light scattering is related to both the size and amount of solute particles that are present. An ELSD has a better limit of

detection than an RI detector and is easier to use with gradient elution. In addition, the solutes that are detected by an ELSD do not have to have any chromophore present, which is an advantage of this device over absorbance detectors [5, 9].

A *conductivity detector* is a detector that can be used in LC to monitor ionic solutes. These solutes are detected by measuring the ability of the mobile phase and its ionic contents to conduct a current when this mixture is placed in an electrical field. As was mentioned earlier, a conductivity detector is used in ion chromatography to aid in the analysis of ions in various samples. A conductivity detector can be used with gradient elution; however, the background signal will change if the ionic strength (and possibly pH) of the mobile phase are not kept constant. It is also necessary with this type of device to keep the overall ionic composition and background conductance of the mobile phase reasonably low so that eluting solute ions can be detected.

A more specific detector that can be used in LC is a *fluorescence detector* [3, 5, 6]. This type of detector looks at the ability of eluting solutes to both absorb and emit light at a given set of wavelengths. This set of wavelengths allows for more selective detection than occurs when using absorbance measurements and allows for a lower background signal, which provides improved limits of detection. This type of device can be used to look at solutes that are naturally fluorescent or that have been converted into a fluorescent derivative. A fluorescence detector can be used with gradient elution but does require relatively pure mobile phases to help maintain good fluorescence signals and a low background response [3, 5, 6, 9].

Electrochemical detectors can also be used in LC, giving a method referred to as *liquid chromatography/electrochemical detection (LC-EC)* [3, 5, 6]. This type of detector is used to monitor solutes that can undergo oxidation or reduction (e.g., aldehydes, ketones, phenols, mercaptans, peroxides, and some carbohydrates). Many of these detectors work by measuring the amount of current that is generated as a solute is oxidized or reduced at a given potential, although other formats are possible as well. The response in this case will depend on the amount of oxidation or reduction that is taking place, which affects the size of the measured current. This type of electrochemical detector can have a quite low limit of detection because of the accuracy and precision with which currents can be measured. Electrochemical detectors and LC/EC can also be used with gradient elution [3, 5, 6, 9, 34].

Liquid chromatography/mass spectrometry (LC/MS) is another important combination that is popular in the separation and analysis of chemicals by LC [3, 5, 6, 9]. This method uses mass spectrometry and mass measurements of ions to measure and identify chemicals based on the molecular ions or fragment ions that are generated for these chemicals. Such a system can be used to look at all or most of the ions that are produced in a "full-scan mode," which results in a general detection method. Alternatively, only a few ions that are characteristic of a particular solute or set of solutes can be examined through "selected ion monitoring," providing a more selective method for detection [3, 5, 9].

LC/MS is often carried out by using electrospray ionization (ESI) to generate ions from the eluting solutes [5, 6, 53]. These ions are then examined and separated based

on their mass-to-charge ratios by using a quadrupole mass analyzer or other type of mass analyzer. ESI can be used in LC/MS to examine substances that range from small polar compounds to proteins. ESI and LC/MS can also be utilized with gradient elution. The use of LC/MS with ESI is particularly useful in work with proteins and peptides, which tend to give ions with high mass-to-charge ratios when examined by other ionization methods. In ESI, many charges are often placed on one protein or peptide, which provides ions with reasonably low mass-to-charge ratios that are easier to measure by common mass analyzers [5, 6, 53].

66.10 OTHER COMPONENTS OF LC SYSTEMS

In addition to the detector, other important components of LC instruments are the pumps and columns. A pump for HPLC should be able to generate pressures up to 5000–6000 psi and flow rates ranging from 0.1 to 10 mL/min with good reproducibility [3, 5, 9, 53]. Even higher pressures are required in UPLC [5, 14]. The two types of pumps that are used in most LC systems are reciprocating pumps and syringe pumps. A reciprocating pump contains a chamber in which a rotating cam causes a piston to move back and forth. Mobile phase is pumped into and out of this chamber by the movement of the piston, and the direction of this flow is set through the use of check valves. A reciprocating pump works well in the milliliter per minute flow rate range and is easy to use with gradient elution. When a lower flow rate or a smaller column is to be used, a syringe pump is usually employed. In this device, a syringe produces flow of the mobile phase by this liquid or solution being passed out of a chamber as a plunger is slowly depressed into the chamber. A syringe pump can achieve constant flow rates in the microliter per minute range. However, this pump is also less convenient to use than a reciprocating pump in work with large volumes of mobile phase or when changing from one mobile to another during gradient elution [3, 5, 9].

There are various column dimensions and flow rates that are used in LC [3, 5, 9, 25]. In HPLC, the columns usually have a length of 5–30 cm and an internal diameter of 4.1 or 4.6 mm. Such columns are used at typical flow rates of 1–3 mL/min [5, 25]. Longer columns or capillaries with smaller internal diameters, such as microbore columns and packed capillaries, are often needed for separations requiring higher efficiencies. For instance, microbore columns often have lengths of 10–100 cm, internal diameters of 1–2 mm, and are used at flow rates of 0.05–0.2 mL/min [25]. These longer and narrower columns also require smaller sample volumes, to avoid overloading the column, while the use of lower flow rates provides a reasonable operating pressure. These latter features can be advantageous in that they can allow work with small amounts of samples. The use of low flow rates is especially useful in LC/MS in that it means less solvent must be removed from solutes during the ionization process and before the resulting ions can be examined by the mass spectrometer [3, 5, 9, 53].

ACKNOWLEDGEMENTS

Portions of this work were supported by the University of Nebraska–Lincoln (UNL) UCARE program, the NSF (under the REU program and grant CMI 1309806), the NSF/EPSCoR program (grant EPS-1004094), and the NIH (grants R01 GM044931 and R01 DK069629). Additional support for R. Matsuda was obtained through a fellowship from the Molecular Mechanisms of Disease Program at UNL.

REFERENCES

1. R.E. Majors and P.W. Carr, “Glossary of liquid-phase separation terms”, *LC-GC*, 19 (2001) 124–162.
2. J. Inczedy, T. Lengyel, and A.M. Ure, *International Union of Pure and Applied Chemistry-Compendium of Analytical Nomenclature: Definitive Rules*, Blackwell Science, Malden, 1997, Chapter 9.
3. C.F. Poole and S.K. Poole, *Chromatography Today*, Elsevier, New York, 1991.
4. B.L. Karger, L.R. Snyder, and C. Horvath, *An Introduction to Separation Science*, John Wiley & Sons, New York, 1973.
5. D.S. Hage and J.D. Carr, *Analytical Chemistry and Quantitative Analysis*, Pearson Prentice Hall, Upper Saddle River, 2011, Chapter 22.
6. D.A. Skoog, F.J. Holler, and T.A. Nieman, *Principles of Instrumental Analysis*, 5th Ed., Brooks Cole, Boston, 1998, Chapter 28.
7. L.S. Ettre, “M.S. Tswett and the invention of chromatography”, *LC-GC*, 21 (2003) 458–467.
8. M.A. Nelson, A. Gates, M. Dodlinger, and D.S. Hage, “Development of a portable immunoextraction/RPLC system for field studies of herbicide residues”, *Anal. Chem.*, 76 (2004) 805–813.
9. W.J. Lough and I.W. Wainer, *High Performance Liquid Chromatography: Fundamentals Principles and Practice*, Blackie Academic, New York, 1995.
10. K.K. Unger, R. Ditz, E. Machtejevas, and R. Skudas, “Liquid chromatography—its development and key role in life science applications”, *Angew. Chem. Int. Ed.*, 49 (2010) 2300–2312.
11. R.E. Majors, “Effect of particle size on column efficiency in liquid-solid chromatography”, *J. Chromatogr. Sci.*, 11 (1973) 88–95.
12. R.E. Majors, “A review of HPLC column packing technology”, *Am. Lab.*, 35 (2003) 46–54.
13. L.S. Ettre, “Csaba Horvath and the development of the first modern high performance liquid chromatograph”, *LC-GC*, 5 (2005) 85–90.
14. J.W. Thompson, J.S. Mellors, J.W. Eschelbach, and J.W. Jorgenson, “Recent advances in ultrahigh-pressure liquid chromatography”, *LC-GC*, 24 (2006) 16–20.
15. J.E. McNair, K.C. Lewis, and J.W. Jorgenson, “Ultrahigh-pressure reversed-phase liquid chromatography in packed capillary columns”, *Anal. Chem.*, 69 (1997) 983–989.

16. F. Svec and C.G. Huber, "Monolithic materials: promises, challenges, and achievements", *Anal. Chem.*, 78 (2006) 2100–2108.
17. G. Guiochon, "Monolithic columns in high-performance liquid chromatography", *J. Chromatogr. A*, 1168 (2007) 101–168.
18. F. Svec, "Porous polymer monoliths: amazingly wide variety of techniques enabling their preparation", *J. Chromatogr. A*, 1217 (2010) 902–924.
19. G. Guiochon and F. Gritti, "Shell particles, trials, tribulations and triumphs", *J. Chromatogr. A*, 1218 (2011) 1915–1938.
20. R. Hayes, A. Ahmed, T. Edge, and H. Zhang, "Core-shell particles: preparation, fundamentals and applications in high performance liquid chromatography", *J. Chromatogr. A*, 1357 (2014) 36–52.
21. R.W. Brice, X. Zhang, and L.A. Colón, "Fused-core, sub-2 microm packings, and monolithic HPLC columns: a comparative evaluation", *J. Sep. Sci.*, 32 (2009) 2723–2731.
22. J.J. Kirkland, F.A. Truszkowski, C.H. Dilks, Jr., and G.S. Engel, "Superficially porous silica microspheres for fast high-performance liquid chromatography of macromolecules", *J. Chromatogr. A*, 890 (2000) 3–13.
23. A. Braithwaite and F.J. Smith, *Chromatographic Methods*, 4th Ed., Chapman & Hall, New York, 1990.
24. R.B. Patel, M.R. Patel, and B.G. Patel, "Experimental aspects and implementation of HPTLC", in M.M. Srivastava (Ed.), *High Performance Thin Layer Chromatography (HPTLC)*, Springer, Germany, 2011, pp. 41–54.
25. S.K. Poole and C.F. Poole, "High performance stationary phases for planar chromatography", *J. Chromatogr. A*, 1218 (2011) 2648–2660.
26. H.E. Hauck and M. Schulz, "Ultrathin-layer chromatography", *J. Chromatogr. Sci.*, 40 (2002) 550–552.
27. H.E. Hauck, O. Bund, W. Fischer, and M. Schulz, "Ultra-thin layer chromatography (UTLC). A new dimension in thin-layer chromatography", *J. Planar Chromatogr.*, 14 (2001) 234–236.
28. L.R. Snyder and J.J. Kirkland, *Introduction to Modern Liquid Chromatography*, 2nd Ed., John Wiley & Sons, New York, 1979.
29. B. Buszewski and S. Noga, "Hydrophilic interaction liquid chromatography (HILIC)—a powerful separation technique", *Anal. Bioanal. Chem.*, 402 (2012) 231–247.
30. P.J. Boersema, S. Mohammed, and A.J.R. Heck, "Hydrophilic interaction liquid chromatography (HILIC) in proteomics", *Anal. Bioanal. Chem.*, 391 (2008) 151–159.
31. Y. Hsieh, "Potential of HILIC-MS in quantitative bioanalysis of drugs and drug metabolites", *J. Sep. Sci.*, 31 (2008) 1481–1491.
32. K. Novakova and H. Vlckova, "A review of current trends and advances in modern bio-analytical methods: chromatography and sample preparation", *Anal. Chim. Acta*, 656 (2009) 8–35.
33. E. Katz, R. Eksteen, P. Schoenmakers, and N. Miller (Eds.), *Handbook of HPLC*, Marcel Dekker, New York, 1998, Chapter 10.
34. B. Ravindranath, *Principles and Practice of Chromatography*, John Wiley & Sons, New York, 1989.

35. D.S. Hage (Ed.), *Handbook of Affinity Chromatography*, 2nd Ed., CRC Press/Taylor & Francis, New York/Boca Raton, 2006.
36. R.R. Walters, "Affinity chromatography", *Anal. Chem.*, 57 (1985) 1099A–1114A.
37. M. Zachariou (Ed.), *Affinity Chromatography: Methods and Protocols*, 2nd Ed., Humana Press, Totowa, 2010.
38. G.T. Hermanson, A.K. Mallia, and P.K. Smith, *Immobilized Affinity Ligand Techniques*, Academic Press, New York, 1992.
39. Q. Yang and P. Lundahl, "Immobilized proteoliposome affinity chromatography for quantitative analysis of specific interactions between solutes and membrane proteins. Interaction of cytochalasin B and D-glucose with the glucose transporter Glut1", *Biochemistry*, 34 (1995) 7289–7294.
40. C.M. Zeng, Y. Zhang, L. Lu, E. Brekkan, A. Lundqvist, and P. Lundahl, "Immobilization of human red cells in gel particles for chromatographic activity studies of the glucose transporter Glut1", *Biochim. Biophys. Acta*, 1325 (1997) 91–98.
41. A.J. Jackson, J. Anguizola, E.L. Pfaunmiller, and D.S. Hage, "Use of entrapment and high-performance affinity chromatography to compare the binding of drugs and site-specific probes with normal and glycated human serum albumin", *Anal. Bioanal. Chem.*, 405 (2013) 5833–5841.
42. M. Komiyama, T. Takeuchi, T. Mukawa, and H. Asanuma, *Molecular Imprinting from Fundamentals to Applications*, Wiley-VCH, Weinheim, 2002.
43. D. Kriz, O. Ramstrom, and K. Mosbach, "Molecular imprinting. New possibilities for sensor technology", *Anal. Chem.*, 69 (1997) 345A–349A.
44. Y.D. Clonis, N.E. Labrou, V. Kotsira, K. Mazitsos, S. Melissis, and G. Gogolas, "Biomimetic dyes as affinity chromatography tools in enzyme purification", *J. Chromatogr. A*, 891 (2000) 33–44.
45. J. Porath, J. Carlsson, I. Olsson, and B. Belfrage, "Metal chelate affinity chromatography, a new approach to protein fraction", *Nature*, 258 (1975) 598–599.
46. H. Block, B. Maertens, A. Priestersbach, N. Brinker, J. Kubicek, R. Fabis, J. Labahn, and F. Schäfer, "Immobilized-metal affinity chromatography (IMAC): a review", *Methods Enzymol.*, 463 (2009) 439–473.
47. D.S. Hage, "A survey of recent advances in analytical applications of immunoaffinity chromatography", *J. Chromatogr. B*, 715 (1998) 3–28.
48. A.C. Moser and D.S. Hage, "Immunoaffinity chromatography: an introduction to applications and recent developments", *Bioanalysis*, 2 (2010) 769–790.
49. W.J. Lough (Ed.), *Chiral Liquid Chromatography*, Blackie and Son, Glasgow, 1989.
50. S. Allenmark, *Chromatographic Enantioseparations: Methods and Applications*, 2nd Ed., Ellis Horwood, New York, 1991.
51. I.M. Chaiken (Ed.), *Analytical Affinity Chromatography*, CRC Press, Boca Raton, 1987.
52. J.E. Schiel, K.S. Joseph, and D.S. Hage, "Biointeraction affinity chromatography: general principles and recent developments", *Adv. Chromatogr.*, 48 (2009) 146–187.
53. G.W. Ewing (Ed.), *Analytical Instrumentation Handbook*, 2nd Ed., Marcel Dekker, New York, 1997.

MASS SPECTROSCOPY MEASUREMENTS OF NITROTYROSINE-CONTAINING PROTEINS

XIANQUAN ZHAN¹ AND DOMINIC M. DESIDERIO²

¹Key Laboratory of Cancer Proteomics of Chinese Ministry of Health, Xiangya Hospital, Central South University, Changsha, P. R. China

²The Charles B. Stout Neuroscience Mass Spectrometry Laboratory, Department of Neurology, College of Medicine, University of Tennessee Health Science Center, Memphis, TN, USA

67.1 INTRODUCTION

67.1.1 Formation, Chemical Properties, and Related Nomenclature of Tyrosine Nitration

Tyrosine nitration is a posttranslational modification (PTM) derived from oxidative/nitrative stress, which results from *in vivo* nitrating agents such as peroxynitrite (ONOO⁻) and nitrogen dioxide [1–4]. Tyrosine nitration is the addition of a nitro group (–NO₂), an electron-withdrawing group, to position 3 of the phenolic ring of a tyrosine residue to form a 3-nitrotyrosine residue in a protein [3]. Nitration of a tyrosine residue significantly changes its physical and chemical properties [3, 5, 6]; for example, (i) the pK_a value of the phenolic hydroxyl group of nitrotyrosine (pK_a ≈ 7.1) is significantly decreased relative to that of tyrosine (pK_a ≈ 10), (ii) the electron density of the phenolic ring of nitrotyrosine is significantly decreased relative to that of tyrosine because the nitro group (–NO₂) is an electron-withdrawing group, and (iii) nitrotyrosine can be reduced to aminotyrosine. Based on nitration of a tyrosine residue, some related nomenclatures were derived: a nitropeptide is defined as a nitrotyrosine-containing peptide or a nitrated peptide; a nitroprotein is defined as a nitrotyrosine-containing protein or a

nitrated protein; a nitroproteome is defined as all nitroproteins in a proteome; and nitroproteomics is defined as the use of proteomics to study the nitroproteome.

67.1.2 Biological Roles of Tyrosine Nitration in a Protein

Tyrosine nitration in a protein in a biological system has extensive biological functions. First, because the nitro group ($-\text{NO}_2$) is an electron-withdrawing group, tyrosine nitration, namely, addition of nitro group to the phenolic ring of the tyrosine residue, decreases the electron density of the phenolic ring of a tyrosine residue in a protein [3, 7]. Nitration yields several direct biological consequences: (i) it shifts the phenolic pK_a value (from ~ 10 for tyrosine) into the physiological pH range (~ 7.1 for 3-nitrotyrosine) to affect chemical properties of a tyrosine residue [3, 6, 8] and (ii) if the tyrosine nitration occurs exactly within an interacting region between a receptor and ligand or between an enzyme and substrate, then the decreased electron density could negatively affect interaction intensity (receptor–ligand and enzyme–substrate) to hinder the functions of that protein [3, 7]. Second, nitration and phosphorylation would compete for the same tyrosine residue when the tyrosine residue is within a tyrosine phosphorylation motif ([R or K]-x2(3)-[D or E]-x3(2)-[Y]) to affect tyrosine phosphorylation signaling pathways and involve important biological processes [7, 9–11]. Third, some studies have demonstrated that tyrosine nitration in a protein in a biological system might be a reversible and dynamic process between nitration and denitration because of the discovery of a putative denitrase [8, 12, 13]. Therefore, tyrosine nitration is not only a pathological consequence, a marker of oxidative injuries, but also involve multiple biological processes such as neurotransmission and redox signaling [4]. Tyrosine nitration can alter protein functions and associated multiple physiological or pathological processes such as tumorigenesis, inflammatory and neurodegenerative diseases, modification of enzymatic activities, and immunogenicity [2, 3, 14–17].

67.1.3 Challenge and Strategies to Identify a Nitroprotein with Mass Spectrometry

In order to understand the biological functions and roles of tyrosine nitration in a protein, an essential step is to identify endogenous nitroproteins and accurately locate each nitrotyrosine site. Mass spectrometry (MS) is the key technique for these tasks. However, MS identification of endogenous nitroproteins and nitrotyrosine sites is severely challenged because of several factors: (i) endogenous nitroproteins occur with an extreme low abundance (1 in $\sim 10^6$ tyrosines), (ii) each MS instrument has its sensitivity limitation that requires a sufficient amount of samples for MS detection, and (iii) various MS behaviors are present in different types of MS analyses; for example, there is a characteristic photodecomposition pattern of a nitro group that is present in UV-laser-based matrix-assisted laser desorption ionization (MALDI)-MS

analysis of a nitroprotein [18–20] but not for electrospray ionization (ESI)-MS [17, 20–23]. As a result, photodecomposition of a nitro group decreases signal intensities of a nitropeptide and complicates interpretation of a mass spectrum; in turn, that characteristic photodecomposition pattern can confirm the existence of a nitro group in a peptide [20].

Several strategies must be developed and used before the use of MS to identify endogenous nitroproteins and nitrotyrosine sites: (i) different chemical derivation techniques [20] are used to convert the nitro group (with various MS behaviors) to an amino group (with stable MS behavior) to resolve the varied MS behaviors of a nitro group and (ii) different enrichment techniques [3, 4, 24] are used to preferentially enrich endogenous nitropeptides or nitroproteins to overcome the extreme low abundance of endogenous nitropeptides/nitroproteins in a biological system and sensitivity limitations of a mass spectrometer. Currently, chemical derivation and targeted enrichment prior to an MS analysis [7, 25, 26] mainly include the following: (i) Nitrotyrosine antibody-based immunoaffinity is used to preferentially enrich nitropeptides [27] or nitroproteins [3, 28]. (ii) Conversion of a nitro group to an amino group is coupled with target enrichment [29]. Briefly, all amines are first acetylated, followed by conversion of nitrotyrosine to aminotyrosine and biotinylation of aminotyrosine. (iii) Conversion of a nitro group to an amino group is coupled with derivatization of the amino group [30]. Briefly, alpha- and epsilon-amino groups in a protein or peptide are protected with $^{13}\text{C}_0/^{13}\text{C}_4$ - or D_0/D_6 -acetic anhydride, nitrotyrosine is reduced to aminotyrosine with sodium dithionite (also known as sodium hydrosulfite), and aminotyrosine is derivatized with 1-(6-methyl[D_0/D_3] nicotinoyloxy) succinimide. (iv) The nitro group in a nitropeptide is reduced to an amino group and dansylated with dansyl chloride, followed by MSⁿ analysis [31, 32]. (v) The “light”- and “heavy”-labeled acetyl groups are used to block *N*-terminal and lysine residues of tryptic nitropeptides, followed by reduction of nitrotyrosine to aminotyrosine with sodium dithionite and derivatization of light- and heavy-labeled aminotyrosine peptides with either isobaric tags for relative and absolute quantification (iTRAQ) or tandem mass tags (TMT), respectively [33]. (vi) Selective chemoprecipitation and subsequent release of tagged species (conversion of nitro group to a small 4-formylbenzylamido tag) are used to analyze nitropeptides with liquid chromatography–tandem mass spectrometry (LC-MS/MS) [34]. (vii) iTRAQ quantitative reagents are used to selectively label nitrotyrosine residues (not primary amines) followed by MS analysis [35]. And (viii) use of combined fractional diagonal chromatography (COFRADIC) [36, 37] peptide sorting is based on a hydrophilic shift after reduction of the nitro group to its amino counterpart, followed by ESI-MS [36] and MALDI-MS [37] identification of a nitropeptide. Moreover, except for proteomics with antinitrotyrosine antibodies and gel-based separation, multidimensional chromatography, precursor-ion scanning, and/or chemical derivatization have also emerged to identify and quantify nitroproteins and nitrotyrosine sites [26, 38].

67.1.4 Biological Significance Measurement of Nitroproteins

MS measurement of nitroproteins must finally serve for real application in a biological system. In order to achieve that goal, several biological significance-related measurements of nitroproteins are worth further study: (i) Quantitative proteomics strategies such as iTRAQ-based quantification should be developed and used to quantify a nitroprotein specific to a pathological condition (or called quantitative nitroproteomics), and the degree of nitration should be quantified in a specific biological condition [33]. (ii) Bioinformatics should be developed and used to locate nitrotyrosine sites within corresponding protein domains and motifs [3, 10] in order to understand in depth the effects of tyrosine nitration on the structure and functions of a protein. (iii) Systems biology methods should be developed and used to elucidate protein-network systems that are involved in nitroproteins [7, 39] for clarification of effects of tyrosine nitration on protein-network systems in a biological condition. (iv) Structural biology should be used to reveal three-dimensional (3D) crystal structure and local primary structure that occur with tyrosine nitration of every biologically important nitroprotein to address the impact of tyrosine nitration on that protein's functions to develop a drug against tyrosine nitration [7, 40, 41]. And (v) body fluids such as serum and cerebrospinal fluid (CSF) are important window to predict and diagnose a disease; body fluid nitroproteomics and nitropeptidomics should be developed and used to discover body fluid biomarkers for prediction, diagnosis, and prognosis of a disease [7, 42].

67.2 MASS SPECTROMETRIC CHARACTERISTICS OF NITROPEPTIDES

67.2.1 MALDI-MS Spectral Characteristics of a Nitropeptide

For UV-laser MALDI-MS analysis of a nitropeptide, high-energy UV-laser light (337 nm) can induce photochemical decomposition of nitro group ($-\text{NO}_2$) to yield a characteristic photochemical decomposition pattern ($[\text{M}+\text{H}]^+$, $[\text{M}+\text{H}-16]^+$, $[\text{M}+\text{H}-30]^+$, and $[\text{M}+\text{H}-32]^+$) in an MS spectrum of a nitropeptide [18, 19, 23]. This photochemical decomposition will decrease the intensity of precursor ion from a nitropeptide and complicate its MS spectrum [20]; meanwhile, recognition of this photochemical decomposition pattern will assist in the interpretation of MS data of a nitropeptide [4, 18–20]. Figure 67.1 shows the formation of a nitrotyrosine and its likely products derived from photochemical decomposition [43].

The nitrotyrosine UV-induced photodecomposition pattern induced by MALDI UV laser has been extensively confirmed in several experiments with a synthetic nitropeptide [A-A-F-G-Y($-\text{NO}_2$)-A-R; $[\text{M}+\text{H}]^+=800.4$] [19], tetranitromethane (TNM)-nitrated bovine serum albumin (BSA) [19], TNM-nitrated angiotensin II ($[\text{M}+\text{H}]^+$, m/z 1092.5)

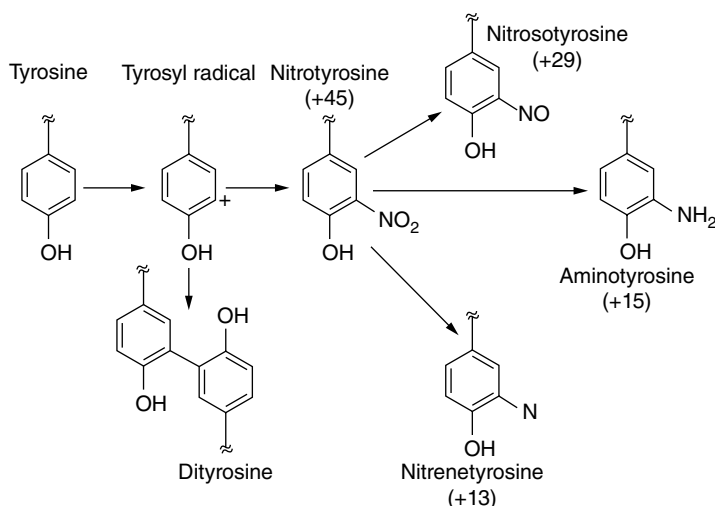


FIGURE 67.1 Formation of dityrosine and nitrotyrosine and photochemical decomposition products of a nitrotyrosine. Source: Turko and Murad [43]. Reproduced with permission of Elsevier.

[18], synthetic leucine enkephalin [LE1: Y-G-G-F-L; molecular weight (MW) = 555.1818 Da], nitro-Tyr-leucine enkephalin [LE2: Y($-\text{NO}_2$)-G-G-F-L, MW = 600.0909 Da], and d_5 -Phe-nitro-Tyr-leucine enkephalin [LE3: Y($-\text{NO}_2$)-G-G-F(d_5)-L, MW = 605.1818 Da] [20], respectively. Here, UV-laser MALDI-MS analysis of synthetic nitropeptide A-A-F-G-Y($-\text{NO}_2$)-A-R ($[\text{M}+\text{H}]^+ = 800.4$) [19] is taken as an example. The MS spectrum shows a typical ion pattern ($[\text{M}+\text{H}]^+$, $[\text{M}+\text{H}-16]^+$, $[\text{M}+\text{H}-14]^+$, $[\text{M}+\text{H}-32]^+$, and $[\text{M}+\text{H}-30]^+$) that corresponded to m/z 800.4, 784.4, 786.4, 768.4, and 770.4, respectively (Fig. 67.2), which results from photochemical decomposition of the nitro group in the $[\text{M}+\text{H}]^+$ ion = 800.4. The $[\text{M}+\text{H}]^+$ ion (m/z 800.4) represents the nitrotyrosine (Tyr- NO_2)-containing peptide; $[\text{M}+\text{H}-16]^+$ ion (m/z 784.4) represents nitrosotyrosine (Tyr-NO)-containing peptide after loss of an oxygen atom from the nitro group; $[\text{M}+\text{H}-14]^+$ ion (m/z 786.4) represents hydroxylaminotyrosine (Tyr-NHOH)-containing peptide after reduction of the nitroso (Tyr-NO) group; $[\text{M}+\text{H}-32]^+$ ion (m/z 768.4) represents triplet nitrenetyrosine (Tyr-N)-containing peptide after loss of two oxygen atoms from the nitro group; and $[\text{M}+\text{H}-30]^+$ ion (m/z 770.4) represents aminotyrosine (Tyr- NH_2)-containing peptide after reduction of the triplet nitrene (Tyr-N) group. Also, intensities of photochemical decomposition products $[\text{M}+\text{H}-32]^+$ and $[\text{M}+\text{H}-30]^+$ ions (Tyr-N and Tyr- NH_2) are much lower than those of the protonated molecule ion $[\text{M}+\text{H}]^+$ (Tyr- NO_2) and photochemical decomposition products $[\text{M}+\text{H}-16]^+$ and $[\text{M}+\text{H}-14]^+$ ions (Tyr-NO and Tyr-NHOH).

Furthermore, MALDI UV-laser-induced photochemical decomposition significantly consumes the protonated molecule ion $[\text{M}+\text{H}]^+$ (Tyr- NO_2) of a nitropeptide to impact detection of endogenous low-abundance nitropeptides/nitroproteins. Figure 67.3 clearly

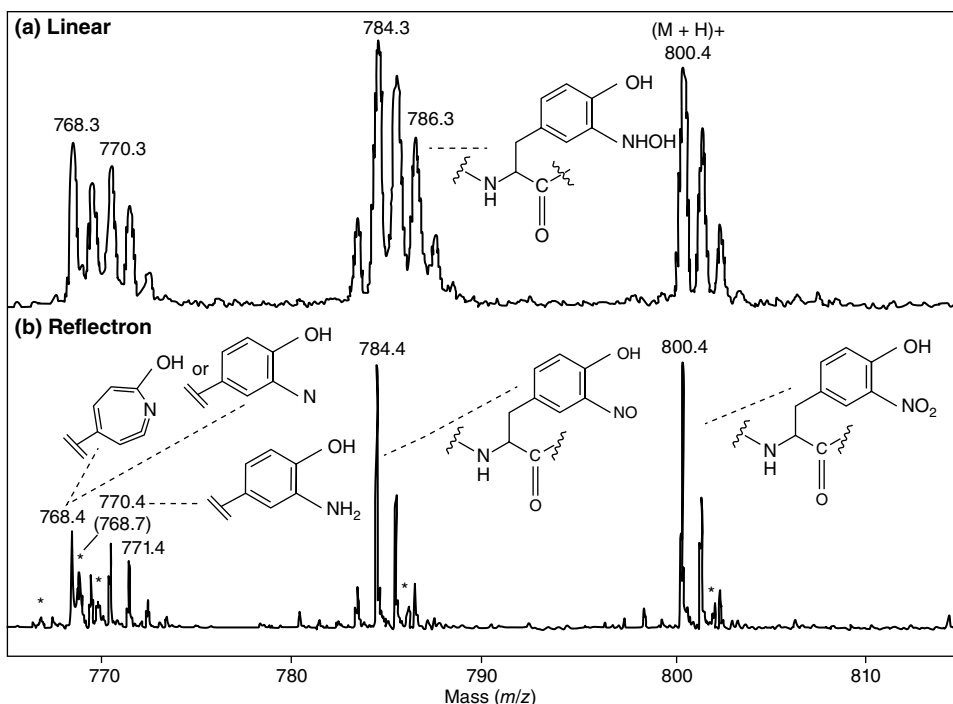


FIGURE 67.2 Photochemical decomposition pattern of synthetic nitropeptide AAFGY($-\text{NO}_2$) AR in a UV-laser MALDI-TOF spectrum in (a) linear mode and (b) reflectron mode. The structures of 3-nitrotyrosine and proposed photochemical decomposition products are shown in the corresponding ions. Several small ions (asterisk) might represent metastable peaks. A weak increase in the abundance of the ion at m/z 771.4 over what would be expected for the ^{13}C isotope peak for the aminotyrosine products at m/z 770.4 in the linear and reflectron spectra suggests that a small amount of a catechol product might have formed as well. Source: Sarver et al. [19]. Reproduced with permission of Springer.

demonstrates that the peak intensity of $[\text{M}+\text{H}]^+$ ion of leucine enkephalin (LE1, $\text{NL}=1.01\text{E}5$) was much higher than that of nitro-Tyr leucine enkephalin (LE2, $\text{NL}=3.25\text{E}4$) and $d(5)$ -Phe-nitro-Tyr leucine enkephalin (LE3, $\text{NL}=9.09\text{E}4$) and that the MS spectrum of nitropeptide (LE2 and LE3) is much more complicated relative to the nonnitrated peptide (LE1) [20].

However, one should note that infrared light-MALDI-Fourier transform ion cyclotron resonance MS (IR-MALDI-FT-ICR-MS) does not fragment the $[\text{M}+\text{H}]^+$ of a nitropeptide to produce an efficient approach to identify protein nitration [44]. The exact reason still remains unknown why MALDI with laser light (337 nm) induces photochemical decompositions of a nitro ($-\text{NO}_2$) group in a nitropeptide (but not with infrared light) and the structures of those photodecomposition products [19]. High-energy UV-laser light (337 nm) might induce loss of one (-16 mass units) or two (-32 mass units) oxygen atoms of the nitro group of a nitropeptide [19].

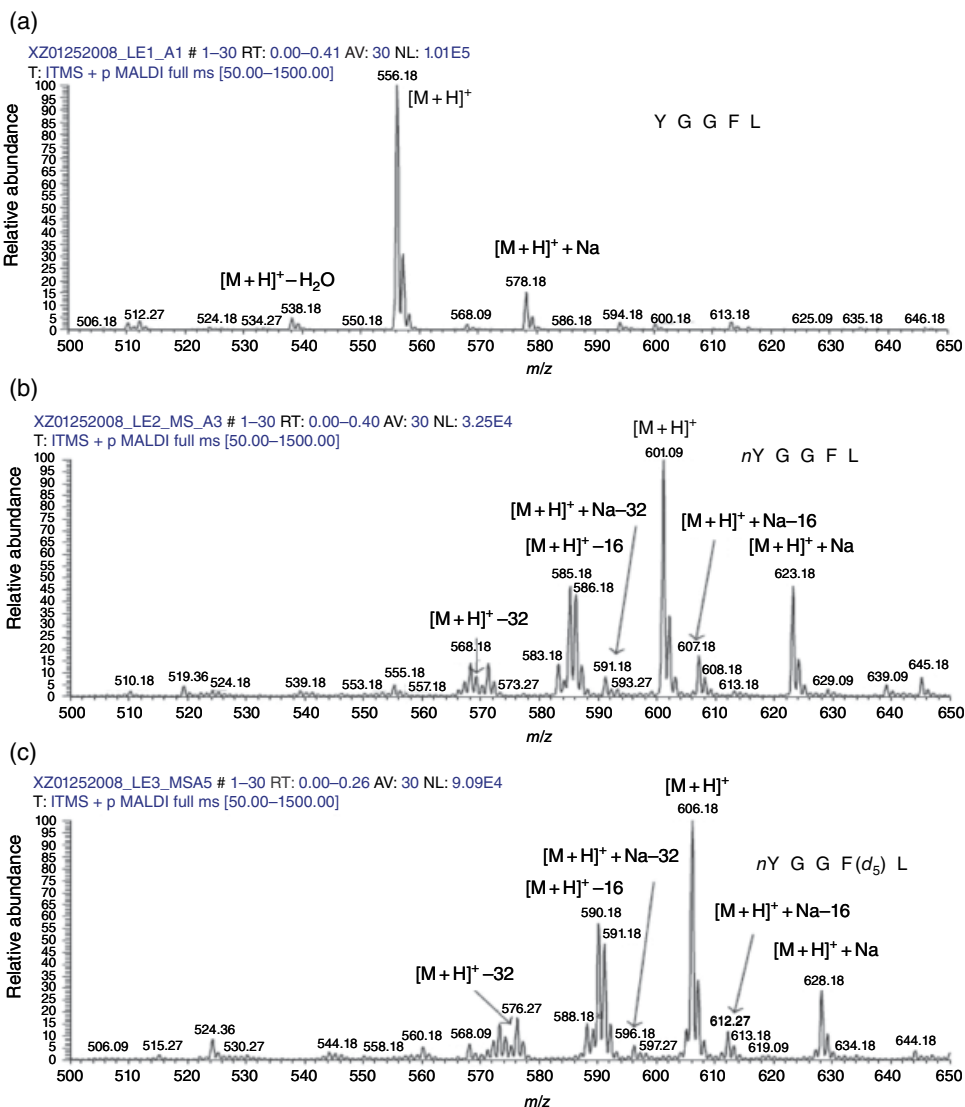


FIGURE 67.3 UV-laser MALDI-MS spectra of LE1 (a), LE2 (b), and LE3 (c). nY = Nitrotyrosine residue. F(d_5) = Phe residue with five 3H (d) atoms. Source: Zhan and Desiderio [20]. Reproduced with permission of Elsevier.

67.2.2 ESI-MS Spectral Characteristics of a Nitropeptide

Compared to the UV-laser MALDI-MS spectrum of a nitropeptide, an ESI-MS spectrum of a nitropeptide does not show decomposition of a nitro group [17–23, 45]. However, an ESI-MS/MS spectrum shows a characteristic immonium ion (m/z 181.06) that is derived from a nitrotyrosine residue to indicate the presence of a nitrotyrosine

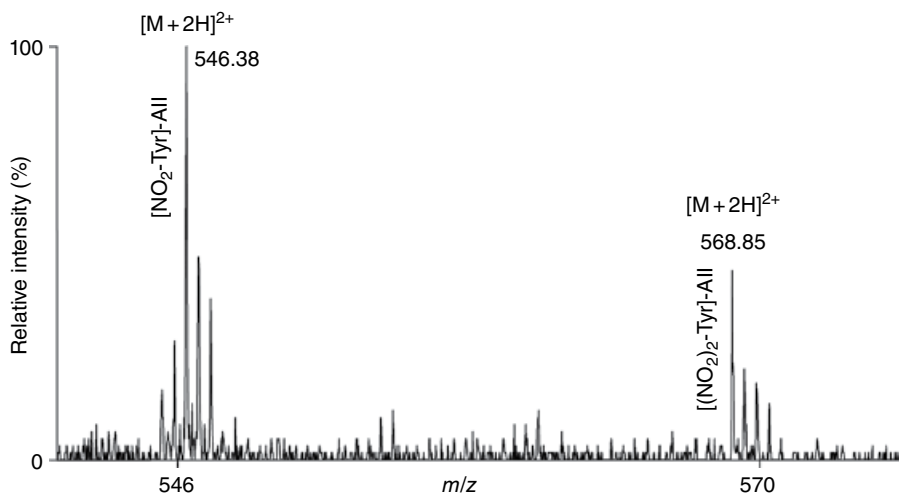


FIGURE 67.4 The ESI-MS spectrum of nitrated angiotensin II to show mononitrated and dinitrated angiotensin II. Source: Petersson et al. [18]. Reproduced with permission of John Wiley & Sons, Inc.

residue. Moreover, precursor-ion scanning based on an immonium ion at m/z 181.06 for nitrotyrosine will accurately identify a nitropeptide/nitroprotein [18].

ESI-MS and ESI-MS/MS spectral characteristics of a nitropeptide and precursor-ion scans for an immonium ion at m/z 181.06 have been confirmed with TNM-nitrated angiotensin II [D-R-V-Y($-\text{NO}_2$)-I-H-P-F; MW = 1090.76 Da] and TNM-nitrated BSA [18]. ESI-MS analysis of TNM-nitrated angiotensin II [D-R-V-Y($-\text{NO}_2$)-I-H-P-F; MW = 1090.76 Da] is taken as an example. First, no chemical composition pattern of a nitro group was found in an ESI-MS spectrum, except for a mononitrated ion ($[\text{M} + 2\text{H}]^{2+}$ m/z 546.38) that represents $[\text{NO}_2\text{-Tyr}]$ -angiotensin II and a dinitrated ion ($[\text{M} + 2\text{H}]^{2+}$ m/z 568.85) that represents $[(\text{NO}_2)_2\text{-Tyr}]$ -angiotensin II (Fig. 67.4). Second, ESI-MS/MS with collision-induced dissociation (CID) fragmentation was used to analyze doubly charged precursor ions of mononitrated angiotensin II at m/z 546.38 and dinitrated angiotensin II at m/z 568.85; characteristic immonium ions occurred at m/z 181.06 (mononitrated tyrosine) and at m/z 226.0 (dinitrated tyrosine) in those ESI-MS/MS spectra (Fig. 67.5). Third, precursor-ion scan spectra based on characteristic immonium ions at m/z 181.06 (mononitrated tyrosine) and m/z 226.0 (dinitrated tyrosine) accurately identified a nitropeptide in a complicated sample (Fig. 67.6).

67.2.3 Optimum Collision Energy for Ion Fragmentation and Detection Sensitivity for a Nitropeptide

The use of UV-laser vMALDI-MS/MS with CID to analyze synthetic peptides LE1 (Y-G-G-F-L; 555.1818 Da), LE2 $[(3\text{-NO}_2)\text{Y-G-G-F-L}]$; 600.0909 Da], and LE3 $[(3\text{-NO}_2)\text{Y-G-G}-(d_5)\text{F-L}]$; 605.1818 Da] found that, first, b- and a-ions were the most

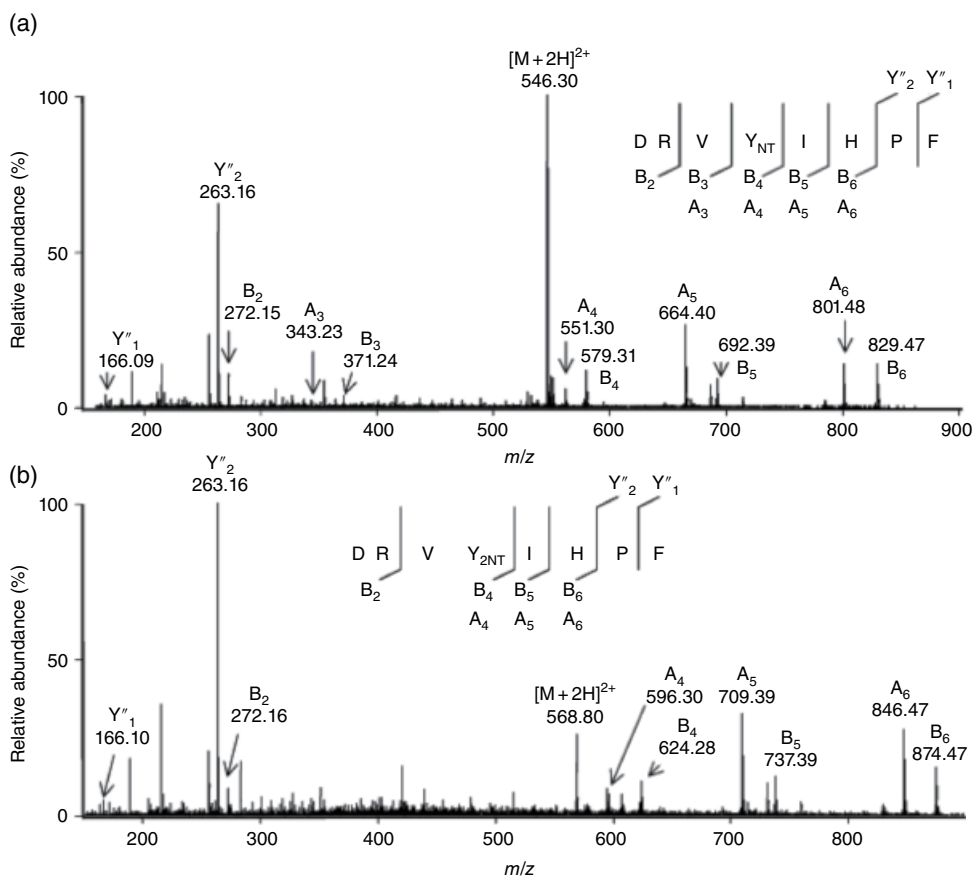


FIGURE 67.5 The MS/MS spectra of mononitrated angiotensin II peptide (precursor ion $[M+2H]^{2+}$ at m/z 546.30) (a) and dinitrated angiotensin II peptide (precursor ion $[M+2H]^{2+}$ at m/z 568.80) (b). Source: Petersson et al. [18]. Reproduced with permission of John Wiley & Sons, Inc.

intense fragment ions relative to y-ions (Fig. 67.7) [20]; those data were confirmed with UV-laser MALDI-MS/MS analysis of nitrated angiotensin II [18]. Second, relative to unmodified peptide (LE1), more collision energy optimized ion fragmentation of the nitropeptide (Fig. 67.8a) but increased intensity of the a_4 -ion and decreased intensity of the b_4 -ion (a -ion = loss of CO from a b -ion) (Fig. 67.8b). Third, optimized UV-laser fluence maximized ion fragmentation of the nitropeptide. Fourth, MS³ analysis confirmed the MS²-derived amino acid sequence; however, MS³ analysis required a greater amount of peptides relative to MS² [20]. Thus, MS³ analysis might not be suitable for routine analysis of endogenous low-abundance nitroproteins. Only when a target is determined can MS³ be used for confirmation. Fifth, to detect a nitropeptide, the amount of peptide must satisfy the sensitivity of a mass spectrometer; for synthetic nitropeptides, the sensitivity of vMALDI-LTQ was 1 fmol for MS detection and 10 fmol for MS² detection [20].

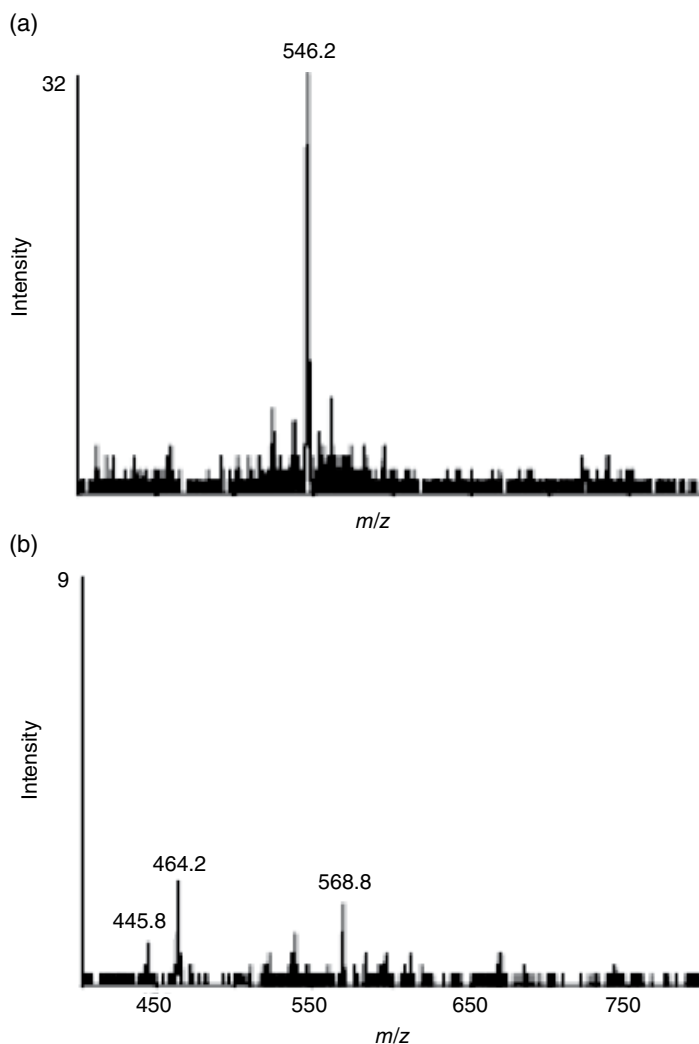


FIGURE 67.6 Precursor-ion scans spectra of nitrated angiotensin II based on immonium ion at m/z 181.06 for mononitrated tyrosine (a) and at m/z 226.0 for dinitrated tyrosine (b). Source: Petersson et al. [18]. Reproduced with permission of John Wiley & Sons, Inc.

67.2.4 MS/MS Spectral Characteristics of a Nitropeptide under Different Ion-Fragmentation Models

For MS/MS analysis of a nitropeptide, ion-fragmentation behaviors differ significantly among CID-, electron-capture dissociation (ECD)-, electron-transfer dissociation (ETD)-, and metastable atom-activated dissociation (MAD)-MS [46, 47]:

- i. CID behavior of a nitropeptide. Studies demonstrated that the presence of nitration did not affect the CID behavior of the peptides.

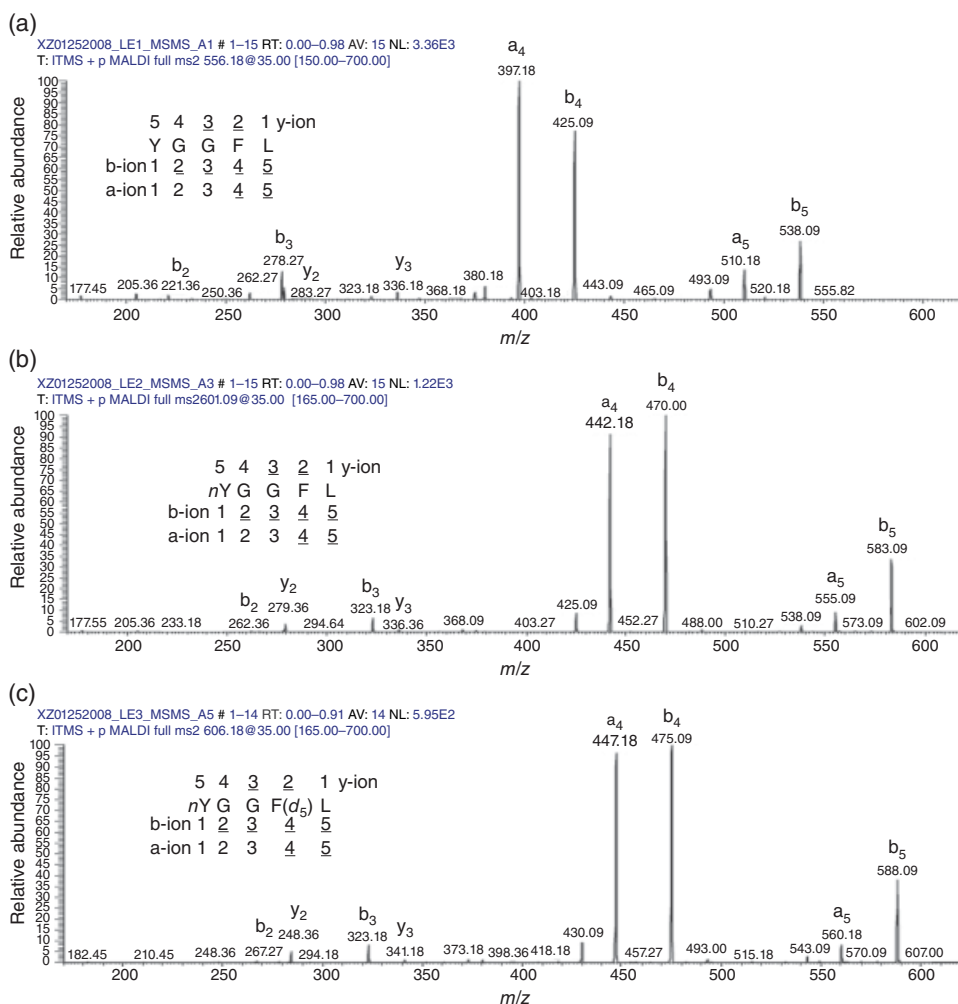


FIGURE 67.7 MS/MS spectra of LE1 (a), LE2 (b), and LE3 (c). nY = Nitrotyrosine residue. F(d_5) = Phe residue with five ^2H (d) atoms. Source: Zhan and Desiderio [20]. Reproduced with permission of Elsevier.

- ii. ECD behavior of a nitropeptide. For doubly charged peptides, production of ECD sequence fragments was severely inhibited with nitration; ECD of triply charged nitropeptides produced some singly charged sequence fragments. ECD of nitropeptides was characterized with multiple losses of small neutral species, including hydroxyl radicals, water, and ammonia. The origin of neutral losses was investigated with activated ion (AI) ECD. Loss of ammonia appears to be the result of noncovalent interactions between a nitro group and protonated lysine side chains [47, 48].
- iii. MAD behavior of a nitropeptide. Some studies found that high kinetic energy helium MAD produced extensive backbone fragmentation with significant

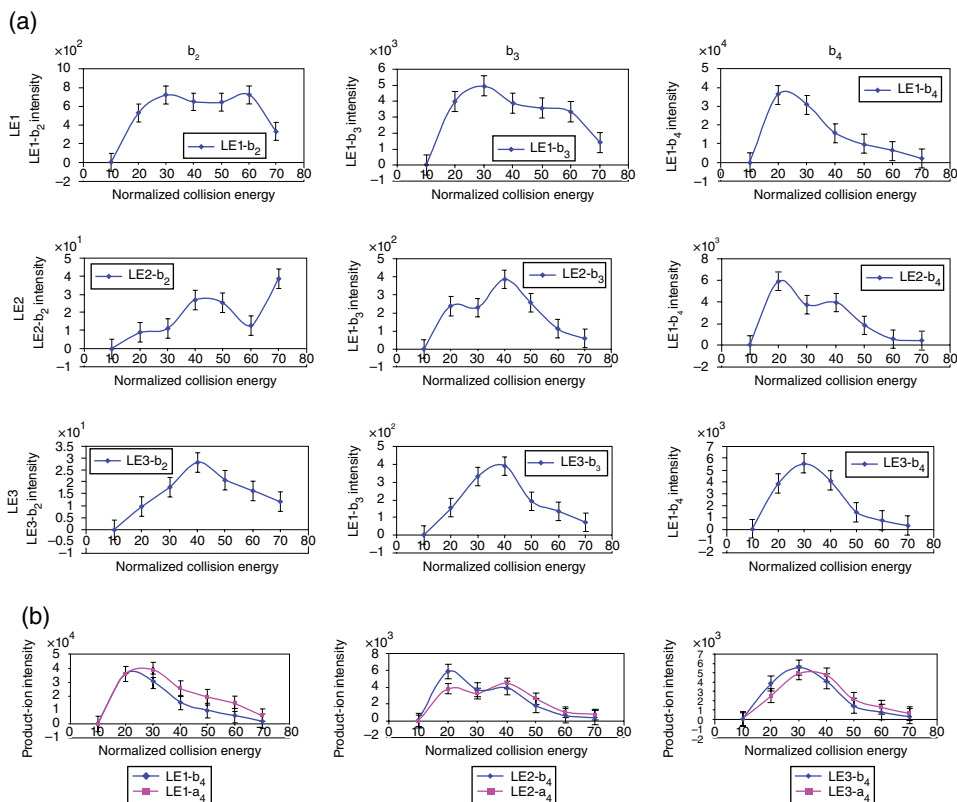


FIGURE 67.8 Effect of collision energy on collision-induced dissociation (CID) fragmentation of nitropeptides. (a) Relationship between collision energy and product-ion intensity ($n=3$). (b) Relationship between collision energy and product ion b_4 and a_4 intensities ($n=3$). Source: Zhan and Desiderio [20]. Reproduced with permission of Elsevier.

retention of PTMs. Although the high electron affinity of a nitrotyrosine moiety quenched radical chemistry and fragmentation in ECD and ETD, MAD does produce numerous backbone cleavages in the vicinity of the nitration. Compared to CID, MAD produced more fragment ions and differentiated I/L residues in nitrated peptides. MAD induced radical ion chemistry even in the presence of strong radical traps and, therefore, offers unique advantages to ECD, ETD, and CID for determination of nitropeptides [46].

- iv. The different types of CID-MS/MS have different abilities to identify nitroproteins [49]. For example, for the same samples, 119 nitropeptides and 23 multiply nitrated nitropeptides were studied with a QSTAR Elite (QTOF) with CID, whereas 197 nitropeptides and 36 multiply nitrated nitropeptides were studied with a dual-pressure ion-trap mass spectrometer (LTQ Velos) with CID (Fig. 67.9) [49]. Therefore, it is essential to choose an appropriate mass spectrometer to analyze nitropeptides/nitroproteins.

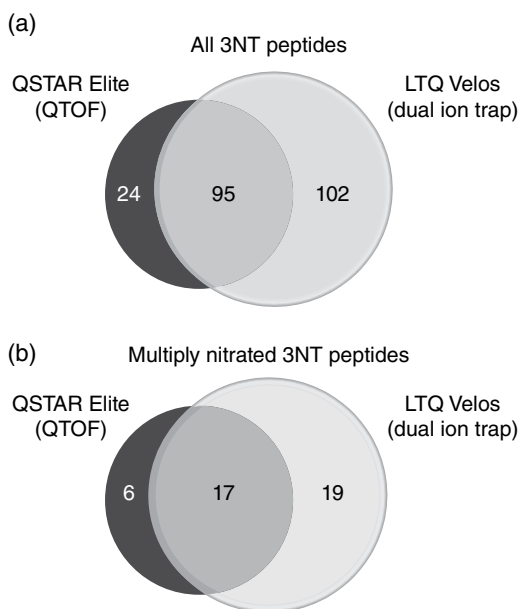


FIGURE 67.9 Overlap analysis of validated 3-nitrotyrosine (3NT)-containing peptides identified with a QSTAR Elite and LTQ Velos. Venn diagram of the overlap of all validated 3NT (a) and multiple 3NT-modified (b) peptides identified with a QSTAR Elite and LTQ Velos. Source: Li et al. [49]. Reproduced with permission of Elsevier.

67.3 MS MEASUREMENT OF *IN VITRO* SYNTHETIC NITROPROTEINS

67.3.1 Importance of Measurement of *In Vitro* Synthetic Nitroproteins

Detection of a nitropeptide is affected by the particular MS characteristics of nitropeptides, the extreme low abundance of *in vivo* nitrotyrosine sites, and MS sensitivity limitations. Moreover, it is much easier to obtain or produce *in vitro* synthetic or nitrated nitropeptides/nitroproteins relative to *in vivo* nitrotyrosine sites. Therefore, it is necessary to use *in vitro* synthetic nitroproteins to develop and establish methods to analyze *in vitro* nitroproteins.

67.3.2 Commonly Used *In Vitro* Nitroproteins and Their Preparation

Some synthetic peptides such as AAFGY($-\text{NO}_2$)AR [19], leucine enkephalin, nitro-Tyr-leucine enkephalin, and d_5 -Phe-nitro-Tyr-leucine enkephalin [20] have been used to study MS characteristics of a nitropeptide. Angiotensin II, BSA, and ovalbumin (OVA) are commonly used standard peptides and proteins that are *in vitro* nitrated with liquid TNM [5, 18, 19, 38, 50], gaseous nitrogen dioxide and ozone ($\text{NO}_2 + \text{O}_3$) [38], or peroxynitrite [51]. In order to simulate *in vivo* proteome situations, some proteomes such as human plasma were also nitrated *in vitro* with TNM followed by nitroproteomic analysis [34].

67.3.3 Methods Used to Measure *In Vitro* Synthetic Nitroproteins

Several well-established nitroproteomic methods based on anti-3-nitrotyrosine antibody and gel-based separations have been used to study *in vitro* prepared standard nitropeptides, nitroproteins, and nitroproteome samples [18, 26]. Methods that involved multidimensional chromatography, diagonal chromatography [36, 37], precursor-ion scanning [18], and/or chemical derivation can identify and quantify protein nitration sites [26]. Of them, chemical derivation is an important step to establish those methods.

Several chemical derivatization methods have been developed to analyze nitropeptide/nitroprotein prior to MS analysis [25]. Because an amino ($-\text{NH}_2$) group in a nitroprotein/nitropeptide is more stable than a nitro ($-\text{NO}_2$) group during MS analysis, all chemical derivation methods of nitrotyrosine reduce nitrotyrosine to aminotyrosine with reducing agents, including $\text{Na}_2\text{S}_2\text{O}_4$ [5, 19, 52], and derive the generated amino group with specific reagents. Those chemical derivatization methods are presented as follows:

- i. A nitrotyrosine residue converted to an aminotyrosine residue *via* reduction can readily discern aminotyrosine peptides in a background of nonnitrated peptides. Thus, aminotyrosine peptides were more stable in a single MS mode and led to easy-to-interpret peptide mass maps [51].
- ii. Dansyl chloride was used to label nitration sites. MS/MS and a precursor-ion scan identified the proteins and determined the nitrotyrosine sites [31, 32].
- iii. A method that specifically enriches nitropeptides to unambiguously identify nitrotyrosine peptides and nitration sites with LC-MS/MS was used to follow conversion of nitrotyrosine to *N*-thioacetyl-aminotyrosine followed with high-efficiency enrichment of sulfhydryl-containing peptides with thiopropyl sepharose beads [24]. Briefly, the derivatization protocol includes the following: (a) all primary amines were acetylated with acetic anhydride to block those primary amines, (b) nitrotyrosine was reduced to aminotyrosine, (c) aminotyrosine was derivatized with *N*-succinimidyl *S*-acetylthioacetate, and (d) *S*-acetyl on *S*-acetylthioacetate was deprotected to form free sulfhydryl groups [24]. This method has been used to analyze *in vitro* nitrated human histone H1.2, BSA, and mouse brain tissue samples [24].
- iv. Although iTRAQ is an effective quantitative proteomics method, it is limited to primary amines. A new strategy based on the use of iTRAQ reagents coupled with MS analysis was developed to selectively label nitrotyrosine residues [35] to simultaneously localize and quantify nitration sites in model proteins and biological systems [35].
- v. A strategy that combined precursor isotopic labeling and isobaric tagging (cPILOT) was developed to increase the multiplexing capability to quantify a nitrotyrosine protein to 12 or 16 samples with TMT or iTRAQ, respectively. Light- and heavy-labeled acetyl groups were used to block *N*-terminal and lysine residues of tryptic peptides. Nitrotyrosine was reduced to aminotyrosine with sodium dithionite, and light- and heavy-labeled aminotyrosine peptides were derivatized with either

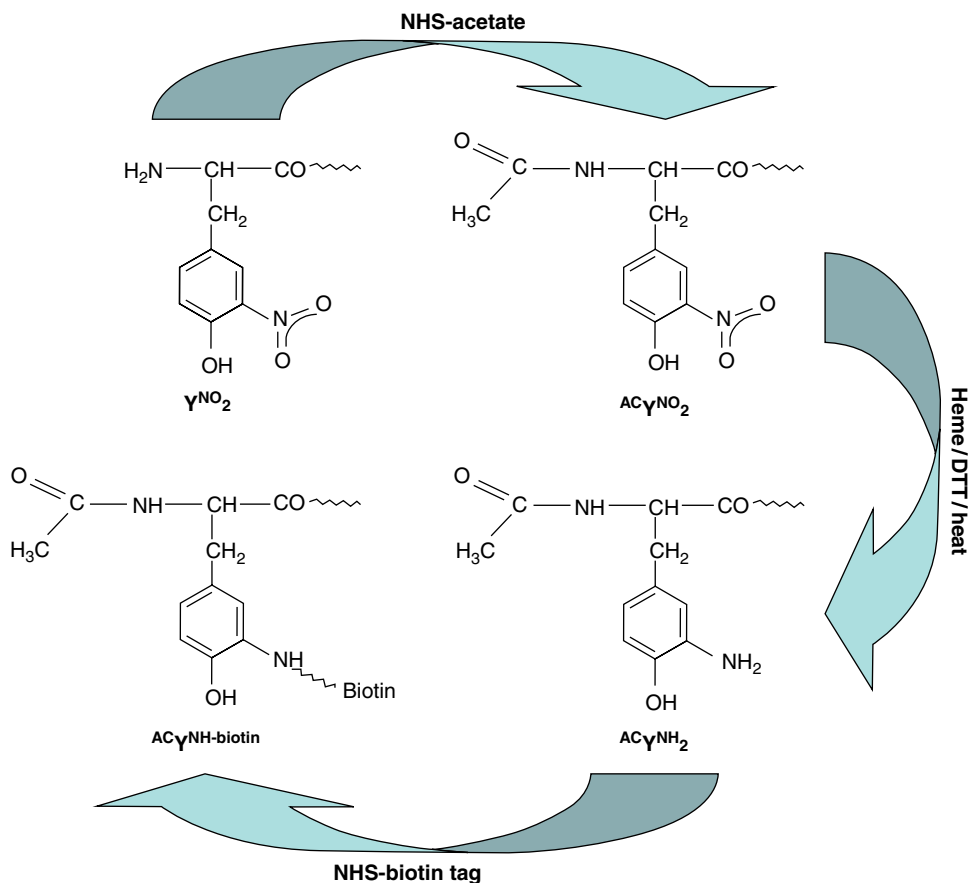


FIGURE 67.10 Reaction scheme of the chemical-labeling method as exemplified with an *N*-terminal nitrotyrosine residue. All amines were blocked with acetylation with acetic acid *N*-hydroxysuccinimide ester (NHS acetate). Nitrotyrosine was reduced to aminotyrosine with heme and DL-dithiothreitol in a boiling-water bath. The reaction sequence was completed with biotinylation of aminotyrosine with NHS-biotin. Source: Abello et al. [29]. Reproduced with permission of Elsevier.

TMT or iTRAQ multiplex reagents [33]. This method demonstrated proof of principle to analyze *in vitro* nitrated BSA and mouse splenic proteins [33].

- vi. An improved chemical-labeling method was designed to enrich nitropeptides independent of sequence context (Fig. 67.10). Briefly, all amines were blocked with acetylation, and nitrotyrosine was converted to aminotyrosine, followed by biotinylation of aminotyrosine [29]. Moreover, the entire reaction was carried out in a single buffer without any sample cleanup or pH changes to minimize any sample loss. Also, a strong cation exchanger was used to remove free biotin, and an immobilized avidin column was used to enrich the labeled peptides, followed by analysis of enriched peptides with LC-MS/MS [29]. This method was approved for *in vitro* nitrated samples [29, 53].

- vii. Because a MALDI UV-laser causes photochemical decomposition of a nitro group in a nitropeptide, a new strategy includes (a) acetylation of *N*-terminal amines and epsilon amines of lysine residues with acetic anhydride, (b) reduction of nitrotyrosine to aminotyrosine with sodium hydrosulfite, and (c) derivatization of aminotyrosine with 1-(6-methyl[d_6/d_3]nicotinoyloxy) succinimide, followed by MALDI-TOF MS analysis [30]. The optimum matrix was sinapinic acid, not 2,5-dihydroxybenzoic acid, for MALDI-MS analysis of a nitropeptide [54].
- viii. Another developed method is the COFRADIC approach [36, 37]. Nitrotyrosine is reduced to aminotyrosine with sodium dithionite and peptides sorted with reversed-phase chromatography based on a hydrophilic shift from nitropeptide (more hydrophilic) to aminotyrosine-containing peptide (more hydrophobic) followed by EDI-MS identification [36] and MALDI-MS [37]. COFRADIC characterized tyrosine nitration in a TNM-nitrated BSA and peroxynitrite-nitrated proteome of human Jurkat cells [36, 37].

Interpretation of MS and MS/MS data of nitropeptides (and especially endogenous) is very challenging. To avoid any risk of linking MS/MS spectra to an incorrect amino acid sequence, the combination of reduction of nitrotyrosine to aminotyrosine and use of the Peptizer algorithm to inspect MS/MS quality-related assumptions [55] has been developed.

The optimal approach to determine the amino acid sequence of an endogenous nitropeptide is a manual approach [3].

67.4 MS MEASUREMENT OF *IN VIVO* NITROPROTEINS

67.4.1 Importance of Isolation and Enrichment of *In Vivo* Nitroprotein/Nitropeptide Prior to MS Analysis

Nitration of tyrosine residue in a protein mainly due to oxidative stress is a low-abundance (1 in $\sim 10^6$ tyrosines) modification in an *in vivo* proteome [14, 56]. MS is the key technique to identify nitropeptides/nitroproteins and to accurately determine nitrotyrosine sites in a nitroprotein [2, 3, 57]. MS sensitivity is the high-femtomole/low-picomole level [20]. Therefore, it is essential to isolate and preferentially enrich nitroproteins/nitropeptides before MS analysis [17, 21, 22, 57].

67.4.2 Methods Used to Isolate and Enrich *In Vivo* Nitroproteins/Nitropeptides

Several enrichment protocols have been used for isolation and preferential enrichment of *in vivo* nitropeptides/nitroproteins from a biological proteome: (i) Two-dimensional gel electrophoresis (2DGE) coupled with nitrotyrosine Western blotting was used to

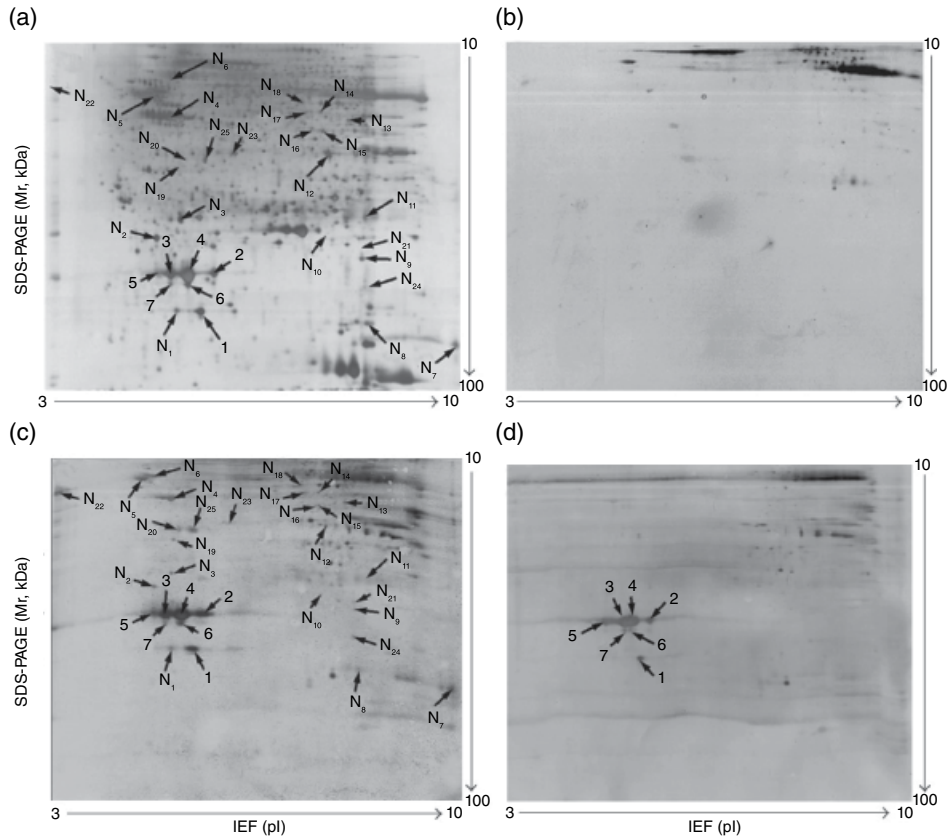


FIGURE 67.11 Two-dimensional Western blotting analysis of anti-3-nitrotyrosine-positive proteins in a human pituitary (70 μ g protein per 2D gel). (a) Silver-stained image on a 2D gel before transfer of proteins onto a PVDF membrane. (b) Silver-stained image on a 2D gel after transfer of proteins onto a PVDF membrane. (c) Western blot image of anti-3-nitrotyrosine-positive proteins (anti-3-nitrotyrosine antibodies + secondary antibody). (d) Negative control of a Western blot to show the cross-reaction of the secondary antibody (only the secondary antibody; no anti-3-nitrotyrosine antibody). Source: Zhan and Desiderio [57]. Reproduced with permission of Elsevier.

analyze nitroproteins in a proteome (Fig. 67.11). Briefly, the nitroproteins in a proteome were arrayed and relatively enriched with 2DGE, transferred to PVDF membrane, and detected with antinitrotyrosine antibody, followed by visualization [2, 7, 16, 57–59]. (ii) Nitrotyrosine affinity column (NTAC) (Fig. 67.12) was used to enrich nitroproteins [3, 7, 60] and to enrich nitropeptides [61]. (iii) A method was used to acetylate all primary amines in a nitropeptide, convert easily a nitrotyrosine residue into aminotyrosine, and then enrich with biotinylation of an aminotyrosine (Fig. 67.10) [5, 25, 29]. (iv) A method was used to acetylate all primary amines, reduce nitrotyrosine to aminotyrosine, derivatize aminotyrosine into a free sulfhydryl group, and then enrich sulfhydryl-containing

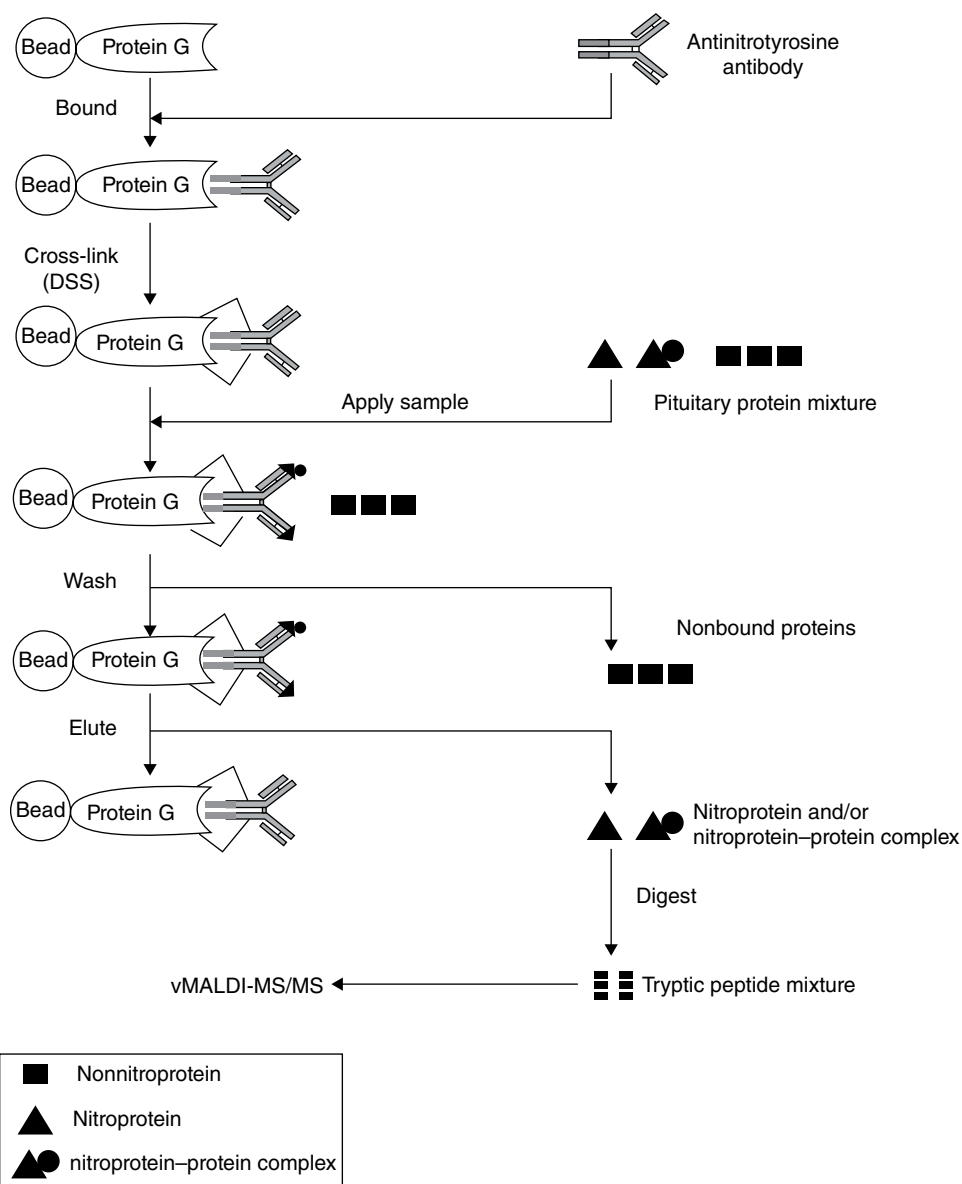


FIGURE 67.12 Experimental flowchart to identify nitroproteins and nitroprotein–protein complexes with NTAC-based MALDI-LTQ MS/MS. The control experiment (without any anti-3-nitrotyrosine antibody) was carried out in parallel with the NTAC-based experiments. Source: Zhan and Desiderio [3]. Reproduced with permission of Elsevier. Zhan, Wang, and Desiderio [7]. CC-BY.

peptides with thiopropyl sepharose beads [24]. (v) A method was utilized for the use of dansyl chloride to label nitration sites, followed with a precursor-ion scan and MS³ analysis [31, 32]. (vi) A method was utilized for the use of a new tagging reagent, (3R, 4S)-1-(4-(aminomethyl) phenylsulfonyl)pyrrolidine-3,4-diol (APPD),

for the selective fluorogenic derivatization of nitrotyrosine residues in peptides (after reduction to aminotyrosine) and boronate affinity enrichment [62]. (vii) COFRADIC was used to sort peptides per the hydrophilic shift after a nitro group was reduced to an amino group and then analyzed with ESI or MALDI-MS [36]. (viii) TMT- or iTRAQ-based quantitative nitroproteomics was used to quantitatively identify nitroproteins/nitropeptides [33, 35]. Briefly, the *N*-terminal and lysine residues of tryptic peptides were blocked with “light-” and “heavy-”labeled acetyl groups, and nitrotyrosine was reduced to aminotyrosine, followed by derivatization of light- and heavy-labeled aminotyrosine-containing peptides with either TMT or iTRAQ multiplex reagents [33, 35]. That method can relatively enrich and quantitatively identify nitroproteins/nitropeptides.

One should note that protocols (i), (ii), and (vii) were used in the identity of *endogenous* nitrotyrosine sites [2, 3, 36, 57, 60]; whereas protocols (iii)–(vii) succeeded mainly with an *in vitro* model peptide or protein and with an *in vitro* nitrated proteome [5, 24, 31, 36, 37]; however, they offer promise to study *in vivo* nitroproteins. Protocols (i)–(vii) mainly focused on the identity of nitropeptides, nitroproteins, and nitrotyrosine sites. However, in order to determine disease-related nitroproteins, it needs to quantitatively identify nitroproteins except for characterization of nitrotyrosine sites and nitroprotein. Protocol (viii) holds promise to achieve that goal because it can relatively enrich nitropeptides, identify nitrotyrosine sites, and quantify nitroproteins; and its sample-multiplexing capabilities has also been enhanced [33, 35].

67.5 MS MEASUREMENT OF *IN VIVO* NITROPROTEINS IN DIFFERENT PATHOLOGICAL CONDITIONS

Tyrosine nitration in a protein is an important oxidative/nitrative stress-mediated modification, which functions in a wide range of cellular, physiological, and pathological processes [63, 64]. Tyrosine nitration can alter activity of a protein and extensively associate pathophysiological conditions. The documented literature demonstrates that endogenous nitroproteins and nitrotyrosine sites have been identified in different types of pathophysiological conditions (Table 67.1) and was summarized here:

- i. Tyrosine nitration in inflammation-related diseases. Studies have demonstrated that tyrosine nitration is extensively associated with inflammatory diseases with identification of nitroproteins in a septic patient’s rectus abdominis muscle [65], bronchial epithelial cells, and bronchoalveolar lavage with asthma [66], Chagas’ disease [67], experimental sepsis [68], and serum sample of a C57BL/6J mouse model with septic shock [36].
- ii. Tyrosine nitration in aging and aging-related diseases. Protein nitration is extensively studied in aging and aging-related diseases with discovery of nitroproteins in aging rat heart [69], rat skeletal muscle [70, 71], and mouse liver [72].

TABLE 67.1 Endogenous Nitroproteins Identified from Different Pathological Conditions

Reference	Specimen	Methods	Nitroprotein and Nitrotyrosine Sites	Remark
<i>(i) Inflammation-related disease</i>				
Lanone et al. [65]	Rectus abdominis muscle from the same control and septic patients	Western blot, MALDI-TOF-PMF, and molecular modeling	Inducible nitric oxide synthase (iNOS) was nitrated at Tyr299, Tyr336, Tyr446, and Tyr698	Analysis coupled with iNOS three-dimensional crystal model
Ghosh et al. [66]	Lung tissues from allergen-induced murine model of asthma	2D-Western blot and LC-ESI-MS/MS	Twenty-seven putative nitrated proteins were identified	Inflammation-related disease. No nitrotyrosine site were identified
Dhiman et al. [67]	Plasma from patients with Chagas' disease	1D- and 2D-Western blot, MALDI-PMF, and LC-ESI-MS/MS	Fifty differentially expressed/ nitrated proteins were identified	Inflammation-related disease. No nitrotyrosine site were identified
Chatterjee et al. [68]	Spleens from LPS-induced systemic inflammation model of C57BL6/J mice	1D-Western blot and LC-ESI-MS/MS	Carboxypeptide B1 (CPB1) was nitrated at specific tyrosine sites	Inflammation-related disease
Ghesquiere et al. [36]	Serum proteome from a C57BL6/J mouse with septic shock	COFRADIC and ESI-MS	α 2-Macroglobulin, apolipoprotein A-I, haptoglobin, and vitamin D-binding protein were nitrated at six specific tyrosine sites	Inflammation-related disease. Nitrotyrosine sites were identified
<i>(ii) Aging and aging-related diseases</i>				
Kanski et al. [69]	Heart from 5-month-old and 26-month-old Fischer 344/BN F1 hybrid rats	1D- and 2D-Western blot, ESI-MS/MS	Forty-eight putative nitrated proteins. Nitration at Tyr105 of the electron-transfer flavoprotein was identified	Endogenous. Heart homogenate and heart mitochondria. Nitration is effects of biological aging. Not every protein's nitrotyrosine site was identified
Kanski, Hong, and Schöneich [70]	Skeletal muscle from 34-month-old Fischer 344/Brown Norway F1 rats	IEF, 1D-Western blot, and ESI-MS/MS	Eleven nitroproteins and twelve nitrotyrosine sites were identified	Endogenous

Sharov et al. [71]	Skeletal muscle from 6-month-old and 34-month-old Fischer 344/Brown Norway F1 hybrid rats	1D-Western blot and HPLC-ESI-MS/MS	Phosphorylase b was found in the accumulation of 3-nitrotyrosine on Tyr113, Tyr161, and Tyr573. Nitration on Tyr 113 was detected in 6-month-old and 34-month-old rat; nitration on Tyr161 and Tyr573 was detected only in 34-month-old rat	Endogenous. Nitration is accumulated with aging
Marshall et al. [72]	Liver from young (19–22 weeks) and old (24 months) C57/BL6 male mice	1D-Western blot and LC-ESI-MS/MS	Six putative nitrated proteins were identified	Nitration is associated with aging. No nitrotyrosine site was identified
<i>(iii) Tumor</i>				
Fiore et al. [73]	Human glioma tissues	Immunohistochemistry, 1DE-MALDI-PMF	Tublin was nitrated at Tyr224 in glioma grade IV but not in grade I and noncancerous brain tissue	Endogenous. Laser-induced decomposition
Nakagawa et al. [74]	C6 rat glioma cell line	HPLC, MALDI-PMF	Cytochrome c was nitrated at Tyr48, Tyr67, and Tyr74	<i>In vitro</i> nitrated with peroxyinitrite
Zhan and Desiderio [2]	Human pituitary postmortem tissue	2D-Western blot and vMALDI-MS/MS	Four nitroproteins and four nitrotyrosine sites were identified	Endogenous. Laser-induced decomposition
Zhan and Desiderio [3]	Human nonfunctional pituitary adenoma tissue	NTAC-vMALDI-MS/MS	Nine nitroproteins, ten nitrotyrosine sites, and three nitroprotein-interacting protein were identified	Endogenous. Laser-induced decomposition
Zhan and Desiderio [57]	Human pituitary postmortem tissue	2D-Western blot and vMALDI-MS/MS	Four nitroproteins and four nitrotyrosine sites were identified	Endogenous. Laser-induced decomposition

(Continued)

TABLE 67.1 (Continued)

Reference	Specimen	Methods	Nitroprotein and Nitrotyrosine Sites	Remark
<i>(iv) Neurodegenerative diseases</i>				
Casoni et al. [75]	Spinal cord from Tg SOD1 G93A mice and Tg SOD1 WT mice	2D-Western blot and MALDI-PMF	Thirty-two nitroproteins and sixteen nitrotyrosine sites	Endogenous. Laser-induced decomposition. Familial amyotrophic lateral sclerosis
Sacksteder et al. [76]	Brain from C57BL/6J mice	SCX-LC-ESI-MS/MS	Twenty-nine nitroproteins and thirty-one nitrotyrosine sites	Endogenous. Links to neurodegenerative disease
Danielson et al. [77]	Dox-inducible MAO-B PC12 cells	LC-ESI-MS/MS	Alpha-synuclein was nitrated at Tyr39	Model of Parkinson's disease
Yoon et al. [78]	Mouse hippocampal cell line HT22	2D-Western blot and MALDI-PMF	Thirteen nitroproteins were detected	Glutamate-treated HT22 cells
Zhang et al. [79]	Brain from C57BL/6J mice	LC-ESI-MS/MS		Endogenous
<i>(v) Cardiovascular system and related diseases</i>				
Chen et al. [80]	Sprague Dawley rat <i>in vivo</i> myocardial regional ischemia-reperfusion model	1DE-LC-ESI-MS/MS	Flavin subunit is nitrated at Tyr56 and Tyr142	Endogenous. Mitochondrial complex II in the posts ischemic myocardium
Ai et al. [81]	Endothelial cell of human coronary arteries	LC-ESI-MS/MS	LDL was nitrated	
Liu et al. [82]	Male C57BL/6 mice myocardial ischemia-reperfusion (I/R) injury model	1D- or 2D-Western blot, LC-ESI-MS/MS	Twenty-three nitroproteins were identified. Ten of them were from mitochondria	Endogenous. No nitrotyrosine sites were identified
<i>(vi) The neurovisual system</i>				
Palamalai et al. [83]	Photoreceptor rod outer segments of cyclic light-reared rats treated or not with the antioxidant	2D-Western blot and LC-ESI-MS/MS	Ten putative nitroproteins were identified	Endogenous. No nitrotyrosine sites were identified

Justilien et al. [60]	Mouse posterior eyecups	NTAC, SDS-PAGE, LC-ESI-MS/MS	Eight nitroproteins and nine nitrotyrosine sites	Endogenous. SOD2 knockdown mouse model of early AMD
Murdaugh et al. [84]	Human Bruch's membrane	HPLC, ESI-MS/MS	A2E was nitrated	Endogenous. Nitro-A2E is a specific biomarker of nitrosative stress in Bruch's membrane, and its concentration is directly related to tissue age
<i>(vii) Diabetes</i>				
Kato et al. [85]	Healthy and diabetic human urine	LC-ESI-MS/MS	Urine nitrotyrosine	Endogenous
<i>(viii) Kidney disease</i>				
Piroddi et al. [86]	Plasma from kidney disease patients	2DE and LC-ESI- MS/MS	Fourteen tentative nitroproteins and seven nitrotyrosine sites were identified	Endogenous
<i>(ix) Plant diseases</i>				
Chaki et al. [87]	Sunflower hypocotyls	2D- Western blot and LC-ESI-MS/MS	Twenty one putative nitroproteins were identified	No nitrotyrosine sites were identified
<i>(x) Others</i>				
Aslan et al. [88]	Liver and kidney from sickle cell disease mouse	Western blot and precipitation, MALDI- PMF, LC-ESI-MS/MS	Actin was nitrated at Tyr91, Tyr198, and Tyr240	Endogenous
Webster, Brockman, and Myatt [89]	Human placenta	1D- Western blot and MALDI-PMF	p38 MAPK was nitrated	No nitrotyrosine sites were identified

(Continued)

TABLE 67.1 (Continued)

Reference	Specimen	Methods	Nitroprotein and Nitrotyrosine Sites	Remark
Ulrich et al. [90]	Human lung tissues and blood samples, aminal granule protein preparation	Western blot and MALDI-PMF	Six nitroproteins and nitrotyrosine sites at Tyr349 in eosinophil peroxidase (EPO) and Tyr33 in other eosinophil cationic protein (ECP) and eosinophil-derived neurotoxin (EDN)	Endogenous
Hamilton et al. [91]	Human plasma	LC-ESI-MS/MS	Low-density lipoprotein (LDL) was nitrated at Tyr276, Tyr666, and Tyr720 of LDL-alpha 1, Tyr2524 of LDL-alpha 2, Tyr4141 of LDL-alpha 3, Tyr3139, Tyr3205, and Tyr3489 of LDL-beta 2	
Zhu et al. [92]	Liver from SOD1 ^{-/-} and WT C57BL/6 mice	1DE, LC-ESI-MS/MS	Ten candidate nitrated proteins were identified	No nitrotyrosine sites were identified
Reed et al. [93]	Traumatic brain injury rats	2D-Western blot and MALDI-PMF	Several nitroprotein such as GSH were identified	
Lee et al. [45]	Hippocampus from smoke inhalation rat model	2D-Western blot and MALDI-PMF or MALDI-MS/MS	Five nitroproteins of mitochondrial proteins were identified	Endogenous. No nitrotyrosine sites were identified
Casanovas et al. [94]	Lipoprotein lipase of bovine and rat	2D-Western blot and LC-ESI-MS/MS	Lipoprotein lipase was nitrated at Tyr95, Tyr164, Tyr316	Endogenous
Sharov et al. [95]	Rabbit muscle	LC-ESI-MS/MS	Glycogen phosphorylase b was nitrated at 28 nitrotyrosine sites	
Ohama and Brautigan [96]	Human peripheral blood mononuclear cells	LC-ESI-MS/MS	Protein phosphatase 2A was nitrated at Tyr284	

Sekar et al. [97]	Mast cells	2D- Western blot and LC-ESI-MS/MS	Aldolase was nitrated	No nitrotyrosine sites were identified
Redondo-Horcajo et al. [98]	Endothelial cells from bovine aorta and mouse lung	LC-ESI-MS/MS	Manganese superoxide dismutase (MnSOD) was nitrated at Tyr34	<i>In vitro</i> nitrated with cyclosporine A
Chen and Chen [99]	Human blood samples from smokers and nonsmokers	LC-ESI-MS/MS under the selected reaction monitoring (SRM) mode	Hemoglobin was nitrated at Tyr24 and Tyr42 (alpha globin) and Tyr130 (beta globin)	Nitration of human hemoglobin is associated with cigarette smoking

- iii. Tyrosine nitration in tumorigenesis. Tumor is another important target that tyrosine nitration is involved in with clear evidences that nitroproteins were identified in human gliomas [73] and rat glioma cell lines [74]; moreover, 8 nitroproteins were discovered in human pituitary control tissue [2, 57], and 9 nitroproteins, 3 nitroprotein-interacted proteins, and 10 nitrotyrosine sites were discovered in a pituitary adenoma [3].
- iv. Tyrosine nitration in neurodegenerative diseases. Studies have discovered nitroproteins in neurodegenerative-related model or disease such as spinal cords of a mouse model of familial amyotrophic lateral sclerosis [75], Parkinson's disease [77], mouse brain [76, 79, 100], and HT22 hippocampal cells [78].
- v. Tyrosine nitration in the cardiovascular system and related diseases. It has been characterized with identification of nitroproteins in the vascularity [81], ischemia–reperfusion injury [80, 82, 101], and mouse heart [79, 100].
- vi. Tyrosine nitration in the neurovisual system. For example, nitroproteins were discovered in outer segments of photoreceptor rod [83], the eyecup of SOD₂ knockdown mouse [60], and human Bruch's membrane [84].
- vii. Tyrosine nitration in diabetes. Diabetes were found to involve tyrosine nitration modification with the discovery of nitroproteins in diabetic mellitus patients [102], a diabetic patient's urine [85], and diabetic rat models [103].
- viii. Tyrosine nitration in kidney disease. It was evidenced with characterization of 14 nitroproteins and 7 nitrotyrosine sites in a kidney disease patient's plasma [86].
- ix. Plant disease. Studies discovered nitroproteins in plant disease [87, 104].
- x. Others. Tyrosine nitration is also discovered to participate in many other pathophysiological processes with identification of nitroproteins in sickle cell disease [88], placenta/preeclampsia [89], murine liver [92], eosinophil granule toxins [90], traumatic brain-injured rat models [93], rat hippocampus after acute inhalation of combustion smoke [45], rabbit muscle [95], hypertriglyceridemia [94], human plasma [91, 105], mononuclear cells from human peripheral blood [96], mast cells [97], endothelial cells [98], and cigarette-associated human hemoglobin [99].

67.6 BIOLOGICAL FUNCTION MEASUREMENT OF NITROPROTEINS

In order to elucidate in depth the biological functions of tyrosine nitration in a protein and biological system, the MS/MS-identified nitroproteins and nitrotyrosine sites must be further analyzed with other strategies such as literature data-based rationalization of biological function, protein domain/motif analysis, systems pathway network, and structural biology analysis. Here, nitroproteins and nitrotyrosine sites from pituitary control and adenoma (Table 67.2) are taken for an example to address those functional analyses.

TABLE 67.2 Nitroproteins and Unnitrated Proteins Identified from Pituitary Adenoma [3] and Control Tissues [2, 57]

Pituitary Adenoma		Pituitary Control	
Protein Name	nY Site	Protein Name	nY Site
<i>Nitrated protein</i>		<i>Nitrated protein</i>	
Rho-GTPase-activating 5 [Q13017] (ARHGAP5)	Y ⁵⁵⁰	Synaptosomal-associated protein (SNAP91)	Y ²³⁷
Leukocyte immunoglobulin-like receptor A4 [P59901]	Y ⁴⁰⁴	Ig alpha Fc receptor [P24071] (FCAR)	Y ²²³
Zinc finger protein 432 [O94892]	Y ⁴¹	Actin [P03996] (ACTA2, ACTG2, ACTC1)	Y ²⁹⁶
PKA beta regulatory subunit [P31321] (PRKAR1B)	Y ²⁰	PKG 2 [Q13237] (PRKG2)	Y ³⁵⁴
Sphingosine-1-phosphate lyase 1 [O95470]	Y ³⁵⁶ , Y ³⁶⁶	Mitochondrial cochaperone protein HscB [Q8IWL3]	Y ¹²⁸
Centaurin beta 1 [Q15027]	Y ⁴⁸⁵	Stanniocalcin 1 [P52823] (STC1)	Y ¹⁵⁹
Proteasome subunit alpha type 2 [P25787] (PSMA2)	Y ²²⁸	Proteasome subunit alpha type 2 (PSMA2)	Y ²²⁸
Interleukin 1 family member 6 [Q9UHA7] (IL1F6)	Y ⁹⁶	Progesterin and adipoQ receptor family member III [Q6TCH7] (PAQR3)	Y ³³
Rhopilin 2 [Q8IUC4] (RHPN2)	Y ²⁵⁸		
<i>Nitroprotein-interacted protein</i>			
Interleukin-1 receptor-associated kinase-like 2 (IRAK-2) [O43187] (IRAK2)			
Glutamate receptor interacting protein 2 [Q9C0E4] (GRIP2)			
Ubiquitin [P62988] (UBB or UBC)			

Source: Zhan and Desiderio [2, 3, 57]. Reproduced with permission of Elsevier.

Note: nY = Nitrotyrosine.

67.6.1 Literature Data-Based Rationalization of Biological Functions

A large volume of literature data analysis of 9 nitroproteins containing 10 nitrotyrosine sites and 3 nonnitrated proteins from a human pituitary adenoma (Table 67.2) [3] demonstrated that 3 nonnitrated proteins (ubiquitin, glutamate receptor-interacting protein 2, and interleukin 1 (IL1) receptor-associated kinase-like 2) interacted with nitroproteins to form 3 nitroprotein–protein complexes, including (i) nitrated beta subunit of cAMP-dependent protein kinase (PKA) complex, (ii) nitrated proteasome–ubiquitin complex, and (iii) nitrated interleukin 1 family member 6–interleukin 1 receptor–interleukin 1 receptor-associated kinase-like 2 (IL1F6–IL1R–IRAK2) [3, 7]. Furthermore, those

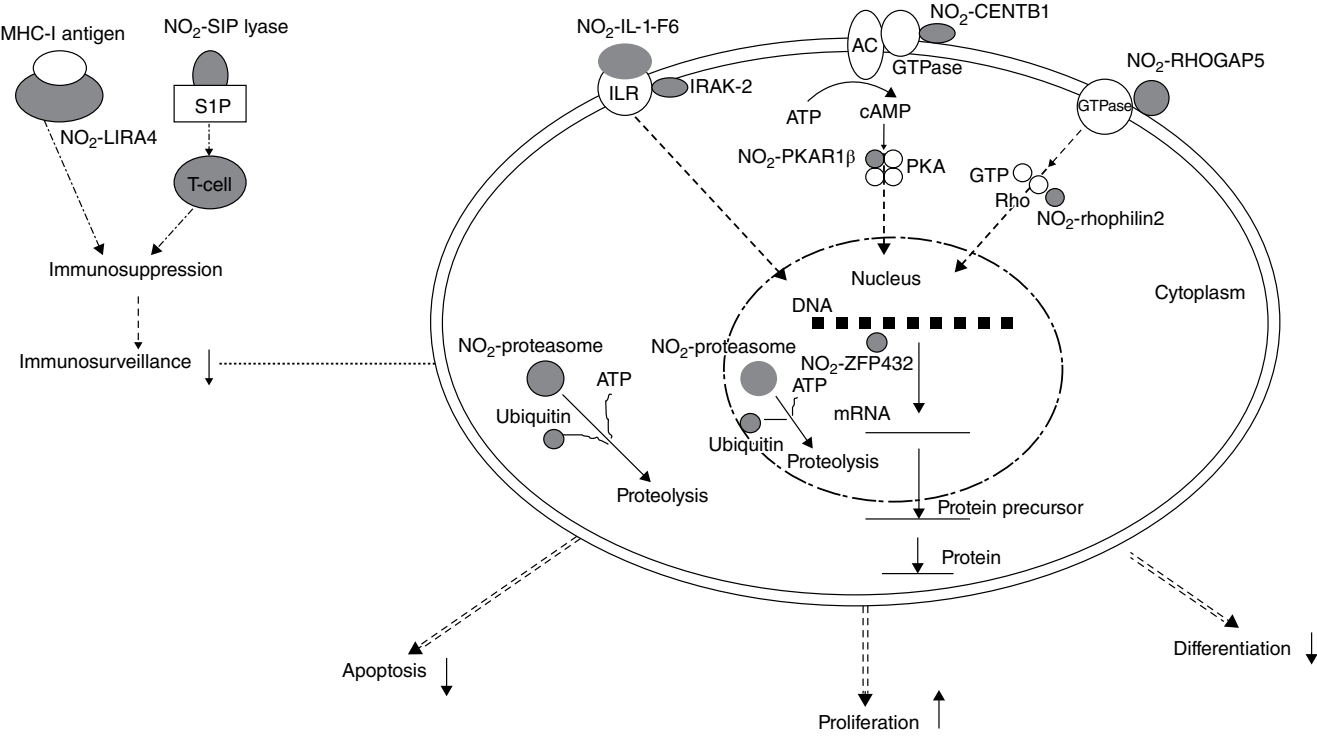


FIGURE 67.13 Experimental data-based model of nitroproteins and their functions in human nonfunctional pituitary adenomas. Source: Zhan and Desiderio [3]. Reproduced with permission of Elsevier.

three nitroprotein–protein complexes and nine nitroproteins were rationalized into an experiment-based biological function system (Fig. 67.13) [3, 7]. Nitrated leukocyte immunoglobulin-like receptor subfamily A member 4 (LIRA4) was associated with the immune system. Nitrated proteasome–ubiquitin complex, an important enzymatic complex, was involved in intracellular nonlysosomal proteolytic pathway [3, 7]. Nitrated sphingosine-1-phosphate lyase 1 (S1P lyase 1) was involved in sphingolipid metabolism to regulate immune system, cell proliferation, survival, and cell death [3, 7]. Nitrated IL1F6 and IRAK2 in the IL1R complex were involved in the cytokine system. Nitrated cAMP-dependent protein kinase type I-beta regulatory subunit (PKAR1-β) and nitrated centaurin beta 1 (CENT-β1) were involved in the PKA signal pathway. Nitrated rhophilin 2 and nitrated Rho-GTPase-activating protein 5 (RHOGAP5) were involved in the GTPase signal pathway. Nitrated zinc finger protein 432 (ZFP432) functioned in transcription regulatory systems [3, 7].

67.6.2 Protein Domain and Motif Analyses

Recognition of protein domains/motifs and location of nitrotyrosine sites into the corresponding domain/motif in a protein will much benefit the accurate clarification of biological activities of tyrosine nitration [7]. The commonly used software programs including Motif Scan (http://myhits.isb-sib.ch/cgi-bin/motif_scan), ProDom (<http://prodom.prabi.fr/prodom/current/html/form.php>), ScanProsite (<http://us.expasy.org/tools/scanprosite/>), Pfam (<http://www.sanger.ac.uk/Software/Pfam/>), and InterProScan (<http://www.ebi.ac.uk/InterProScan/>) were effective in the determination of significant domains/motifs in a nitroprotein and in location of each nitrotyrosine site within a protein domain/motif, for insights into the effect of tyrosine nitration on protein functions [3, 7]. Our studies demonstrated that most nitrotyrosine sites occurred within important domains/motifs in a protein [3] (Fig. 67.14), which hints that protein functions are altered by tyrosine nitration. For instance, nitrated S1P lyase 1 (Fig. 67.14) in a human pituitary adenoma is a key enzyme that catalyzes decomposition of S1P; two nitrations ($\text{NO}_2\text{-}^{356}\text{Y}$ and $\text{NO}_2\text{-}^{366}\text{Y}$) were discovered within the enzyme activity region to decrease the interaction intensity of enzyme–substrate (S1P lyase 1/S1P) and to alter enzymatic activities of S1P lyase 1 [3, 7].

67.6.3 Systems Pathway Analysis

Because each protein in a proteome functions in a multiple, complex, and interacting systematic network but does not work alone [39], it is necessary to rationalize tyrosine nitration within those complex pathway system networks and to address pathway network variations due to tyrosine nitration. Currently, lots of pathway network analysis software have been developed such as Ingenuity Pathway Analysis (IPA) (<http://www.ingenuity.com/>) and MetaCore Pathway Analysis programs (<http://www.genego.com/metacore.php/>). IPA was used to determine signaling networks that involve nitroproteins from a

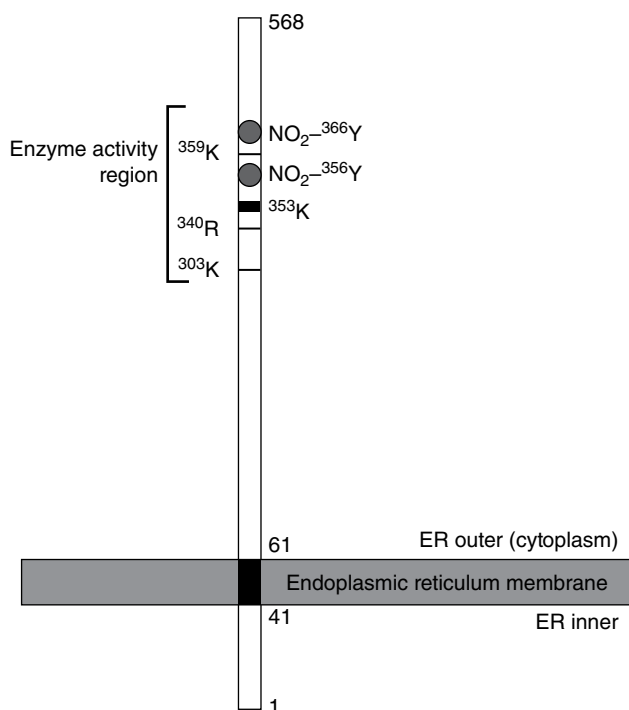


FIGURE 67.14 Nitration site and functional domains of sphingosine-1-phosphate lyase 1. Source: Zhan and Desiderio [3]. Reproduced with permission of Elsevier.

human pituitary control and adenoma tissues (Table 67.2) [3, 7]. As a result, those nitroproteins and their complexes from pituitary adenoma were involved in the IL1 and tumor necrosis factor (TNF) signaling networks (Fig. 67.15a), which function in cancer, cell cycle, and reproductive system diseases [39]; whereas those nitroproteins from control pituitary were involved in actin cellular skeleton and transforming growth factor beta 1 (TGF β 1) signaling networks (Fig. 67.15b), which function in gene expression, cellular development, and connective tissue development. Both adenoma and control networks include a beta-estradiol signal pathway, which hints that hormone metabolism is involved in pituitary normal physiology and adenoma pathology. Furthermore, pathway analysis revealed three important signaling pathway network systems (oxidative stress, cell-cycle dysregulation, and MAPK-signaling abnormality) in a pituitary adenoma that involve tyrosine nitration [7, 39]. Those pathway network data clearly elucidate biological functions and roles of tyrosine nitration in pituitary tumorigenesis.

67.6.4 Structural Biology Analysis

The electron density of a phenolic ring of a tyrosine residue is decreased by tyrosine nitration, which causes diminishing interaction intensity between receptor and ligand or between enzyme and substrate [2]. Thus, the spatial position of a nitrotyrosine in a

protein might obviously influence on biological functions and roles of tyrosine nitration. 3D spatial structure of a protein can clearly determine its biological functions. Therefore, reconstruction of 3D spatial structure of a nitroprotein with X-ray crystallography data would be very easy to clarify effects of tyrosine nitration on 3D structure of a nitroprotein. Also, based on tyrosine nitration site, domain, and 3D structure, it is possible for one to design a small drug toward 3D structure and domain that contains tyrosine nitration [7]. For example, the nicotinamide adenine dinucleotide (NAD⁺) binding assay demonstrated that nitrated glyceraldehyde-3-phosphate dehydrogenase (GAPDH) did not bind NAD⁺ [40]; while X-ray crystal structure was used to interpret effects of tyrosine nitration on the capability of NAD⁺ binding in GAPDH [40]. MS analysis of nitrated GAPDH determined Tyr³¹¹ and Tyr³¹⁷ were the only sites of nitration; and the distances between Tyr³¹¹ and Tyr³¹⁷ and the cofactor NAD⁺ were revealed by X-ray crystal structure to be less than 7.2 and 3.7 Å, respectively. Those data imply that nitration of these two residues might affect NAD⁺ binding [40]. Another example is the use of X-ray crystal structure of mammalian succinate ubiquinone reductase (SQR or complex II) to effectively interpret the association of tyrosine nitration (Tyr¹⁴²) of the flavin subunit with *S*-glutathionylated cysteine residue Cys⁹⁰ of mitochondrial complex II in a postischemic myocardium [80]. Briefly, X-ray crystal structure of SQR indicates a Rossman-type fold with four major domains in flavin subunit and Tyr¹⁴² in the major helix (residues 136–158) of a floating subdomain (residues 105–196). Moreover, Tyr¹⁴² is highly surface exposed and situated in the hydrophilic environment, which suggests that this tyrosine should be susceptible to nitration with OONO⁻. X-ray crystal structure of SQR also indicates that Tyr¹⁴² is approximately 20 Å away from isoalloxazine ring of flavin adenine dinucleotide (FAD) and that Cys⁹⁰ is located within the part of the *N*-terminal beta barrel subdomain (residue 53–104) of the large FAD-binding domain, near the AMP moiety of FAD (~7.7 Å), where major catalysis of electron transfer and O₂⁻ production occurs. Thus, *S*-glutathionylation of Cys⁹⁰ seems likely to induce a conformational change near the floating subdomain (residues 105–196) to increase shielding effect on Tyr¹⁴² and to render Tyr¹⁴² less accessible to OONO⁻ oxidation [80]. Therefore, 3D structure of a protein can accurately interpret effects of tyrosine nitration on that protein's structure and functions.

67.7 PITFALLS OF NITROPROTEIN MEASUREMENT

Although measurement strategies of nitroproteins have extensively been developed, one must clearly realize that no highly reliable, high-sensitivity, high-throughput, and high-reproducibility approach exists to analyze the extremely challenging endogenous tyrosine nitration in a proteome. Therefore, different approaches are still under development. Those pitfalls include the following: (i) Nitrotyrosine antibody-based immunoaffinity methods such as 2D-Western blotting and NTAC succeeded to identify

endogenous nitrotyrosine sites; however, an overwhelming amount of nonnitrated tryptic peptides negatively affects nitropeptide characterization. For that reason, we suggest development of immunoaffinity enrichment of tryptic nitropeptides—not nitroproteins—prior to MS analysis. (ii) Until now, most methods based on chemical derivatization (as described earlier) are used only for *in vitro* experiments and not for endogenous nitrotyrosine sites. Although the COFRADIC-based characterization of nitropeptides succeeded in a serum proteome, sensitivity and throughput were very low, and it has not been used extensively in *endogenous* tissue nitroproteomes. Therefore, development of better nitrotyrosine analysis methods is necessary in the following aspects—alone or in combination: (i) derivatize a nitro to amino group to stabilize MS behaviors, (ii) develop specific amino group tags to enrich nitrotyrosine peptides, (iii) enrich nitrotyrosine—or aminotyrosine peptides are better than nitrotyrosine—or aminotyrosine proteins for sensitivity, (iv) improve liquid chromatography isolation, (v) develop super high-sensitivity mass spectrometers, (vi) choose an appropriate ion source and collision model to fragment nitropeptide or aminopeptides, and (vii) develop reliable software for data analysis. The combined multiple aspects among items (i)–(vii) are recommended to maximize coverage of endogenous nitrotyrosine sites in a proteome.

67.8 CONCLUSIONS

Protein tyrosine nitration is an important oxidative-/nitrative-mediated modification and associates a wide range of pathophysiological conditions [2–4, 7, 20, 57]. Moreover, evidence suggests the presence of a denitrase in mammalian tissues although a denitrase has not been isolated and its enzymatic activity not confirmed. Thus, tyrosine nitration can be considered as reversible; and tyrosine nitration is not only a result from oxidative damage, but it also participates in pathophysiological processes [106]. Nitration dynamically alters protein functions [107], including activation or inactivation [108–110]. MS-based identification of nitroproteins and nitrotyrosine sites is essential to understand biological roles of this modification [111–113]. However, it is analytically very challenging to identify *endogenous* nitroproteins and nitrotyrosine sites due to nitration's low abundance in biological samples and its multiple mass spectrometric behaviors among MALDI UV laser, ESI, CID, ECD, ETD, and MAD. Endogenous nitroproteins/nitropeptides must be enriched prior to MS analysis. Several enrichment methods have been developed, including immunoaffinity enrichment, biotin-affinity enrichment, and COFRADIC. Nitrotyrosine sites have been found in many different pathophysiological conditions. TMT- or iTRAQ-based quantitative nitroproteomics are needed to quantify disease-key nitroproteins/peptides. Protein domain/motif analysis, systems pathway analysis, and structural biological analysis of nitroproteins are significantly needed to elucidate the biological roles of tyrosine nitration.

NOMENCLATURE

1DE	one-dimensional electrophoresis
2D	two-dimensional
2DE	two-dimensional electrophoresis
2DGE	two-dimensional gel electrophoresis
3D	three-dimensional
APPD	(3R, 4S)-1-(4-(aminomethyl)phenylsulfonyl)pyrrolidine-3,4-diol
BSA	bovine serum albumin
CENT- β 1	centaurin beta 1
CID	collision-induced dissociation
COFRADIC	combined fractional diagonal chromatography
cPILOT	combined precursor isotopic labeling and isobaric tagging
ECD	electron-capture dissociation
ESI	electrospray ionization
ETD	electron-transfer dissociation
FAD	flavin adenine dinucleotide
FT-ICR	Fourier transform ion cyclotron resonance
GAPDH	glyceraldehyde-3-phosphate dehydrogenase
HPLC	high-performance liquid chromatography
IEF	isoelectric focusing
IL1	interleukin 1
IL1F6	interleukin 1 family member 6
IL1R	interleukin 1 receptor
IPA	Ingenuity Pathway Analysis
IR	infrared
IRAK2	interleukin 1 receptor-associated kinase-like 2
iTRAQ	isobaric tags for relative and absolute quantification
LC	liquid chromatography
LE1	leucine enkephalin
LE2	nitro-Tyr leucine enkephalin
LE3	<i>d</i> (5)-Phe-nitro-Tyr leucine enkephalin
LIRA4	leukocyte immunoglobulin-like receptor subfamily A member 4
MAD	metastable atom-activated dissociation
MALDI	matrix-assisted laser desorption/ionization
MAPK	mitogen-activated protein kinase
MS	mass spectrometry
MS/MS	tandem mass spectrometry
MW	molecular weight
<i>m/z</i>	mass-to-charge ratio
NAD ⁺	nicotinamide adenine dinucleotide
NHS acetate	acetic acid <i>N</i> -hydroxysuccinimide ester
NTAC	nitrotyrosine affinity column

OONO ⁻	peroxynitrite
OVA	ovalbumin
PKA	cAMP-dependent protein kinase
PKAR1- β	cAMP-dependent protein kinase type I-beta regulatory subunit
PMF	peptide mass fingerprint
PTM	posttranslational modification
RHOGAP5	Rho-GTPase-activating protein 5
S1P	sphingosine-1-phosphate
SCX	strong cation exchange
SDS-PAGE	sodium dodecyl sulfate polyacrylamide gel electrophoresis
SQR	succinate ubiquinone reductase
SRM	selected reaction monitoring
TGF β 1	transforming growth factor beta 1
TMT	tandem mass tags
TNF	tumor necrosis factor
TNM	tetranitromethane
TOF	time of flight
UV	ultraviolet
ZFP432	zinc finger protein 432

ACKNOWLEDGMENTS

The authors acknowledge the financial support from the National Natural Science Foundation of China (Grant No. 81272798 and 81572278 to X.Z.), the Xiangya Hospital Funds for Talent Introduction (to X.Z.), the Hunan Provincial Natural Science Foundation of China (Grant No. 14JJ7008 to X. Z.), China “863” Plan Project (Grant No. 2014AA020610-1 to X. Z.), and the National Institutes of Health, United States (RR16679; NS 42843 to D.M.D.).

REFERENCES

1. Scaloni A. 2006. Mass spectrometry approaches for the molecular characterization of oxidatively/nitrosatively modified proteins, in *Redox Proteomics: From Protein Modification to Cellular Dysfunction and Diseases* (Dalle-Donne I, Scaloni A, Butterfield DA, Eds.), John Wiley & Sons, Inc., Hoboken, NJ, pp. 59–100.
2. Zhan X, Desiderio DM. 2004. The human pituitary nitoproteome: detection of nitrotyrosyl-proteins with two-dimensional Western blotting, and amino acid sequence determination with mass spectrometry. *Biochem Biophys Res Commun* 325: 1180–1186.
3. Zhan X, Desiderio DM. 2006. Nitroproteins from human pituitary adenoma tissue discovered with a nitrotyrosine affinity column and tandem mass spectrometry. *Anal Biochem* 354: 279–289.

4. Zhan X, Desiderio DM. 2009a. Mass spectrometric identification of in vivo nitrotyrosine sites in the human pituitary tumor proteome. *Methods Mol Biol* 566: 137–63.
5. Ghesquiere B, Goethals M, Van Damme J, Staes A, Timmerman E, Vandekerckhove J, Gevaert K. 2006. Improved tandem mass spectrometric characterization of 3-nitrotyrosine sites in peptides. *Rapid Commun Mass Spectrom* 20: 2885–2893.
6. Yee CS, Seyedsayamdost MR, Chang MC, Nocera DG, Stubbe J. 2003. Generation of the R2 subunit of ribonucleotide reductase by intein chemistry: insertion of 3-nitrotyrosine at residue 356 as a probe of the radical initiation process. *Biochemistry* 42: 14541–14552.
7. Zhan X, Wang X, Desiderio DM. 2013. Pituitary adenoma nitroproteomics: current status and perspectives. *Oxid Med Cell Longev* 2013: 580710.
8. Irie Y, Saeki M, Kamisaki Y, Martin E, Murad F. 2003. Histone H1.2 is a substrate for denitrase, an activity that reduces nitrotyrosine immunoreactivity in proteins. *Proc Nat Acad Sci U S A* 100: 5634–5639.
9. Mallozzi C, D'Amore C, Camerini S, Macchia G, Crescenzi M, Petrucci TC, Di Stasi AM. 2013. Phosphorylation and nitration of tyrosine residues affect functional properties of synaptophysin and dynamin I, two proteins involved in exocytosis of synaptic vesicles. *Biochim Biophys Acta* 1833: 110–121.
10. Zhan X, Desiderio DM. 2011. Nitroproteins identified in human ex-smoker bronchoalveolar lavage fluid. *Aging Dis* 2: 100–115.
11. Zhan X, Du Y, Crabb JS, Kern TS, Crabb JW. 2007. Identification of nitrated proteins in diabetic rat retina. *Invest Ophthalmol Vis Sci* 48: E-abstract 4962.
12. Aulak KS, Koeck T, Crabb JW, Stuehr DJ. 2004. Dynamics of protein nitration in cells and mitochondria. *Am J Physiol* 286: H30–H38.
13. Koeck T, Fu X, Hazen SL, Crabb JW, Stuehr DJ, Aulak KS. 2004. Rapid and selective oxygen-regulated protein tyrosine denitration and nitration in mitochondria. *J Biol Chem* 279: 27257–27262.
14. Haddad IY, Pataki G, Hu P, Galliani C, Beckman JS, Matalon S. 1994. Quantitation of nitrotyrosine levels in lung sections of patients and animals with acute lung injury. *J Clin Invest* 94: 2407–2413.
15. Halliwell B, Zhao K, Whiteman M. 1999. Nitric oxide and peroxynitrite. The ugly, the uglier and the not so good: a personal view of recent controversies. *Free Radic Res* 31: 651–669.
16. Miyagi M, Sakaguchi H, Darrow RM, Yan L, West KA, Aulak KS, Stuehr DJ, Hollyfield JG, Organisciak DT, Crabb JW. 2002. Evidence that light modulates protein nitration in rat retina. *Mol Cell Proteomics* 1: 293–303.
17. Yeo WS, Lee SJ, Lee JR, Kim KP. 2008. Nitrosative protein tyrosine modifications: biochemistry and functional significance. *BMB Rep* 41: 194–203.
18. Petersson AS, Steen H, Kalume DE, Caidahl K, Roepstorff P. 2001. Investigation of tyrosine nitration in proteins by mass spectrometry. *J Mass Spectrom* 36: 616–625.
19. Sarver A, Scheffler K, Shetlar MD, Gibson BW. 2001. Analysis of peptides and proteins containing nitrotyrosine by matrix-assisted laser desorption/ionization mass spectrometry. *J Am Soc Mass Spectrom* 12: 439–448.

20. Zhan X, Desiderio DM. 2009b. MALDI-induced fragmentation of leucine enkephalin, nitro-Tyr leucine enkephalin, and *d*(5)-Phe-nitro-Tyr leucine enkephalin. *Int J Mass Spectrom* 287: 77–86.
21. Kim JK, Lee JR, Kang JW, Lee SJ, Shin GC, Yeo WS, Kim KH, Park HS, Kim KP. 2011. Selective enrichment and mass spectrometric identification of nitrated peptides using fluorinated carbon tags. *Anal Chem* 83: 157–163.
22. Lee JR, Lee SJ, Kim TW, Kim JK, Park HS, Kim DE, Kim KP, Yeo WS. 2009b. Chemical approach for specific enrichment and mass analysis of nitrated peptides. *Analyt Chem* 81: 6620–6629.
23. Lee SJ, Lee JR, Kim YH, Park YS, Park SI, Park HS, Kim KP. 2007. Investigation of tyrosine nitration and nitrosylation of angiotensin II and bovine serum albumin with electrospray ionization mass spectrometry. *Rapid Commun Mass Spectrom* 21: 2797–2804.
24. Zhang Q, Qian WJ, Knyushko TV, Clauss TR, Purvine SO, Moore RJ, Sacksteder CA, Chin MH, Smith DJ, Camp DG 2nd, Bigelow DJ, Smith RD. 2007. A method for selective enrichment and analysis of nitrotyrosine-containing peptides in complex proteome samples. *J Proteom Res* 6: 2257–2268.
25. Dekker F, Abello N, Wisastra R, Bischoff R. 2012. Enrichment and detection of tyrosine-nitrated proteins. *Curr Protoc Protein Sci* DOI:10.1002/0471140864.ps1413s69.
26. Feeney MB, Schöneich C. 2013. Proteomic approaches to analyze protein tyrosine nitration. *Antioxid Redox Signal* 19: 1247–1256.
27. Dr Gusanu M, Petre BA, Przybylski M. 2011. Epitope motif of an anti-nitrotyrosine antibody specific for tyrosine-nitrated peptides revealed by a combination of affinity approaches and mass spectrometry. *J Pept Sci* 17: 184–191.
28. Sultana R, Reed T, Butterfield DA. 2009. Detection of 4-hydroxy-2-nonenal- and 3-nitrotyrosine-modified proteins using a proteomics approach. *Methods Mol Biol* 519: 351–361.
29. Abello N, Barroso B, Kerstjens HAM, Postma DS, Bischoff R. 2010. Chemical labeling and enrichment of nitrotyrosine-containing peptides. *Talanta* 80: 1503–1512.
30. Tsumoto H, Taguchi R, Kohda K. 2010. Efficient identification and quantification of peptides containing nitrotyrosine by matrix-assisted laser desorption/ionization time-of-flight mass spectrometry after derivation. *Chem Pharm Bull* 58: 488–494.
31. Amoresano A, Chiappetta G, Pucci P, D’Ischia M, Marino G. 2007. Bidimensional tandem mass spectrometry for selective identification of nitration sites in proteins. *Anal Chem* 79: 2109–2117.
32. Amoresano A, Chiappetta G, Pucci P, Marino G. 2008. A rapid and selective mass spectrometric method for the identification of nitrated proteins. *Methods Mol Biol* 477: 15–29.
33. Robinson RA, Evans AR. 2012. Enhanced sample multiplexing for nitrotyrosine-modified proteins using combined precursor isotopic labeling and isobaric tagging. *Anal Chem* 84: 4677–4686.
34. Prokai-Tatrai K, Guo J, Prokai L. 2011. Selective chemoprecipitation and subsequent release of tagged species for the analysis of nitropeptides by liquid chromatography-tandem mass spectrometry. *Mol Cell Proteomics* 10 DOI:10.1074/mcp.M110.002923.

35. Chiappetta G, Corbo C, Palmese A, Galli F, Piroddi M, Marino G, Amoresano A. 2009. Quantitative identification of protein nitration sites. *Proteomics* 9: 1524–1537.
36. Ghesquiere B, Colaert N, Helsens K, Dejager L, Vanhaute C, Verleysen K, Kas K, Timmerman E, Goethals M, Libert C, Vandekerckhove J, Gevaert K. 2009. In vitro and in vivo protein-bound tyrosine nitration characterized by diagonal chromatography. *Mol Cell Proteomics* 8: 2642–2652.
37. Larsen TR, Bache N, Gramsbergen JB, Roepstorff P. 2011. Identification of nitrotyrosine containing peptides using combined fractional diagonal chromatography (COFRADIC) and off-line nano-LC-MALDI. *J Am Soc Mass Spectrom* 22: 989–996.
38. Zhang Y, Yang H, Posch U. 2011. Analysis of nitrated proteins and tryptic peptides by HPLC-chip-MS/MS: site-specific quantification, nitration degree, and reactivity of tyrosine residues. *Anal Bioanal Chem* 399: 459–471.
39. Zhan X, Desiderio DM. 2010a. Signaling pathway networks mined from human pituitary adenoma proteomics data. *BMC Med Genomics* 3: 13.
40. Palamalai V, Miyagi M. 2010. Mechanism of glyceraldehyde-3-phosphate dehydrogenase inactivation by tyrosine nitration. *Protein Sci* 19: 255–262.
41. Seeley KW, Stevens SM Jr. 2012. Investigation of local primary structure effects on peroxynitrite-mediated tyrosine nitration using targeted mass spectrometry. *J Proteomics* 75: 1691–1700.
42. Zhan X, Desiderio DM. 2010b. The use of variations in proteomes to predict, prevent, and personalize treatment for clinically nonfunctional pituitary adenomas. *EPMA J* 1: 439–459.
43. Turko IV, Murad F. 2005. Mapping sites of tyrosine nitration by matrix-assisted laser desorption/ionization mass spectrometry. *Methods Enzymol* 396: 266–275.
44. Petre BA, Youhnovski N, Lukkari J, Weber R, Przybylski M. 2005. Structural characterization of tyrosine-nitrated peptides by ultraviolet and infrared matrix-assisted laser desorption/ionization Fourier transform ion cyclotron resonance mass spectrometry. *Eur J Mass Spectrom* 11: 513–518.
45. Lee HM, Reed J, Greeley GH Jr, Englander EW. 2009a. Impaired mitochondrial respiration and protein nitration in the rat hippocampus after acute inhalation of combustion smoke. *Toxicol Appl Pharmacol* 235: 208–215.
46. Cook SL, Jackson GP. 2011. Characterization of tyrosine nitration and cysteine nitrosylation modifications by metastable atom-activation dissociation mass spectrometry. *J Am Soc Mass Spectrom* 22: 221–232.
47. Jones AW, Mikhailov VA, Iniesta J, Cooper HJ. 2010. Electron capture dissociation mass spectrometry of tyrosine nitrated peptides. *J Am Soc Mass Spectrom* 21: 268–277.
48. Jones AW, Cooper HJ. 2010. Probing the mechanisms of electron capture dissociation mass spectrometry with nitrated peptides. *Phys Chem Chem Phys* 12: 13394–13399.
49. Li B, Held JM, Schilling B, Danielson SR, Gibson BW. 2011. Confident identification of 3-nitrotyrosine modifications in mass spectral data across multiple mass spectrometry platforms. *J Proteomics* 74: 2510–2521.
50. Sokolovsky M, Riordan JF, Vallee BL. 1966. Tetranitromethane. A reagent for the nitration of tyrosyl residues in proteins. *Biochemistry* 5: 3582–3589.

51. Fujigaki H, Saito K, Lin F, Fujigaki S, Takahashi K, Martin BM, Chen CY, Masuda J, Kowalak J, Takikawa O, Seishima M, Markey SP. 2006. Nitration and inactivation of IDO by peroxynitrite. *J Immunology* 176: 372–379.
52. Sokolovsky M, Riordan JF, Vallee BL. 1967. Conversion of 3-nitrotyrosine to 3-aminotyrosine in peptides and proteins. *Biochem Biophys Res Commun* 27: 20–25.
53. Abello N, Kerstjens HA, Postma DS, Bischoff R. 2009. Protein tyrosine nitration: selectivity, physicochemical and biological consequences, denitration, and proteomics methods for the identification of tyrosine-nitrated proteins. *J Proteome Res* 8: 3222–3238.
54. Sheeley SA, Rubakhin SS, Sweedler JV. 2005. The detection of nitrated tyrosine in neuropeptides: a MALDI matrix-dependent response. *Anal Bioanal Chem* 382: 22–27.
55. Ghesquiere B, Helsen K, Vandekerckhove J, Gevaert K. 2011. A stringent approach to improve the quality of nitrotyrosine peptide identifications. *Proteomics* 11: 1094–1098.
56. Shigenaga MK, Lee HH, Blunt BC, Christen S, Shigeno ET, Yip H, Ames BN. 1997. Inflammation and NO(X)-induced nitration: assay for 3-nitrotyrosine by HPLC with electrochemical detection. *Proc Natl Acad Sci U S A* 94: 3211–3216.
57. Zhan X, Desiderio DM. 2007. Linear ion-trap mass spectrometric characterization of human pituitary nitrotyrosine containing proteins. *Int J Mass Spectrom* 259: 96–104.
58. Aulak KS, Miyagi M, Yan L, West KA, Massillon D, Crabb JW, Stuehr DJ. 2001. Proteomic method identifies proteins nitrated in vivo during inflammatory challenge. *Proc Natl Acad Sci U S A* 98: 12056–12061.
59. Butt YK, Lo SC. 2008. Detecting nitrated proteins by proteomic technologies. *Methods Enzymol* 440: 17–31.
60. Justilien V, Pang JJ, Renganathan K, Zhan X, Crabb JW, Kim SR, Sparrow JR, Hauswirth WW, Lewin AS. 2007. SOD2 knockdown mouse model of early AMD. *Invest Ophthalmol Vis Sci* 48: 4407–4420.
61. Petre BA, Ulrich M, Stumbaum M, Bernevic B, Moise A, Döring G, Przybylski M. 2012. When is mass spectrometry combined with affinity approaches essential? A case study of tyrosine nitration in proteins. *J Am Soc Mass Spectrom* 23: 1831–1840.
62. Dremina ES, Li X, Galeva NA, Sharov VS, Stobaugh JF, Schöneich C. 2011. A methodology for simultaneous fluorogenic derivatization and boronate affinity enrichment of 3-nitrotyrosine containing peptides. *Anal Biochem* 418: 184–196.
63. Dalle-Donne I, Scaloni A, Butterfield DA (Eds.). 2006. *Redox Proteomics: From Protein Modifications to Cellular Dysfunction and Diseases*, John Wiley & Sons, Inc., Hoboken, NJ.
64. Dalle-Donne I, Scaloni A, Giustarini D, Cavarra E, Tell G, Lungarella G, Colombo R, Rossi R, Milzani A. 2005. Proteins as biomarkers of oxidative/nitrosative stress in diseases: the contribution of redox proteomics. *Mass Spectrom Rev* 24: 55–99.
65. Lanone S, Manivet P, Callebort J, Launay JM, Payen D, Aubier M, Boczkowski J, Mebazaa A. 2002. Inducible nitric oxide synthase (NOS2) expressed in septic patients is nitrated on selected tyrosine residues: implications for enzymic activity. *Biochem J* 366: 399–404.
66. Ghosh S, Janocha AJ, Aronica MA, Swaidani S, Comhair SAA, Xu W, Zheng L, Kaveti S, Kinter M, Hazen SL, Erzurum SC. 2006. Nitrotyrosine proteome survey in asthma identifies oxidative mechanism of catalase inactivation. *J Immunology* 176: 5587–5597.

67. Dhiman M, Nakayasu ES, Madaiah YH, Reynolds BK, Wen JJ, Almeida IC, Garg NJ. 2008. Enhanced nitrosative stress during *Trypanosoma cruzi* infection causes nitrotyrosine modification of host proteins: implications in Chagas' disease. *Am J Pathol* 173: 728–740.
68. Chatterjee S, Lardinois O, Bonini MG, Bhattacharjee S, Stadler K, Corbett J, Deterding LJ, Tomer KB, Kadiiska M, Mason RP. 2009. Site-specific carboxypeptidase B1 tyrosine nitration and pathophysiological implications following its physical association with nitric oxide synthase-3 in experimental sepsis. *J Immunol* 183: 4055–4066.
69. Kanski J, Behring A, Pelling J, Schöneich C. 2005a. Proteomic identification of 3-nitrotyrosine-containing rat cardiac proteins: effects of biological aging. *Am J Physiol Heart Circ Physiol* 288: H371–H381.
70. Kanski J, Hong SJ, Schöneich C. 2005b. Proteomic analysis of protein nitration in aging skeletal muscle and identification of nitrotyrosine-containing sequences in vivo by nano-electrospray ionization tandem mass spectrometry. *J Biol Chem* 280: 24261–24266.
71. Sharov VS, Galeva NA, Kanski J, Williams TD, Schöneich C. 2006. Age-associated tyrosine nitration of rat skeletal muscle glycogen phosphorylase b: characterization by HPLC-nano-electrospray-tandem mass spectrometry. *Exp Gerontol* 41: 407–416.
72. Marshall A, Lutfeali R, Raval A, Chakravarti DN, Chakravarti B. 2013. Differential hepatic protein tyrosine nitration of mouse due to aging-effect on mitochondrial energy metabolism, quality control machinery of the endoplasmic reticulum and metabolism of drugs. *Biochem Biophys Res Commun* 430: 231–235.
73. Fiorce G, Di Cristo C, Monti G, Amoresano A, Columbano L, Pucci P, Cioffi FA, Cosmo AD, Palumbo A, d'Ischia M. 2006. Tubulin nitration in human gliomas. *Neurosci Lett* 394: 57–62.
74. Nakagawa H, Komai N, Takusagawa M, Miura Y, Toda T, Miyata N, Ozawa T, Ikota N. 2007. Nitration of specific tyrosine residues of cytochrome C is associated with caspase-cascade inactivation. *Biol Pharm Bull* 30: 15–20.
75. Casoni F, Basso M, Massignan T, Gianazza E, Cheroni C, Salmona M, Bendotti C, Bonetto V. 2005. Protein nitration in a mouse model of familial amyotrophic lateral sclerosis: possible multifunctional role in the pathogenesis. *J Biol Chem* 280: 16295–16304.
76. Sacksteder CA, Qian WJ, Knyushko TV, Wang HW, Chin MH, Lacan G, Melega WP, Camp DG 2nd, Smith RD, Smith DJ, Squier TC, Bigelow DJ. 2006. Endogenously nitrated proteins in mouse brain: links to neurodegenerative disease. *Biochemistry* 45: 8009–8022.
77. Danielson SR, Held JM, Schilling B, Oo M, Gibson BW, Andersen JK. 2009. Preferentially increased nitration of alpha-synuclein at tyrosine-39 in a cellular oxidative model of Parkinson's disease. *Anal Chem* 81: 7823–7828.
78. Yoon SW, Kang S, Ryu SE, Poo H. 2010. Identification of tyrosine-nitrated proteins in HT22 hippocampal cells during glutamate-induced oxidative stress. *Cell Prolif* 43: 584–593.
79. Zhang X, Monroe ME, Chen B, Chin MH, Heibeck TH, Schepmoes AA, Yang F, Petritis BO, Camp DG 2nd, Pounds JG, Jacobs JM, Smith DJ, Bigelow DJ, Smith RD, Qian WJ. 2010. Endogenous 3,4-dihydroxyphenylalanine and dopaquinone modifications on protein tyrosine: links to mitochondrially derived oxidative stress via hydroxyl radical. *Mol Cell Proteomics* 9: 1199–1208.

80. Chen CL, Chen J, Rawale S, Varadharaj S, Kaumaya PPT, Zweier JL, Chen YR. 2008. Protein tyrosine nitration of the Flavin subunit is associated with oxidative modification of mitochondrial complex II in the post-ischemic myocardium. *J Biol Chem* 283: 27991–28003.
81. Ai L, Rouhanizadeh M, Wu JC, Takabe W, Yu H, Alavi M, Chu Y, Miller J, Heistad DD, Hsiai TK. 2008. Shear stress influences spatial variations in vascular Mn-SOD expression. *Am J Physiol Cell Physiol* 294: C1576–C1585.
82. Liu B, Tewari AK, Zhang L, Green-Church KB, Zweier JL, Chen YR, He G. 2009. Proteomic analysis of protein tyrosine nitration after ischemia reperfusion injury: mitochondria as the major target. *Biochem Biophys Acta* 1794: 476–485.
83. Palamalai V, Darrow RM, Organisciak DT, Miyagi M. 2006. Light-induced changes in protein nitration in photoreceptor rod outer segments. *Mol Vis* 12: 1543–1551.
84. Murdaugh LS, Wang Z, Del Priore LV, Dillon J, Gaillard ER. 2010. Age-related accumulation of 3-nitrotyrosine and nitro-A2E in human Bruch's membrane. *Exp Eye Res* 90: 564–571.
85. Kato Y, Dozaki N, Nakamura T, Kitamoto N, Yoshida A, Naito M, Kitamura M, Osawa T. 2009. Quantification of modified tyrosines in healthy and diabetic human urine using liquid chromatography/tandem mass spectrometry. *J Clin Biochem Nutr* 44: 67–78.
86. Piroddi M, Palmese A, Pilolli F, Amoresano A, Pucci P, Ronco C, Galli F. 2011. Plasma nitroproteome of kidney disease patients. *Amino Acids* 40: 653–667.
87. Chaki M, Valderrama R, Fernández-Ocaña AM, Carreras A, López-Jaramillo J, Luque F, Palma JM, Pedrajas JR, Begara-Morales JC, Sánchez-Calvo B, Gómez-Rodríguez MV, Corpas FJ, Barroso JB. 2009. Protein targets of tyrosine nitration in sunflower (*Helianthus annuus* L.) hypocotyls. *J Exp Bot* 60: 4221–4234.
88. Aslan M, Ryan TM, Townes TM, Coward L, Kirk MC, Barnes S, Alexander CB, Rosenfeld SS, Freeman BA. 2003. Nitric oxide-dependent generation of reactive species in sickle cell disease. Actin tyrosine induces defective cytoskeletal polymerization. *J Biol Chem* 278: 4194–4204.
89. Webster RP, Brockman D, Myatt L. 2006. Nitration of p38 MAPK in the placenta: association of nitration with reduced catalytic activity of p38 MAPK in pre-eclampsia. *Mol Hum Reprod* 12: 677–685.
90. Ulrich M, Petre A, Youhnovski N, Prömm F, Schirle M, Schumm M, Pero RS, Doyle A, Checkel J, Kita H, Thiagarajan N, Acharya KR, Schmid-Grendelmeier P, Simon HU, Schwarz H, Tsutsui M, Shimokawa H, Bellon G, Lee JJ, Przybylski M, Döring G. 2008. Post-translational tyrosine nitration of eosinophil granule toxins mediated by eosinophil peroxidase. *J Biol Chem* 283: 28629–28640.
91. Hamilton RT, Asatryan L, Nilsen JT, Isas JM, Gallaher TK, Sawamura T, Hsiai TK. 2008. LDL protein nitration: implication for LDL protein unfolding. *Arch Biochem Biophys* 479: 1–14.
92. Zhu JH, Zhang X, Roneker CA, McClung JP, Zhang S, Thannhauser TW, Ripoll DR, Sun Q, Lei XG. 2008. Role of copper, zinc-superoxide dismutase in catalyzing nitrotyrosine formation in murine liver. *Free Radic Biol Med* 45: 611–618.
93. Reed TT, Owen J, Pierce WM, Sebastian A, Sullivan PG, Butterfield DA. 2009. Proteomic identification of nitrated brain proteins in traumatic brain-injured rats treated postinjury with gamma-glutamylcysteine ethylester: insights into the role of elevation of glutathione as a potential therapeutic strategy for traumatic brain injury. *J Neurosci Res* 87: 408–417.

94. Casanovas A, Carrascal M, Abián J, López-Tejero MD, Llobera M. 2009. Lipoprotein lipase is nitrated in vivo after lipopolysaccharide challenge. *Free Radic Biol Med* 47: 1553–1560.
95. Sharov VS, Galeva NA, Dremina ES, Williams TD, Schöneich C. 2009. Inactivation of rabbit muscle glycogen phosphorylase b by peroxynitrite revisited: does the nitration of Tyr613 in the allosteric inhibition site control enzymatic function? *Arch Biochem Biophys* 484: 155–166.
96. Ohama T, Brautigan DL. 2010. Endotoxin conditioning induces VCP/p97-mediated and inducible nitric-oxide synthase-dependent Tyr284 nitration in protein phosphatase 2A. *J Biol Chem* 285: 8711–8718.
97. Sekar Y, Moon TC, Slupsky CM, Befus AD. 2010. Protein tyrosine nitration of aldolase in mast cells: a plausible pathway in nitric oxide-mediated regulation of mast cell function. *J Immunol* 185: 578–587.
98. Redondo-Horcajo M, Romero N, Martínez-Acedo P, Martínez-Ruiz A, Quijano C, Lourenço CF, Movilla N, Enríquez JA, Rodríguez-Pascual F, Rial E, Radi R, Vázquez J, Lamas S. 2010. Cyclosporine A-induced nitration of tyrosine 34 MnSOD in endothelial cells: role of mitochondrial superoxide. *Cardiovasc Res* 87: 356–365.
99. Chen HJ, Chen YC. 2012. Reactive nitrogen oxide species-induced post-translational modifications in human hemoglobin and the association with cigarette smoking. *Anal Chem* 84: 7881–7890.
100. Bigelow DJ, Qian WJ. 2008. Quantitative proteome mapping of nitrotyrosines. *Methods Enzymol* 440: 191–205.
101. Tao RR, Huang JY, Shao XJ, Ye WF, Tian Y, Liao MH, Fukunaga K, Lou YJ, Han F, Lu YM. 2013. Ischemic injury promotes Keap1 nitration and disturbance of antioxidative responses in endothelial cells: a potential vasoprotective effect of melatonin. *J Pineal Res* 54: 271–281.
102. Safinowski M, Wilhelm B, Reimer T, Weise A, Thomé N, Hänel H, Forst T, Pfützner A. 2009. Determination of nitrotyrosine concentrations in plasma samples of diabetes mellitus patients by four different immunoassays leads to contradictory results and disqualifies the majority of the tests. *Clin Chem Lab Med* 47: 483–488.
103. Lu N, Zhang Y, Li H, Gao Z. 2010. Oxidative and nitrative modifications of alpha-enolase in cardiac proteins from diabetic rats. *Free Radic Biol Med* 48: 873–881.
104. Cecconi D, Orzetti S, Vandelle E, Rinalducci S, Zolla L, Delledonne M. 2009. Protein nitration during defense response in *Arabidopsis thaliana*. *Electrophoresis* 30: 2460–2468.
105. Hui Y, Wong M, Zhao SS, Love JA, Ansley DM, Chen DD. 2012. A simple and robust LC-MS/MS method for quantification of free 3-nitrotyrosine in human plasma from patients receiving on-pump CABG surgery. *Electrophoresis* 33: 697–704.
106. Smallwood HS, Lourette NM, Boschek CB, Bigelow DJ, Smith RD, Pasa-Tolic L, Squier TC. 2007. Identification of a denitrase activity against calmodulin in activated macrophages using high-field liquid chromatography—FTICR mass spectrometry. *Biochemistry* 46: 10498–10505.

107. Mani AR, Moore KP. 2005. Dynamic assessment of nitration reactions in vivo. *Methods Enzymol* 396: 151–159.
108. Lin HL, Kenaan C, Zhang H, Hollenberg PF. 2012. Reaction of human cytochrome P450 3A4 with peroxynitrite: nitrotyrosine formation on the proximal side impairs its interaction with NADPH-cytochrome P450 reductase. *Chem Res Toxicol* 25: 2642–2653.
109. Lin HL, Myshkin E, Waskell L, Hollenberg PF. 2007. Peroxynitrite inactivation of human cytochrome P450 2B6 and 2E1: heme modification and site-specific nitrotyrosine formation. *Chem Res Toxicol* 20: 1612–1622.
110. Yamakura F, Kawasaki H. 2010. Post-translational modifications of superoxide dismutase. *Biochim Biophys Acta Proteins Proteomics* 1804: 318–325.
111. Kanski J, Schöneich C. 2005. Protein nitration in biological aging: proteomic and tandem mass spectrometric characterization of nitrated sites. *Methods Enzymol* 396: 160–171.
112. Spickett CM, Pitt AR. 2012. Protein oxidation: role in signaling and detection by mass spectrometry. *Amino Acids* 42: 5–21.
113. Tsikas D. 2012. Analytical methods for 3-nitrotyrosine quantification in biological samples: the unique role of tandem mass spectrometry. *Amino Acids* 42: 45–63.

FLUORESCENCE SPECTROSCOPY

YEVGEN POVROZIN AND BENIAMINO BARBIERI

ISS, Champaign, IL, USA

The number of fluorescence technique applications has been continuously growing over the last 20 years. While initially intended as an analytical tool for the determination of the presence of specific molecules in solutions, fluorescence is now routinely used in biochemistry and biophysics for studying molecular interactions and dynamics, both in solutions and in cells; in clinical immunoassays for the determination of the presence of specific antibodies and antigens; in drug discovery; in life sciences for DNA sequencing; and in nanotechnology and materials sciences for identification and characterization of new materials.

The reasons of the continuing increase in popularity are multiple: on one hand, it is due to the improvements in the sensitivity of the instrumentation that allows now for the observation of single-molecule events on a routine basis; on the other hand, the interface of the instrumentation with the personal computer has increased the automation of the data collection and the sophistication of the data analysis. A third reason for its increased success is due to the introduction in the past 30 years of innumerable and specific chemical probes used as markers for compounds that either do not display fluorescence or only emit a low level of it. The extent of the applications has benefited from the development of the green fluorescent protein (GFP) family that allows for the expression of fluorescent proteins in cells and tissues, a feature that allows the experimenter to follow the whereabouts of proteins in live cells and even tissues in live animals.

Paradoxically, the capabilities of the instrumentation coupled to the computation power of the computer brings new challenges to the field, as novel practitioners are not always aware of the potential pitfalls that lie behind an experiment. In the past few years several articles and books have been published on the subject describing in detail the applications of the fluorescence techniques to the chemical and life sciences. A brief article cannot cover such details; our goal is rather to reiterate the fundamental principles of the technique and to mention some of the common pitfall that a user of the technique may encounter.

68.1 OBSERVABLES MEASURED IN FLUORESCENCE

Fluorescence is generally referred to as the emission of photons from a sample following the absorption of photons. There are other means for producing fluorescence in a sample (bioluminescence, sonoluminescence, and electroluminescence), but in the following we will refer exclusively to the phenomenon originated by the absorption of light.

Fluorescence is part of a general class of phenomena named *luminescence*; it is distinguished by the *phosphorescence* as the latter takes, typically, a time of the order of 1 μ s (10^{-6} s) or longer, while the former takes a time of the order of 1 ns (10^{-9} s). As we will see in the following, the distinction between the two is described using the more precise terminology of quantum mechanics.

The main five parameters measured in fluorescence spectroscopy are:

1. Excitation spectrum
2. Emission spectrum
3. Decay times (fluorescence lifetimes)
4. Quantum yield
5. Anisotropy (or polarization)

Recent advancements in fluorescence microscopy have introduced the measurement of additional parameters (diffusion correlation times, brightness), but we will limit our discussion in this chapter to the five parameters listed earlier, which are measurable using a spectrofluorometer.

The description of the fluorescence measurable parameters is best understood with the introduction of the Perrin–Jabłoński diagram, which is a quantum mechanics representation of the energy levels of a molecular structure.

68.2 THE PERRIN–JABŁOŃSKI DIAGRAM

Figure 68.1 shows a classic representation of the electronic levels of a molecule in solution or in the gas phase (in solid phase the energy levels collapse into “bands” although the basic concepts are still valid).

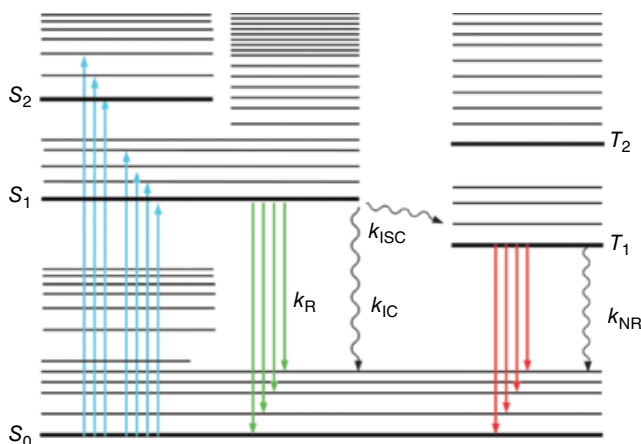


FIGURE 68.1 Perrin-Jabłoński energy diagram for a molecular structure. Singlet states are indicated by S_0, S_1, \dots , and triplet states by T_1, T_2, \dots . Internal conversion rate is k_{IC} ; intercrossing conversion rate between singlet and triplet states is k_{ISC} ; the fluorescence decay rate is k_R , while the nonfluorescence rate is k_{NR} .

The energy levels occupied by an electron are named “singlet states,” and the letters S_0, S_1, S_2, \dots , indicate the ground state, the first excited state, etc.; upon absorption of a photon, an electron moves from the ground state S_0 to the excited states. Associated with each electronic level, there are several vibrational and rotational levels, which differ in energy by a smaller amount than the corresponding electronic levels.

Moreover, there are energy transitions that are not directly allowed (forbidden transitions). They are identified as “triplet states” and indicated by T_1, T_2, \dots , etc.; they also feature associated vibrational and rotational levels.

The absorption probability of a photon in each electronic level is described within the framework of quantum mechanics (energy separation between the levels and momentum and spin of the various levels). The molecules interact when in the presence of photons of the appropriate photon energy E , where

$$E = h\nu = h \frac{c}{\lambda} \quad (68.1)$$

In the relation, h is the Planck constant ($6.626 \times 10^{-34} \text{ Js}$), c is the speed of light ($2.9979 \times 10^8 \text{ ms}^{-1}$), while ν and λ are the frequency and wavelength of the electromagnetic wave describing the photon.

For absorption to occur, E has to be of the order of magnitude of the separation between the excited level and the ground state; that is,

$$E \approx E_{S_1} - E_{S_0} \quad (68.2)$$

Let us consider a population of N molecules in a solution. Upon absorption of photons (group of line to the left in the figure), a fraction of the molecules undergo a

transition from the ground state S_0 to the upper electronic states, S_1 , S_2 , the final state depending ultimately by the energy of the absorbed photon. The absorption process takes an amount of time of the order of the femtosecond (10^{-15} s) or shorter.

Once in the excited electronic level, the molecules relax fairly rapidly (about 10^{-12} s) to the lowest level of the first excited state S_1 ; hence, they decay with rate k_R to emit fluorescence (group of lines in the middle in the figure). The characteristic time of the fluorescence is of the order of 1 ns (10^{-9} s).

There are additional decay routes that are not necessarily associated with the emission of photons; they are indicated by k_{IC} (internal conversion between two electronic states of the same spin multiplicity) and k_{ISC} (intersystem crossing conversion between the S levels and T levels). It is noteworthy that the excited level T_1 (triplet state) emits photons; this process is usually termed “phosphorescence,” and its characteristic time, as mentioned previously, is of the order of 1 μ s (10^{-6} s) and longer.

The Perrin–Jabłoński diagram (Fig. 68.1) is instrumental to determine the law describing the decay time of fluorescence. If N_1 is the population of the excited level S_1 , upon absorption of photons the population of the level changes is described by the relation

$$\frac{dN_1}{dt} = -(k_R + k_{NR})N_1 + f_1 \quad (68.3)$$

where f_1 is a function that describes the process of the excitation photons (pulsed source, continuous-wave (CW) source, etc.). By solving the equation (and disregarding f_1), we find that

$$N_1 = N_1(0)e^{-(t/\tau_s)} \quad (68.4)$$

where τ_s the decay time of the excited state S_1 is defined as

$$\tau_s = \frac{1}{k_R + k_{NR}} \quad (68.5)$$

The fluorescence quantum yield is the fraction of excited molecules that return to the ground state with the emission of fluorescence. From direct examination of the Perrin–Jabłoński diagram, one simply divides the rate of radiative emission k_R by the total rates of deactivation, which includes both the radiative and nonradiative contributions:

$$\Phi = \frac{k_R}{k_R + k_{NR}} \quad (68.6)$$

By using the definition of decay times, the quantum yield can also be expressed in terms of lifetimes:

$$\Phi = \frac{\tau_S}{\tau_R} \quad (68.7)$$

One can say that the quantum yield is the ratio of the number of emitted photons over the total number of absorbed photons.

The five measurable parameters of fluorescence are usually used to describe these processes, namely, the range in wavelengths of the absorption and emission of photons (excitation and emission spectra), the orientation changes during the time the molecules are in the excited states between absorption and emission of the photons (anisotropy or polarization), the fraction of photons emitted over the number of photons absorbed (quantum yield), and the emission rate (decay times). After a brief overview of the instrumentation, we will examine in detail the measurement of the five parameters.

68.3 INSTRUMENTATION

The peculiar parameters that characterize fluorescence are measured using “spectrofluorometers”; sometimes, instruments for the measurement of excitation and emission spectra are termed “spectrofluorimeters,” while the ones for the measurements of the decay times are termed “spectrofluorometers.” Yet, the distinction is not anymore as clearly demarked as several instruments allow, in the same unit, to measure both the steady-state (excitation and emission spectra) and the dynamic (decay times and rotational correlation times) properties of the fluorescence.

Usually, in all of the instruments, the fluorescence is collected at an angle of 90° with respect to the optical axis set by the excitation light beam. This geometry maximizes the efficiency of the emission collection and reduces the background due to the excitation light.

It is worthy to mention that absorption spectra can be measured using a spectrophotometer. In this type of instrument, the light detector is placed on the same optical axis of the excitation light beam, and the instrument detects the amount of light that is being transmitted (i.e., not absorbed) through the sample. A spectrophotometer measures the difference in the intensity of two signals (typically, sample transmittance is compared to 100% transmittance); instead, a spectrofluorometer measures a signal (the fluorescence) over a zero background.

The key elements of a spectrofluorometer are the light source, the monochromator, and the light detector.

68.3.1 Light Source

The typical light source utilized in a spectrofluorometer is a high-pressure xenon arc lamp. The bulb of this lamp includes xenon at high pressure that is excited to higher level by the electrical arc established by the current running through the electrodes. The emitted light is a continuous spectrum from (depending upon the models and geometries) about 250 up to 1100 nm. Figure 68.2 displays the spectrum of the lamp utilized by ISS. Although the spectrum is relatively flat up to about 800 nm, several sharp resonances are present above that wavelength.

It is worth noting that a variation of this lamp is the Hg–Xe lamp, which contains traces of mercury; this element displays resonances at around 295 nm, and this feature allowed for its use as an excitation source for the proteins containing tryptophan.

In the past several years, lasers have replaced the xenon arc lamp, specifically for time-resolved applications. Although they emit radiation only at specific wavelengths, their brightness is order of magnitude higher than that of the lamp. In addition, they can be pulsed with fairly narrow pulse widths (about 50 ps for the laser diodes). A recent advancement is the supercontinuum laser (or white laser) that delivers any wavelength in the range from 390 up to 2000 nm, featuring 5 ps pulse width and (in the model made by Fianium Ltd) the option of selecting the repetition rate up to 40 MHz.

Light-emitting diodes (LEDs) are also utilized as light sources especially in the region from 240 to 350 nm, where lasers are not available (with exceptions at 266, 315, 325 nm). They are compact, relatively inexpensive, and the source of choice when building an instrument dedicated to a specific application.

68.3.2 Monochromator

Monochromators are utilized to select the wavelength used for irradiating the sample when using a xenon arc lamp; in the collection channel of a spectrofluorometer, they are utilized to record the range of wavelengths emitted by a fluorophore (emission

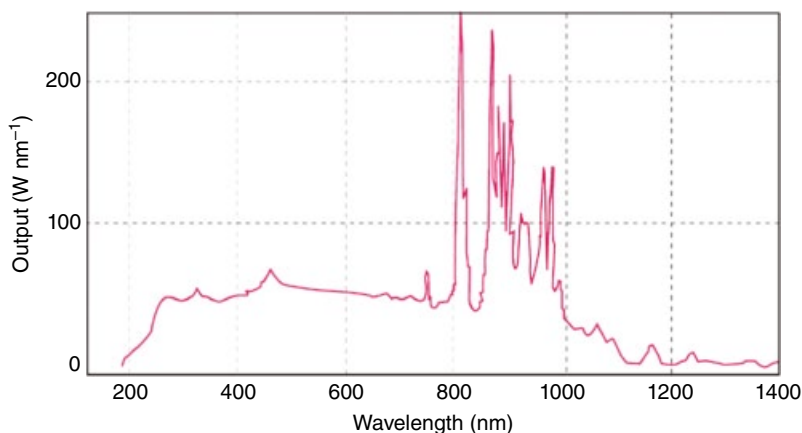


FIGURE 68.2 Spectral distribution for the 300W xenon arc lamp. Source: Reproduced with permission of ISS.

spectrum, see Section 68.5.2). The simplest monochromator includes a diffraction grating and slits at the entrance and at the output. Light impinging at an angle on the grating is diffracted at a series of angles; usually, the first angle (or first order) is selected for the measurement.

It is important to realize that the transmission of the light traversing a monochromator is affected by two parameters:

1. The wavelength—the grating has a specific transmission curve, and some wavelengths are transmitted with a higher efficiency than other wavelengths, a feature to remember when collecting excitation and emission spectra.
2. The polarization status of the radiation—the grating of the monochromator transmits differently radiation with different planes of polarization.

Moreover, it is important to remember that when a monochromator is set to deliver radiation at wavelength λ , it also delivers radiation at 2λ (second order); as an example, if the excitation monochromator is set at 300 nm, it delivers radiation at 600 nm too. Typically the intensity of the second order is about 1/10 the intensity of the first order; still this amount is sufficient to contaminate the emission spectrum. The second order can be eliminated with a judicious selection of filters.

A characterization of every monochromator is the amount of stray light, that is, radiation present at any wavelength other than the specific wavelength the monochromator is set at. The stray light is usually measured as the amount of light that is transmitted outside the band pass of the 632.8 nm HeNe laser line. For typical holographic gratings it is 10^{-5} the intensity of the line. While this amount is not typically important for the study of fluorophores in thin solutions, it becomes important when the sample is in a turbid solution or even a solid state. Different strategies are available for the minimization of the stray light, the first being a judicious selection of the grating. Gratings are classified depending upon their fabrication process: the ruled gratings and holographic gratings, with the latter displaying less stray light inhomogeneity as the grooves are formed through the interference process of two laser beams in a photosensitive material, while in the former the grooves are formed mechanically.

Gratings can be arranged in different designs to build a monochromator, the two more popular being the Czerny–Turner and the Seya–Namioka.

68.3.3 Light Detectors

In all the instruments the fluorescence signal is converted into current by a photomultiplier tube (PMT) or photodiode (instruments for lifetime measurements may utilize other types of detectors too, such as hybrid PMTs, microchannel plate detectors, or streak cameras). PMTs are sensitive within a set wavelength range that is determined by the material used in the photocathode. Figure 68.3 displays the region of sensitivity for the PMT Model R928 by Hamamatsu. The PMT can be utilized in the region from

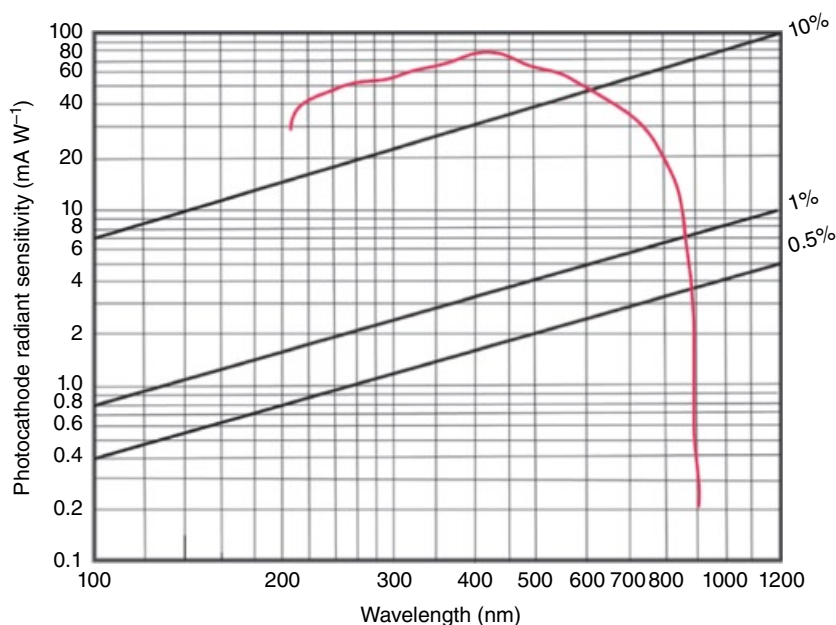


FIGURE 68.3 Wavelength range for a photomultiplier tube model R928 by Hamamatsu.

about 230 nm to about 830 nm. It is apparent that even within the operational wavelength region, the sensitivity is not the same; the nonlinearity of the sensitivity introduces an artifact in the data such that a correction to the data has to be introduced.

A spectrofluorometer includes other optical elements such as lenses and mirrors; moreover polarizers are utilized for anisotropy measurements. The operational region of the instrument is given by the superposition of the wavelength transmission of the various elements of the instruments. Even within this region, the variation in transmission has to be taken into account when measuring the fluorescence parameters. The procedures will be outlined in the measurement sections later.

Figure 68.4 displays the technical diagram of the K2 multifrequency phase fluorometer (MPF) made by ISS, an instrument capable of measuring all of the relevant fluorescence parameters.

The standard light source is a 300 W xenon arc lamp. CW lasers, pulsed lasers (including the multiphoton laser), and LEDs can be coupled to the K2 as well; typically these sources are utilized for the measurement of the decay times of fluorescence.

The light emitted by the source travels through the excitation channel that comprises the monochromator, a filter holder, and the polarizer holder; the monochromator selects the wavelength of the light that excites the sample. The fluorescence emitted by the sample is collected through the left or the right channels; the right channel includes the emission monochromator.

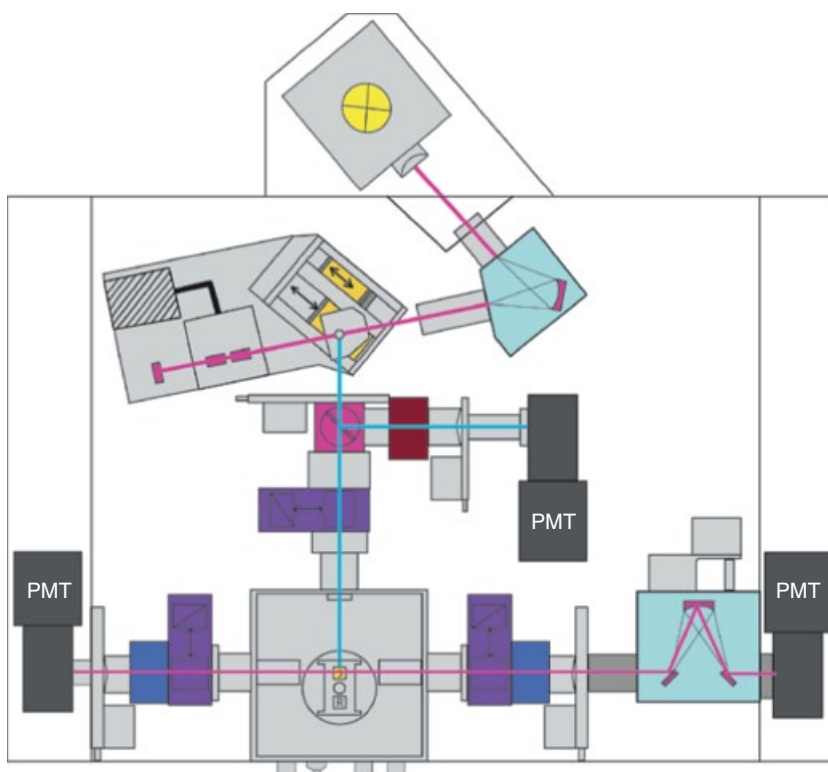


FIGURE 68.4 K2 multifrequency phase and modulation spectrofluorometer. Source: Reproduced with permission of ISS.

The instrument includes polarizer holders, filter holders, shutters for blocking the light from reaching the sample, and the detectors. All of these components are required for automated measurement acquisition.

68.3.4 Instrumentation for Steady-State Fluorescence: Analog and Photon Counting

Two general schemes are utilized to process the signal collected by the PMT: in one scheme, named *analog detection*, the signal from the PMT goes through a current-to-voltage converter, an amplifier, and finally, it is digitized by an analog-to-digital converter (ADC). The signal is then displayed on, and/or stored in, the computer.

In another scheme, named *photon counting detection*, the signal from the PMT goes through an amplifier discriminator that allows for the selection of pulses over a set threshold. A counter in the processing unit counts the number of photons collected per seconds by the detector. This parameter is then displayed by the software on, or stored in, the computer.

Although the advantage of analog detection is in the capability of processing signals within a high dynamic range and fast response, its overall sensitivity is lower than the sensitivity of photon counting detection. Ultimately the choice of one scheme over the other depends upon the specific application.

68.3.5 The Measurement of Decay Times: Frequency-Domain and Time-Domain Techniques

The instrumentation for the measurement of fluorescence decay times is broadly classified as belonging to one of two groups, time-domain and frequency-domain techniques.

The time-domain technique includes the single photon counting, the multiscaler, and the time-correlated single photon counting (TCSPC); the TCSPC is usually the technique utilized more often. The frequency-domain technique comes in an *analog* version (AFD) and a *digital* version (DFD), which has just been introduced.

In TCSPC, a photon is counted within a set time period with a high precision (Fig. 68.5). The time period is defined by the intervals between the pulses of the

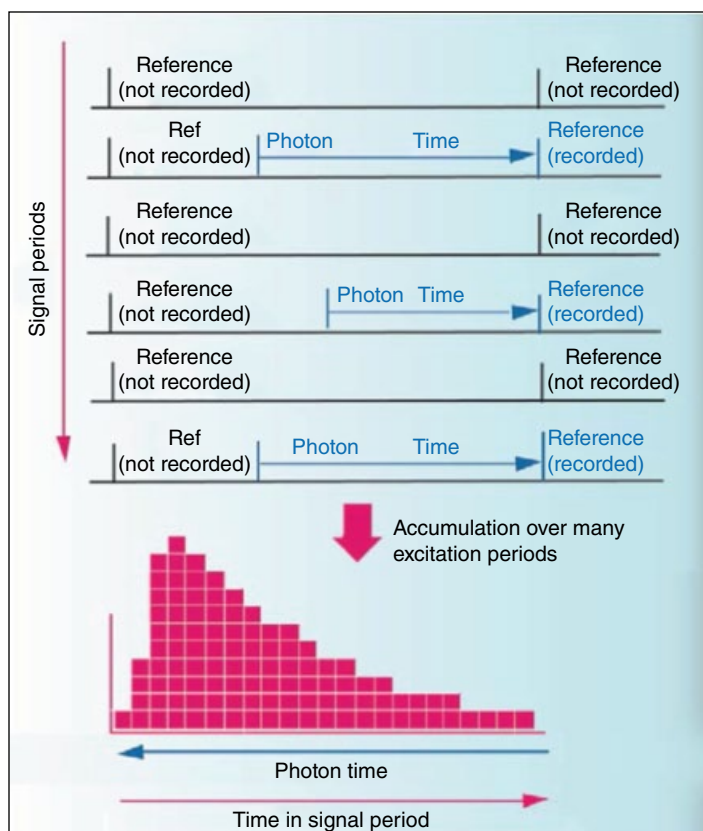


FIGURE 68.5 Principle of start-stop mechanism utilized in TCSPC data acquisition.

excitation light (repetition rate of the light source), and the precision is given by the acquisition electronics (mainly the time-to-amplitude converter (TAC) and the ADC components). For instance, when using an excitation light, emitting pulses at 80 MHz, the time period is the distance between two such pulses (12.5 ns). Typically, the repetition rate of some light sources can be set by the user.

At the arrival of each pulse on the light detector, a high precision timer is triggered, which records how much time has passed between the arrival of the excitation pulse and the emitted photon.

The TAC unit produces a signal proportional to the arrival time of the photon; different arrival time records are grouped in different memory locations (bins) of computer memory.

To interpret the lifetime time information obtained by a TCSPC instrument, a histogram of the arrival time records is built. For a single exponential decay, a curve similar to the one of Equation 68.4 is collected, and the decay time τ is determined using a minimization technique to fit the experimental data to the theoretical decay model.

The frequency-domain technique is more versatile as it can perform either with pulsed sources used for TCSPC or with the modulation of the excitation light source: the modulated excitation results in a modulated fluorescence with a phase and modulation, which is dependent on the lifetime of the excited fluorophores.

The instruments utilized in frequency-domain technique are called MPF or, simply, frequency-domain fluorimeters. The underlying operational principle of an MPF is illustrated by Figure 68.6 for a CW source. The excitation light $E(t)$ is modulated at a frequency ω ; its modulation is characterized by an alternating component AC_{EX} and an

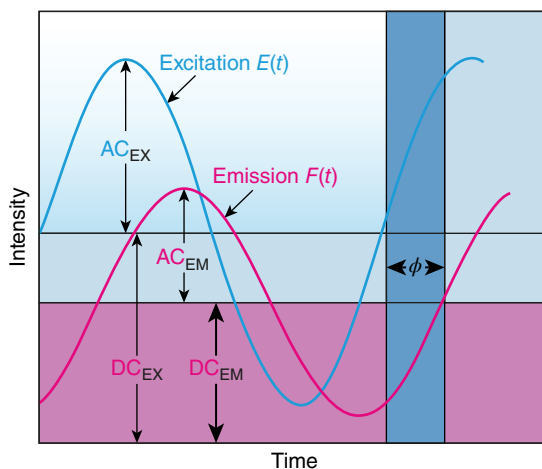


FIGURE 68.6 Schematics of the excitation and emission light in frequency-domain spectroscopy; the emission light is phase shifted and demodulated with respect to the excitation light.

average component DC_{EX} . The fluorescence light is modulated at the same frequency ω , but its phase is delayed by the quantity ϕ , and the overall modulation $(AC/DC)_{\text{EM}}$ is less than the original modulation of the excitation light. A frequency-domain instrument measures the phase shift ϕ and the demodulation m of the fluorescence; both quantities are related to the decay time (see Equations 68.8 and 68.9). For a single exponential decay, the decay time is related to the phase angle and to the modulation by the following relations:

$$\tau_p = \frac{1}{\omega} \tan \phi \quad (68.8)$$

$$\tau_M = \frac{1}{\omega} \sqrt{\frac{1}{m^2} - 1} \quad (68.9)$$

Such measurements are repeated at several different values of the modulation frequency ω ranging typically from two or three for a single exponential decay to up to 20–25 for multiple exponential decays. The decay times τ_i are determined using a minimization technique to fit the experimental data.

The first modern frequency-domain instrument has been introduced by Spencer and Weber in 1969 [1]. In this instrument the light source is modulated at a frequency ω , and the light detector is modulated at a frequency $(\omega + \Delta\omega)$; the two frequencies are being provided by phase-locked frequency synthesizers. The approach is also known as “heterodyning.” The output signal includes components at the sum (2ω) and the difference ($\Delta\omega$) frequency; the low signal component $\Delta\omega$, called the “cross-correlation frequency,” which is typically in the range from 1 Hz to 20 kHz, is utilized to determine the phase shift and the demodulation of the fluorescence. From the phase and modulation of the $\Delta\omega$ frequency, the phase and the modulation of the fluorescence can be determined relative to that of a reference lifetime.

68.4 FLUOROPHORES

Generally fluorophores are divided into *intrinsic* and *extrinsic*. Intrinsic fluorophores are the natural components of a system (typically biological macromolecule) that exhibit fluorescence that can be measured, for instance, the aromatic amino acids tyrosine, tryptophan, and phenylalanine of the proteins, NADH, the flavins, and the porphyrins-based compounds such as chlorophylls. Extrinsic probes include all those molecules that are foreign to the system or were added to it artificially (fluorescent probes and labels—organic dyes, quantum dots, or biological fluorophores), such as fluorescein and 1,8-anilinonaphthalene sulfonic acid (ANS), which are introduced by the experimenter. Such molecules can be covalently linked to the molecule under study or noncovalently as in the case for diphenylhexatriene (DPH) used to study membranes.

68.5 MEASUREMENTS

68.5.1 Excitation Spectrum

The excitation spectrum displays the emission intensity distribution at one wavelength while scanning the excitation wavelength over a range of wavelengths (Fig. 68.7). Practically, for the acquisition of the excitation spectrum, the emission monochromator of the spectrofluorometer is set at a fixed wavelength (in the sample emission range), and the excitation monochromator is scanned over a range of wavelengths (the range that corresponds to the sample absorption range). Referring to the Perrin–Jabłoński diagram of Figure 68.1, when acquiring the excitation spectrum, one detects photons emitted by the molecules at a set wavelength (represented by one of the lines in the group at the middle of the figure) while scanning the wavelength of the radiation (energy of photons) sent to the sample from high energy to low energy (the group of lines at the left of the figure).

If there are no changes that occur to the molecule in the excited state, then the excitation spectrum closely resembles the absorption spectrum acquired with a spectrophotometer; yet, in most instances, it does not: in order for the two to match, a suitable correction of the instrumental factors has to be applied. The main culprit of the differences is due to the lamp; it features a peculiar emission spectrum of its own, that is, the intensity of the emitted radiation is not constant at all the wavelengths. In order to

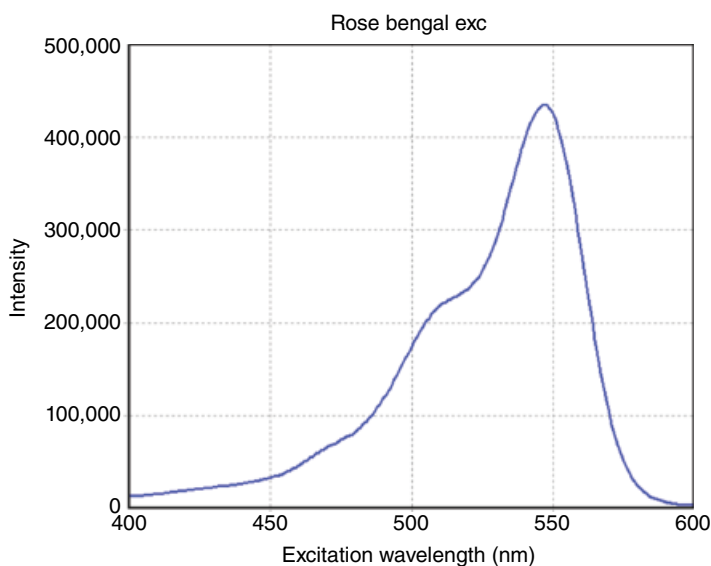


FIGURE 68.7 Excitation spectrum of rose bengal in a water solution, acquired using the K2 spectrofluorometer. The spectrum was acquired by scanning the excitation monochromator from 400 to 600 nm in steps of 1 nm; at each position data were acquired for 1 s. The fluorescence was observed at 610 nm. Source: Reproduced with permission of ISS.

correct for this effect, a small fraction of the excitation light is diverted in the reference channel of the spectrofluorometer (Fig. 68.4) where it passes through the quantum counter and it is collected by the reference detector. The quantum counter, usually a stable fluorophore at a high concentration in solution, delivers a number of photons proportional to the absorbed signal; therefore, at each wavelength, we have a signal proportional to the signal emitted by the lamp; this signal is utilized to correct the fluorescence signal collected in the emission channel. Although this correction addresses most of the concerns, it does not completely correct the excitation spectrum as the beam splitter utilized to divert part of the excitation light into the reference channel reflects differently the two planes of polarization. For a full correction to be implemented, one should place a cuvette with a scattering solution in the sample compartment and acquire an emission spectrum over the wavelength range of interest and then acquire the emission spectrum of the fluorophore and divide it by the spectrum of the scatterer. In this way, the excitation spectrum is fully corrected. Practically, the correction introduced by using the quantum counter and the reference channel is sufficient; one should nonetheless specify the experimental conditions when publishing the spectrum.

68.5.2 Emission Spectrum

The emission spectrum of a fluorophore is most likely the most popular experimental measurement carried out in fluorescence (Fig. 68.8). The spectrum is acquired by setting the excitation wavelength at a fixed value (one of the lines belonging to the group to the left of the Fig. 68.1) and then by scanning the emission monochromator over a range of emission wavelengths (group of lines in the middle of the Fig. 68.1).

There are a few general rules that apply to emission spectra:

1. The emission of fluorescence occurs at wavelengths longer than the excitation wavelength (Stokes shift).
2. The shape of the emission spectrum does not change by changing the excitation wavelength.
3. The emission spectrum is a mirror image of the excitation spectrum of lower energy.

An examination of Figure 68.1 explains as to why the first rule holds. When the molecules are excited, they relax to the lowest vibrational level of the excited states, and from there, they emit fluorescence. Fluorescence photons have a lower energy than excitation photons (i.e., the fluorescence occurs at longer wavelengths than the excitation). Hence, we also gather that the shape of the emission spectrum does not change by changing the excitation wavelength. Finally, rule 3 establishes that the emission spectrum ($S_1 \rightarrow S_0$ transition) is a mirror image of the absorption transition involving the same levels ($S_0 \rightarrow S_1$ transition). If the excitation spectrum includes transitions to higher levels, the emission spectrum will not be a mirror image of the excitation. There are exceptions to the mirror image rule: for instance, when *p*-terphenyl is excited the

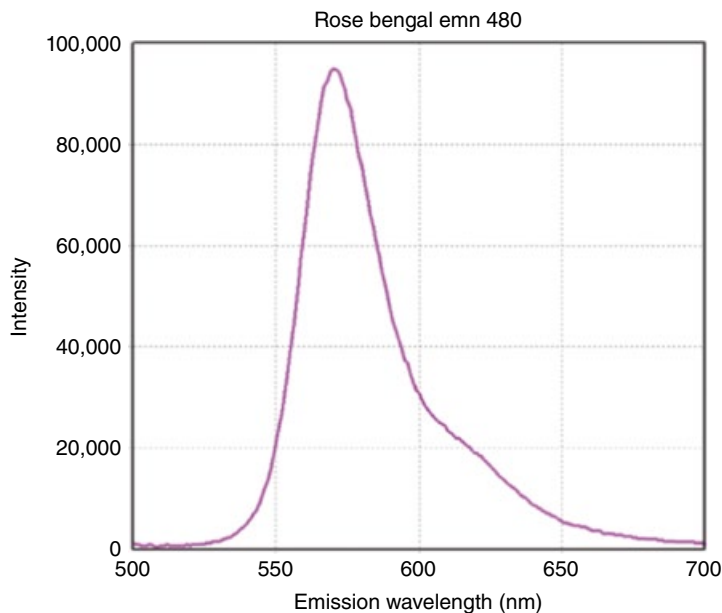


FIGURE 68.8 Emission spectrum of rose bengal in a water solution, acquired using the K2 spectrofluorometer. The excitation monochromator was set at 490 nm. The emission spectrum was acquired by scanning the emission monochromator from 500 to 700 nm in steps of 1 nm; at each position data were acquired for 1 s. Source: Reproduced with permission of ISS.

nuclei undergo a geometric rearrangement upon absorption, and the emission spectrum shows the additional vibrational structure. Excited-state reactions can also result in emission spectra that mark a departure from the mirror rule and so the formation of complexes (for instance, pyrene).

As for the excitation spectrum, the emission spectrum is affected by experimental artifacts, namely, the transmission of the emission monochromator and the sensitivity of the light detector: The transmission of the monochromator varies with the wavelengths, and moreover, it features different transmissions for the two planes of polarization of the light (see later for the definition of light polarization); the sensitivity of the light detector varies with the wavelength. All these variations have to be accounted for in order to acquire a “true” emission spectrum. To this respect, one distinguishes between technical spectrum (the spectrum acquired by an instrument) and the corrected spectrum (the technical spectrum that has been corrected for the experimental artifacts). Manufacturers typically provide correction files for an instrument; these factors are embedded in the software, and corrected spectra can be acquired online; or spectra can be corrected afterward. Practically, one does not need to correct a spectrum unless it is meant for publications; even in that event, it is completely acceptable to specify that the spectrum is a *technical* spectrum rather than a *corrected* one. There are some instances when corrected spectra are required; when calculating the quantum yield of a fluorophore, one has to calculate the area under the spectrum; the spectrum has to be corrected for providing the proper value. Another instance occurs when using the

Förster resonance energy transfer (FRET), a useful tool for estimating the distances between two interacting and close fluorophores.

Besides the instrumental artifacts, the emission spectra are sometimes distorted by experimental artifacts that a practitioner of the field needs to be aware of, namely:

1. Background fluorescence
2. The second order of the monochromator
3. The Raman spectrum of water

Background fluorescence occurs when the fluorophore is diluted in a solution, and the solvent (e.g., buffer) emits some fluorescence of its own at the emission wavelength utilized in the experiment; the resulting emission spectrum is the superposition of the individual spectra of the solvent and the fluorophore. In this case, one can acquire the emission spectrum of the solvent alone and subtract it from the emission spectrum of the solution in order to obtain the emission spectrum of the fluorophore.

We mentioned about the second order in the paragraph covering the monochromators: when a monochromator is set to deliver radiation at wavelength λ , it also delivers radiation at 2λ (second order); although the intensity is about 1/10 of the intensity of the first order, it is sufficient to introduce distortions when measuring turbid solutions and solid samples. The second order can be eliminated with a judicious selection of filters.

Finally, when working with water as a solution, the Raman peaks are present at a wavelength that is 3400 cm^{-1} longer than the excitation wavelength:

$$\lambda_{\text{Ex}}^{-1} - \lambda_{\text{R}}^{-1} = 3400\text{ cm}^{-1} \quad (68.10)$$

As an example, when exciting at 300 nm an emission peak appears at 334 nm; when exciting at 350 nm, an emission peak appears at 397 nm. Note that, while the position of the peak is fixed in unit of wavenumbers ($1/\lambda$), the position varies when dealing in wavelengths (λ); the change in the peak position with the change of the excitation wavelength allows for the user to discern the peak from other peaks or artifacts. The intensity of the Raman peak provides a simple tool to verify the status of the light source of the spectrofluorometer; measured periodically, one can have a pretty good idea of the derating of the xenon arc lamp and make a decision as to when to replace the lamp.

68.5.3 Decay Times of Fluorescence

The fact that the decay times of many fluorophores are in the range of 1–30 ns is truly amazing as this time scale is typical of molecular interactions in biological systems (enzyme conformational shifts, rotational motions in proteins, photosynthetic reactions, etc.) in physiologically active systems.

The decay time is affected by many parameters of the microenvironment (temperature, ions, polarity, viscosity, electric fields), and this is the reason it is widely utilized

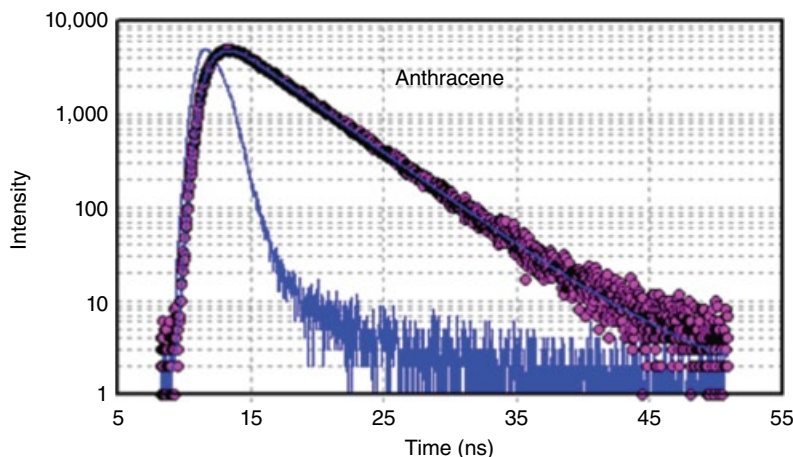


FIGURE 68.9 Decay curve of anthracene in ETOH using a TCSPC instrument (ChronosBH, by ISS). Source: Reproduced with permission of ISS.

for studying molecular interactions. For instance, the decay time of ANS in water is about 100 ps; when ANS is bound to a protein, the lifetime is 8–10 ns. The lifetime of ethidium bromide is 1.8 ns in water; it is 22 ns when bound to DNA and 37 ns when bound to tRNA.

Finally, the lifetimes can be used as an analytical tool as well for the characterization of the presence of specific dyes or simply for the quantitation of complex fluorescent mixtures (the type of crude oil provided by a well, the dye in a hair spray or a soap, the production process of paper, the counterfeiting of banknotes and of drugs, etc.).

Back in 1962, Strickler and Berg [2] published a relation to estimate *a priori* the excited-state lifetime of a fluorescent molecule. Yet, its usefulness is limited because of the variation of lifetimes due to the experimental conditions. That is, the best way to know the lifetime of a fluorophore is to measure it directly.

Figure 68.9 displays the decay time of anthracene in ETOH using the ChronosBH, a TCSPC instrument, by ISS. The light source is a pulsed LED emitting at 335 nm. A high-pass filter (WG 385, 50% transmission at 385 nm) was used to separate the fluorescence. A single lifetime of 4.2 ns was determined using the fitting routine of the software.

Figure 68.10 displays the decay time of anthracene in ETOH using the ChronosFD, a frequency-domain instrument. Phase and modulation data were acquired at 14 different modulation frequencies ranging from 2 MHz to about 250 MHz. The light source is a pulsed LED emitting at 370 nm. A high-pass filter (WG 389, 50% transmission at 385 nm) was used to separate the fluorescence. A single lifetime of 4.2 ns was determined using the fitting routine of the software. In both techniques the decay times are recovered by using a fitting algorithm (least square analysis), the algorithm of the theoretical functions that best minimize the differences with the experimental points. Other approaches are available for the data analysis, such as the maximum entropy method (MEM) and the phasor analysis.

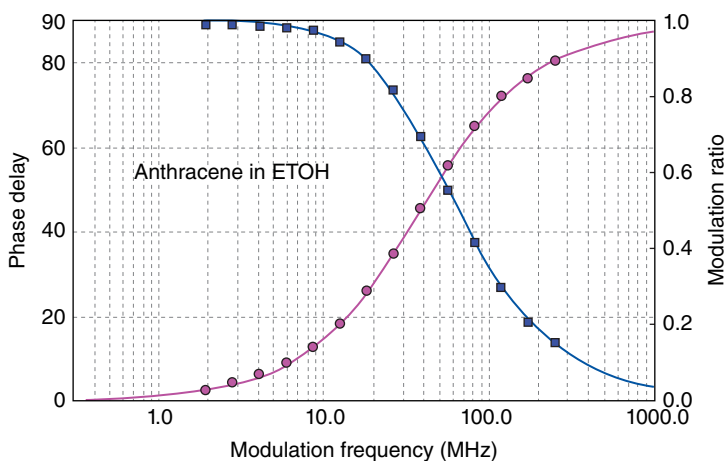


FIGURE 68.10 Decay curve of anthracene in ETOH using a frequency-domain instrument (ChronosFD, by ISS). Source: Reproduced with permission of ISS.

TABLE 68.1 Quantum Yield Values of Selected Molecules

Molecule	Wavelength Range (nm)	Temperature (°C)	Solvent	Quantum Yield
Benzene	270–300	20	Ethanol	0.04
Anthracene	360–480	20	Ethanol	0.27
Tryptophan	300–380	25	H ₂ O	0.14
Rhodamine 101	600–650	20	Ethanol	1.0

68.5.4 Quantum Yield

The quantum yield is a parameter that varies widely from molecule to molecule. A few examples are reported in Table 68.1. Clearly, when looking for a fluorescent probes, there are advantages in selecting one featuring a high quantum yield!

We refer the reader to the literature listed in Further Reading for the measurement of the quantum yield. We only recollect that there is a direct mode and a relative mode. The direct mode encompasses the use of the integrating sphere, an accessory of the spectrofluorometer that allows for the determination of the number of photons emitted by a sample. The relative mode allows for the determination of the quantum yield of a sample by comparison to a reference of known quantum yield. Both measurements require particular attention to the details.

68.5.5 Anisotropy and Polarization

Anisotropy (or polarization) is a popular application of fluorescence spectroscopy as it allows for the measurement of the rotation of molecules as well as of their shape and size and the rigidity of molecular structures.

A light beam is described as an electromagnetic wave with an electric vector \vec{E} and a magnetic vector \vec{B} perpendicular between them; both are also perpendicular to the

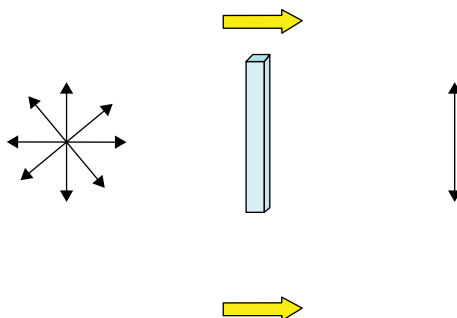


FIGURE 68.11 An unpolarized light beam traverses a polarizer; a plane of polarization is selected.

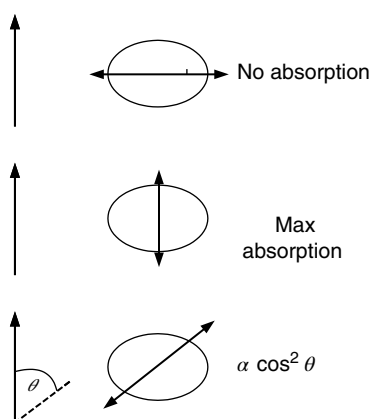


FIGURE 68.12 Molecules with the electric dipole featuring a component parallel to the direction of the electric field of the excitation light have a probability for absorption of a photon.

direction of propagation of the light beam \vec{k} . Natural light can be described as the superposition of innumerable single wave representations. When working with natural light a particular direction of the electric vector \vec{E} can be selected by using a polarizer; such wave is said to be “polarized” (Fig. 68.11).

Polarized light can be utilized for interesting experiments and applications. When polarized light with the proper energy illuminates an ensemble of molecules (Fig. 68.12), only molecules with the excited-state dipole moment \vec{M}_A (or transition moment) oriented in the same direction of the electrical field (polarization) can absorb the photons. If the direction of polarization of the excited beam and the direction of the dipole moment of the molecule are perpendicular to each other, no absorption takes place. In intermediate cases, the probability of the absorption is proportional to $\cos^2\theta$, where θ is the angle between the vector \vec{E} of the exciting light and the vector \vec{M} of the transition moment dipole (Fig. 68.12).

Because of the preferential absorption rules of the molecules, a polarized light introduces a *photoselection* of the molecules. As the distribution of the excited fluorophores is anisotropic, the fluorescence is anisotropic too. Any change in the direction of the

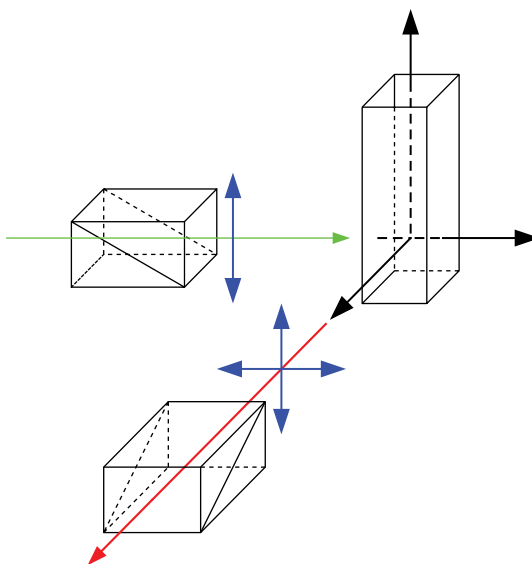


FIGURE 68.13 Experimental setup for anisotropy measurements. The spectrofluorometer has a polarizer in the excitation channel, and a second polarizer in the emission channel. The intensity of the fluorescence reaching the light detector is measured for the different orientation of the polarizers (see Equation 68.11).

transition moment \vec{M}_A during the time the molecule spends in the excited level will result in a decrease of the anisotropy, that is, the overall polarization of the fluorophore solution will decrease. The decrease in the anisotropy can be due to several reasons:

- Difference in direction between the absorption and emission transition moments. This happens as the transition moments of the excited states S_1 and S_2 may not be the same; yet, molecules emit from the lowest vibrational level of S_1 .
- Brownian motion. Molecules in the excited state enter into collisions with the molecules of the solvent or with the molecules of the same species, and as a result, the direction of the emission transition moment changes.
- Energy transfer to another molecule featuring a different orientation.

Anisotropy is measured using a spectrofluorometer equipped with polarizers; one polarizer is mounted in the excitation beam (Fig. 68.13), and a second polarizer is inserted in the emission channel. The anisotropy is defined as

$$r = \frac{I_{VV} - gI_{VH}}{I_{VV} + 2gI_{VH}} \quad (68.11)$$

And the polarization is

$$P = \frac{I_{VV} - gI_{VH}}{I_{VV} + gI_{VH}} \quad (68.12)$$

The two parameters, anisotropy and polarization, describe the same phenomenon; they are related to each other by

$$P = \frac{3r}{r+2} \quad (68.13)$$

(In the following description we will refer to anisotropy only.) In the relations earlier, I_{VV} is the measured fluorescence intensity with the polarizer in the excitation channel in the (V)ertical position and the polarizer in the emission channel in the (V)ertical position; I_{VH} is the measured fluorescence intensity with the polarizer in the excitation channel in the (V)ertical position and the polarizer in the emission channel in the (H)orizontal position.

The number g , called the *g-factor*, is given by $g = I_{HV}/I_{HH}$, where the letters V and H refer to the positions of the polarizers in the excitation and emission channel, respectively. The *g-factor* corrects the anisotropy values for the artifact introduced by the instrument; as is the case for emission spectra, the instrument has different transmission properties for the two planes of polarization.

Figure 68.14 displays the excitation polarization spectrum for erythrosine (top line until 500 nm); the other line represents the excitation spectrum in the range from 300 to 530 nm. The fluorescence is collected at 550 nm. The polarization is negative for wavelengths below 360 nm and then rises sharply up to 400 nm and stays almost constant above 400 nm. The reason for this behavior is due to the fact that the excitation at the short wavelengths favors the transition $S_0 \rightarrow S_2$, while at the longer wavelengths the transition $S_0 \rightarrow S_1$ is the one excited: as the fluorescence is always emitted by the lowest vibrational level of S_1 , it is an indication of the different orientation of the transition moments of the excited levels S_1 and S_2 . Practically, when using anisotropy measurements one has to select and specify the excitation wavelength (and choose a wavelength displaying a high value of polarization).

What are the values that the anisotropy can assume? In order to answer this question, one has to introduce the emission transition moment \vec{M}_E and distinguish the two cases:

1. \vec{M}_E and \vec{M}_A are parallel.
2. \vec{M}_E and \vec{M}_A are not parallel.

Without going into the details of the calculations (the interested reader can consult the book by Valeur cited in References), we note that for the case of the two moments being parallel and in absence of any motion, it is $r_0 = 0.4$; this value is called the *fundamental anisotropy*. When the two moments are not parallel, the values are confined in the range

$$-0.2 \leq r_0 \leq 0.4 \quad (68.14)$$

The case of the decrease of anisotropy due to Brownian motion collisions is a very interesting one for its practical applications. This is the case when molecules in the

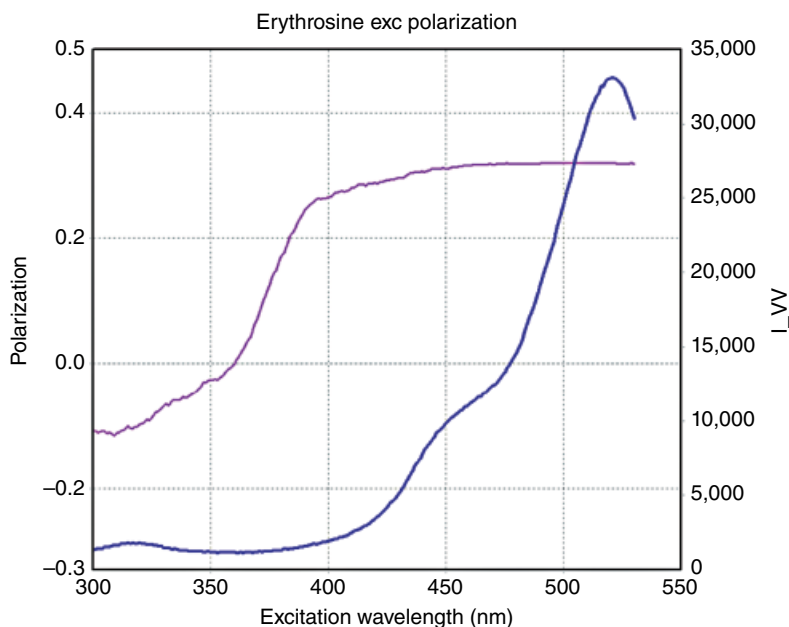


FIGURE 68.14 Excitation polarization spectrum for erythrosine (top line until 500nm); the other line represents the excitation spectrum in the range from 300 to 530 nm. The fluorescence is collected at 550 nm.

excited state rotate due to collisions with the solvent. The amount of the depolarization depends upon the value of the decay time of the molecule, the size of the molecule, and the viscosity and temperature of the solvent. In fact, let us suppose that the decay time is of the same order of the rotational time; it is found that the anisotropy decays, for a spherical molecule, according to the following relation:

$$r(t) = r_0 \exp(-6D_r t) \quad (68.15)$$

where D_r is the rotational diffusion coefficient. From the Stokes–Einstein relation $D_r = RT/6V\eta$, V is the hydrodynamic volume of the molecule, η is the solvent viscosity, R is the gas constant, and T is the absolute temperature. D_r can be determined by resolving Equation 68.15 using time-resolved fluorescence techniques. Alternatively, if the decay is a single exponential decay, it can be solved using steady-state technique. As

$$\bar{r} = \frac{1}{\tau} \int_0^{\infty} r(t) \exp\left(-\frac{t}{\tau}\right) dt \quad (68.16)$$

By direct substitution one finds

$$\frac{1}{\bar{r}} = \frac{1}{r_0} (1 + 6D_r \tau) \quad (68.17)$$

TABLE 68.2 Selected Applications of Anisotropy Measurements

Spectroscopy	Separation of Excited States
Polymers	Local viscosity Molecular orientation Chain dynamics
Immunology	Antigen–antibody reactions Immunoassays
Molecular biology	Proteins interactions Nucleic acid–protein interactions Biological membranes Micellar systems

This is the Perrin equation; it allows for the evaluation of the decay times by measurements of the steady-state polarization! In some literature, the quantity $\tau_c = 1/6D_r$, called the rotational correlation time, is used. This case is strictly valid for a spherical molecule. When the more complex shape of a general ellipsoid is considered, the motion is described by three rotational diffusion coefficients associated with each of the rotational axis. The relation between the rotational correlation times and the rotational diffusion coefficients is no longer simple. The anisotropy decay is described by

$$r(t) = \beta_1 e^{-t(4D_1+2D_2)} + \beta_2 e^{-t(D_1+5D_2)} + \beta_3 e^{-t(6D_2)} \quad (68.18)$$

where

$$\begin{aligned} \tau_1 &= \frac{1}{(4D_1 + 2D_2)} \\ \tau_2 &= \frac{1}{(D_1 + 5D_2)} \\ \tau_3 &= \frac{1}{(6D_2)} \end{aligned} \quad (68.19)$$

In this expression the quantities $\beta_1, \beta_2, \beta_3$ represent expressions for the angles between the absorption and emission dipoles and the axes of the ellipsoid; D_1 and D_2 are the diffusion coefficients around the axis of symmetry and equatorial axes, respectively.

There are physical conditions where a probe is restricted to motion within an angle, for instance, the case of a probe in a membrane. In these cases, the anisotropy does not decay to zero. A hindered rotator is described by the following expression:

$$r(t) = (r_0 - r_\infty) \exp\left(\frac{-t}{\tau_c}\right) + r_\infty \quad (68.20)$$

Table 68.2 lists a few applications of the technique that spans from the physical chemistry research all the way to clinical applications.

68.6 CONCLUSIONS

Fluorescence is a sensitive technique that, although started as an analytical tool, is used more and more for the study of molecular interactions *in vitro* and in cells; in fact, it is nowadays capable of detection of single molecules on a routine basis. The fluorescence decay time of typical fluorophores falls in a window (1–20 ns) suitable for the observation of several molecular processes of biological relevance. The spectral properties of fluorophores are changed by several processes including collisions with other molecules, rotational diffusion, and formation of complexes; moreover, the fluorescence properties are sensitive to changes of the environment such as pH, electrical fields, concentration, temperature, and polarity. These features have expanded the applications of fluorescence to fields as diverse as the development of sensors for monitoring the presence of specific analytes (O₂, ions) *in vitro* and *in situ* to the development of sensors for the measure of physical parameters (materials under high pressure, mechanical properties of materials). A variety of research instruments is available for the measurement of the general and specific parameters of the fluorescence. Dedicated instruments are utilized for the measurements in specific immunoassays (polarimeters), in drug discovery (microwell plates and microarrays), cell sorting (cytofluorometers), and genome sequencing.

REFERENCES

1. R.D. Spencer and G. Weber, 1969. Measurements of subnanosecond fluorescence lifetimes with crosscorrelation phase fluorometer. *Ann. N. Y. Acad. Sci.* 158, 361–376.
2. J.S. Strickler and R.A. Berg, 1962. Relationship between absorption intensity and fluorescence lifetime of molecules. *J. Chem. Phys.* 37, 814–822.

FURTHER READING

- W. Becker, 2005. *Advanced Time-Correlated Single Photon Counting Techniques*; Springer-Verlag, Berlin/Heidelberg.
- D.M. Jameson, 2014. *Introduction to Fluorescence*; CRC Press/Taylor & Francis Group, Boca Raton.
- J.R. Lakowicz, 2006. *Principles of Fluorescence Spectroscopy*, 3rd Edition; Springer-Verlag, New York.
- B. Valeur, 2005. *Molecular Fluorescence*; Wiley-VCH Verlag GmbH, Weinheim.

X-RAY ABSORPTION SPECTROSCOPY

GRANT BUNKER

Department of Physics, Illinois Institute of Technology, Chicago, IL, USA

69.1 INTRODUCTION

X-ray photons are products of natural (e.g., astrophysical) processes that also find many uses as the basis of powerful research tools in science, medicine, and engineering. In addition to the familiar use of X-rays for industrial and medical radiography, X-ray absorption spectroscopy is of great utility for providing information on molecular and electronic structure in chemistry and in condensed matter physics, materials science, biology, geology, and other fields. This chapter describes what is measured, how the measurement is accomplished—sources, X-ray optics, detectors, measurement modes, and samples—and how to minimize errors. The quantum physics of X-ray absorption processes is beyond the scope of this chapter on measurement, as is the detailed analysis of the spectra. For further information on fundamentals, see, for example, [1–3].

69.2 BASIC PHYSICS OF X-RAYS

In this section we briefly describe the basic physics describing the interaction of X-rays with matter. These processes (scattering and absorption) are central to understanding the mode of operation and characteristics of the X-ray optics that are used to manipulate X-ray beams, as well as the interactions of X-rays with materials, and methods of detection of X-rays.

69.2.1 Units

Although SI units are commonly used in X-ray science, angstrom units ($1 \text{ \AA} = 10^{-10} \text{ m} = 1 \text{ nm}$) also are used, because this unit corresponds well to atomic dimensions, a property that is important for both diffraction and spectroscopic experiments. The wavelength range of X-rays is much shorter than that of visible light, which is approximately 0.4–0.7 micrometer (μm), about 10^4 times longer wavelength than typical X-ray wavelengths.

Electron volt (eV) units ($1 \text{ eV} \approx 1.60 \times 10^{-19} \text{ J}$) often are used to describe energies, because the spacings between energy levels in atoms are on the order of eV or greater—typically 10^4 times greater in the case of X-rays, in which case keV (10^3 eV) units are often used. 1 eV is the magnitude of energy change when one electron charge is moved through an electric potential difference of 1 V.

69.2.2 X-Ray Photons and Their Properties

X-radiation is the portion of the electromagnetic radiation spectrum with a wavelength that is shorter than ultraviolet and “vacuum ultraviolet” radiation. X-ray wavelengths are on the order of 1 nanometer (nm) or shorter, typically about 1 \AA , or 0.1 nm.

In classical physics an electromagnetic wave in free space consists of oscillating electric (\vec{E}) and magnetic (\vec{B}) fields that are mutually perpendicular and also perpendicular to the direction of motion. The fields oscillate in phase at a frequency f (with angular frequency $\omega = 2\pi f$) and wavelength λ , where the speed of light is $c = f\lambda$, and $|\vec{E}| = |\vec{B}|c$. The wave vector \vec{k} (where $|\vec{k}| = 2\pi/\lambda$) points along the direction of propagation ($\vec{E} \times \vec{B}$). The X-ray polarization vector \hat{e} is a unit vector in the direction of \vec{E} .

When considering X-ray absorption by atoms, it must be recognized that light, including X-rays, comes in photons. It is generally necessary to use a quantum mechanical and special relativistic description to understand the interaction of light with matter, but when a large number of photons are present in a beam, it may be sufficient for some purposes to treat the beam as a classical electromagnetic field. This approximation is commonly made in X-ray diffraction experiments.

The quantum of electromagnetic field is a photon, a massless particle that carries energy $E = hc/\lambda$, where h is Planck’s constant and c is the speed of light; photon momentum $p = E/c = h/\lambda$; and the photon’s angular momentum along the direction of motion has the quantized values $\pm\hbar$, with $\hbar = h/2\pi$. The product $hc \approx 12398.5 \text{ eV \AA}$, so that a 12.4 keV photon has a wavelength of $1 \text{ \AA} = 0.1 \text{ nm}$ and a 6.2 keV photon has a wavelength of 2 \AA . In this chapter we shall largely restrict our attention to energies from around 1 to 100 keV. The rest energy of an electron is approximately $mc^2 = 511 \text{ keV}$.

Physically X-rays are fundamentally no different than gamma rays, which also are high-energy photons; gamma rays generally have their origins in nuclear processes, whereas X-ray photons generally are produced by transitions of electrons between atomic energy levels and by accelerating charges. Both are produced in various energetic astrophysical processes.

69.2.3 X-Ray Scattering and Diffraction

The simplest interaction between a photon and an atom is the scattering of a photon from electrons in an atom. Although they are charged particles, protons do not scatter significantly compared to electrons because their mass is much greater. For this reason X-ray scattering from atoms depends only on the electron distribution and not (directly) on the nuclear charge.

If a photon scatters from a single weakly bound (quasi-free) electron, the momentum change of the photon causes the electron to recoil. The kinetic energy of the electron carries off some energy so that the scattered (outgoing) photon has a lower energy than the incoming photon. The change in wavelength between the scattered and incoming photons is given by

$$\lambda' - \lambda = \frac{h}{mc}(1 - \cos \theta),$$

where θ is the scattering angle, the angle between the directions of incoming and scattered photon directions. This is called “Compton scattering” and it is a type of inelastic scattering: the scattered photon has a lower energy than the incoming photon. The constant $h/mc = hc/mc^2 \approx 12.4 \text{ keV } \text{\AA} / 511 \text{ keV} \approx 0.0243 \text{ \AA}$ is called the Compton wavelength.

The incident photon can also transfer part of its momentum to the whole atom, in which case its recoil is minimal, and the energy change between incoming and outgoing wavelengths in that case is negligible. This is elastic scattering: the incident and scattered photons have the same energy. Both elastic and inelastic scattering depend on the number of electrons in an atom and their spatial distribution within it. Multiple (say, N) electrons may coherently participate in the scattering process, generating a scattered wave of amplitude N times as strong as that from a single electron. The intensity of the wave varies as the square of the amplitude, giving N^2 times the beam intensity as that from a single electron. The number of electrons that coherently scatter depends on the wavelength and spatial distribution of electrons in the sample.

If a photon (or beam of photons) scatters from a material with spatial periodicities, such as a crystal, the coherent scattering will be greatly enhanced at certain angles of incidence: this is called X-ray diffraction. It is an X-ray analog of using diffraction grating with visible light. The atoms within a crystal are arranged in planes with various separations d_{hkl} , where the integer (“Miller”) indices h, k, l serve to identify which sets of planes one is referring to. The geometric condition for “Bragg diffraction” is $n\lambda = 2d_{hkl} \sin \theta_B$, where λ is the X-ray wavelength and θ_B is the angle (“Bragg angle”) between the incoming wave direction and the diffracting planes and n is a positive integer. The geometry of Bragg diffraction is shown in Figure 69.1.

Bragg diffraction from crystals is an important means of studying structures of materials such as semiconductors, alloys, minerals, and biological molecules such as proteins, lipids, and nucleic acids; it also provides a means to select specific wavelengths (energies) for X-ray absorption experiments using X-ray monochromators. X-ray

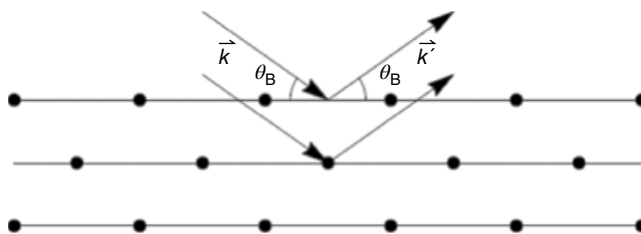


FIGURE 69.1 Bragg diffraction from atomic planes of crystal.

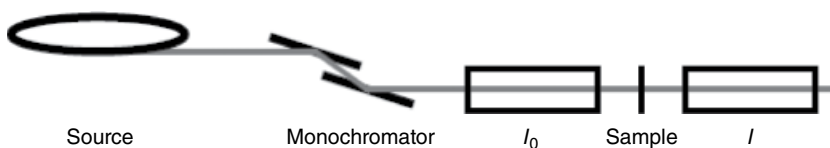


FIGURE 69.2 Schematic transmission mode X-ray absorption measuring apparatus (not to scale).

absorption fine structure (XAFS) is a distinct and complementary technique because it uses photoelectron (not X-ray) interference to probe the sample, and it does not require translational order (e.g., crystalline structure) to determine molecular structure.

69.2.4 X-Ray Absorption

Light can be absorbed by a material when the energy of the photons $E = h\nu = \hbar\omega$ is sufficient to create excitations in it. Microwaves excite transitions between rotational states of molecules; infrared light excites transitions between vibrational states of molecules. Visible, ultraviolet, and X-ray photons excite transitions between electronic states of molecules and atoms. The energies at which specific transitions of materials can be excited are characteristic of the structure and bonding of the atoms, molecules, or crystalline domains of which it's composed.

An X-ray photon may be absorbed when it has an energy that corresponds to a difference between two electronic quantum states in an atom or molecule. For X-rays the transitions are between a deeply bound core electronic state in an atom (such as the lowest energy level, the $1S$) and an empty final state of suitable symmetry (e.g., a P state). The energies of these transitions are characteristic of the atoms in the material.

A typical setup for measuring X-ray absorption spectra is shown in Figure 69.2. The X-ray source (e.g., synchrotron radiation source) generally produces a range of energies from which a specific narrow band of energy is selected by a monochromator, typically one employing Bragg diffraction from set of crystals. Following the monochromator the beam flux (photons/s) is measured with a partially transparent detector I_0 (labeled by the current output signal it produces), and the beam is then transmitted through the sample, and the transmitted flux I is then measured. Often focusing optics, shutters, and beam defining slits are placed between the monochromator and the I_0 detector.

X-rays passing through a homogeneous uniform sample are attenuated exponentially, so that $I/I_0 = G(E)\exp(-\mu(E)x)$, where $\mu(E)$ is the X-ray linear attenuation coefficient of the material, E is the X-ray beam (photon) energy, x is the sample thickness (path length through the sample), and $G(E)$ accounts for detector sensitivities and absorption by air paths and detector windows. This equation is equivalent to $\mu(E)x = \ln(I_0/I) + \ln G(E)$; the second term ($\ln G(E)$) ordinarily gives a slowly varying additive background that can be readily subtracted out numerically or better measured independently by removing the sample from the beam path.

The quantity $\mu(E)$ measures the attenuation of X-rays as they pass through the sample because of absorption and scattering. In most cases of practical interest absorption is the dominant process, but scattering is not generally negligible.

69.2.5 Cross Sections and Absorption Edges

The ability of an atom to absorb or scatter is characterized by atomic *cross sections*, which are energy-dependent quantities that have dimensions of area. A beam of photons of intensity N [photons/s/m²] hitting a thin target of total absorption cross section σ_{tot} [m²] absorbs photons at a rate $N \times \sigma_{\text{tot}}$ [photons/s] where σ_{tot} is the sum of the absorption cross sections of the atoms that are illuminated by the beam.

The *barn* (10^{-28} m²) is a non-SI unit to describe cross sections, commonly used in nuclear physics. It is also common to specify cross sections per mass or per atom.

If the cross sections σ for a pure element are specified on a per-mass basis, we can write $\mu(E) = \rho\sigma(E)$ where ρ is the mass density of the material in the sample. For a sample consisting of a mixture or compound of different elements, we have approximately $\mu(E) = \rho \sum_i f_i \sigma_i(E)$ and f_i and σ_i are, respectively, the mass fraction m_i/M_{total} and cross section of the i th atomic species in the sample. Approximate atomic cross sections are tabulated and readily available online. The ones shown here are from Chantler et al. [4].

A log-log plot of the specific cross section for gold is shown in Figure 69.3. It consists essentially of step increases in absorption at certain energies, with straight lines between them. The jumps are called *absorption edges*; they correspond to specific atomic excitations. The highest energy one at about 80 keV comes from exciting electrons out of the $n=1$ atomic shell (K-edge); the next lower ones in energy between about 12 and 14 keV are due to excitations from the $n=2$ atomic levels (L-edges LI, LII, LIII); the next lower ones in energy are the *M-edges*. Between the edges the curves are approximately straight lines on a log-log plot, which implies a power law dependence. A good rule of thumb is that, between the absorption edges, absorption varies as $1/E^3$, so the cross section (and absorption coefficient μ) decreases by a factor of 8 at twice the energy (if there are no absorption edges in that region). It will be noticed (Figs. 69.3, 69.4, and 69.5) that for heavy elements (gold, copper) the scattering cross sections are much less than absorption cross sections except at high energies; for lighter elements such as carbon, absorption and scattering are equal at about 20 keV.

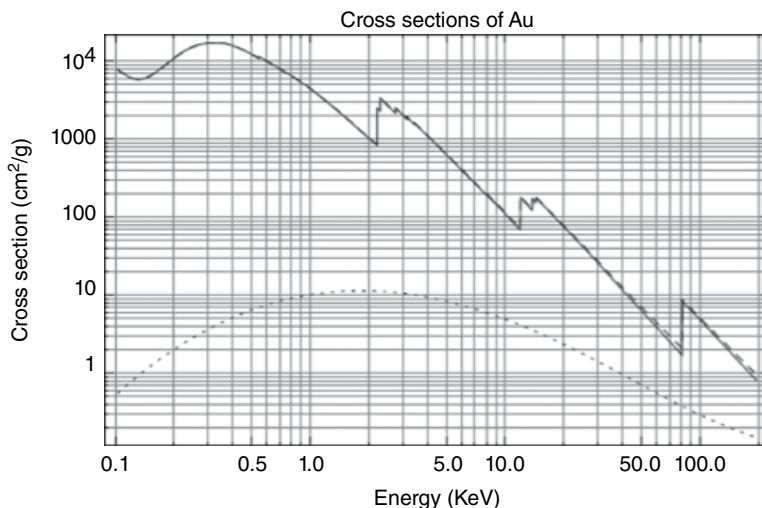


FIGURE 69.3 Log-log plot of the (semiempirical) X-ray absorption cross section of gold ($Z=79$) versus X-ray energy. The K , L_1 , L_2 , L_3 , and M -edges are shown; fine structure is *not* shown. The solid line is the photoelectric absorption, and the dotted line is total elastic+inelastic scattering cross section; the dashed line is the sum of absorption and scattering.

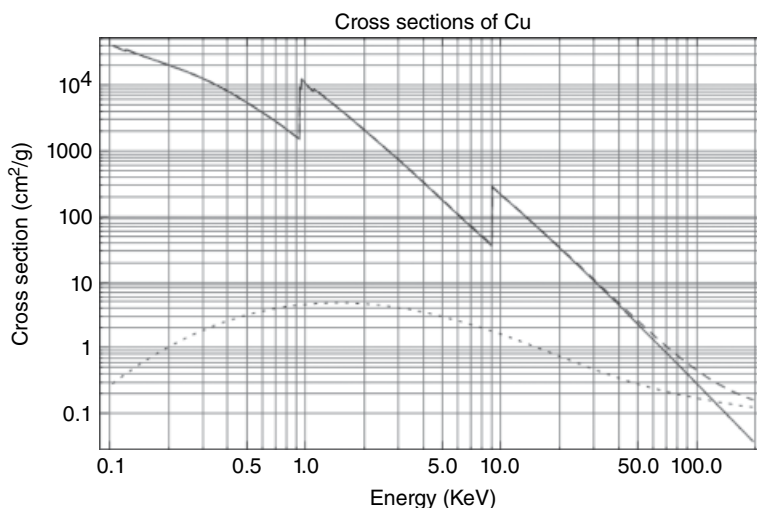


FIGURE 69.4 Log-log plot of the (semiempirical) X-ray absorption cross section of copper ($Z=29$) versus X-ray energy. The K , L_1 , L_2 , L_3 , and M -edges are visible; fine structure is *not* shown. The solid line is the photoelectric absorption, and the dotted line is total elastic+inelastic scattering cross section; the dashed line is the sum of absorption and scattering.

The energies at which the absorption edges occur are characteristic of the atomic number of the element. For example, the calcium ($Z=20$) K-absorption edge is at 4.04 keV, iron ($Z=26$) is at 7.11 keV, zinc ($Z=30$) is at 9.66 keV, and molybdenum ($Z=42$) is at 20.0 keV. Empirically the K-edge energy depends on atomic number Z as

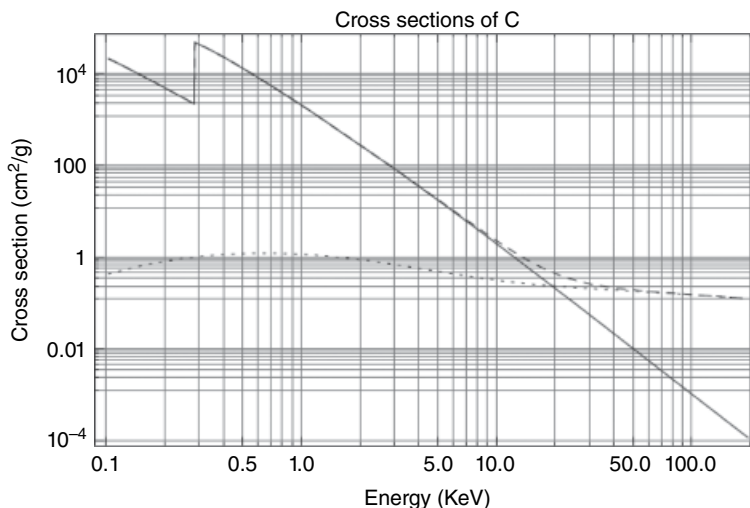


FIGURE 69.5 Log-log plot of the (semiempirical) X-ray absorption cross section of carbon ($Z=6$) versus X-ray energy. The K -edge is visible; fine structure is *not* shown. The solid line is the photoelectric absorption, and the dotted line is total elastic + inelastic scattering cross section; the dashed line is the sum of absorption and scattering.

$E_K \propto Z^{2.16}$. Similarly the L-edge energies are fairly smooth functions of the atomic number. Historically the simple dependence of core-level energies with atomic number was important in completing the periodic table. This predictability and the substantial separation between edges allow experimenters to easily “tune into” particular atomic species for study. Good sources of data on absorption edges, fluorescence energies, and absorption cross sections are the X-ray data booklet [5], the servers at NIST [4], and the online server at IIT [6].

The X-ray absorption cross section curves shown in Figures 69.3, 69.4, and 69.5 are composite averaged representations of various experiments and theoretical computations. They do not represent the XAFS, which consists of peaks and periodic oscillations that are observed in experimental absorption edge spectra. Once measured, experimental XAFS spectra can be analyzed to provide information on average distances to atoms around the element of interest (whose absorption edge it is). This is beyond the scope of this chapter but it is described in [3].

69.3 EXPERIMENTAL REQUIREMENTS

The accurate measurement of XAFS spectra places stringent requirements on the apparatus, because one must not only measure absorption spectra as a function of incident X-ray photon energy to a precision of 10^{-3} to 10^{-4} of the size of the edge step (the difference in absorption above and below the edge), but the incident photon beam

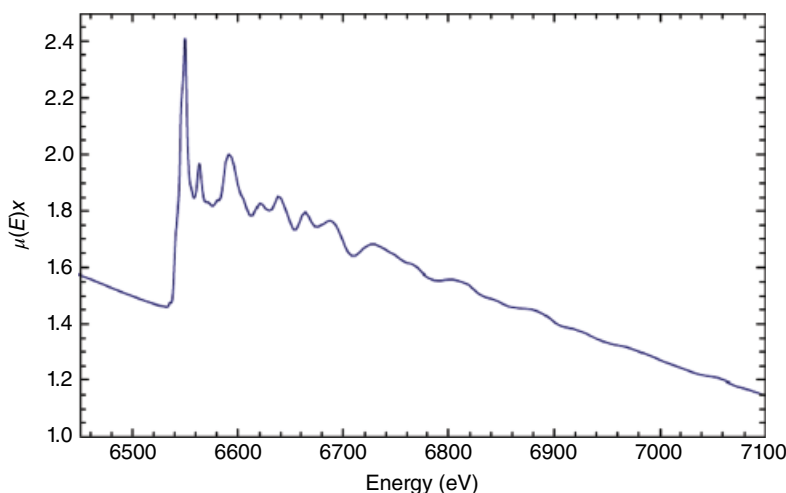


FIGURE 69.6 Plot of experimental transmission mode $\mu(E)x$ data for manganese oxide (MnO) measured at 80 K. The absorption edge occurs at approximately 6540 eV.

energy must be smoothly varied over an extended energy range of about 1 keV above the edge. Preserving positional stability of the beam and shape of the beam is also important. Furthermore certain systematic errors must be eliminated, as described below; otherwise conclusions regarding the structure will be incorrect. For this reason here we will concentrate on XAFS spectra.

XAFS spectra consist of absorption peaks within approximately 20 eV of the absorption edge, some on the order of 1 eV in width, plus much slower oscillations extending about 1 keV above the edge. The width of the near-edge peaks depends on the structure and also the core hole lifetime of the edge in question, because the finite lifetime of the excited state after the X-ray absorption event introduces an uncertainty principle-related energy broadening, which smooths the spectrum.

An example of an XAFS spectrum can be seen in Figure 69.6, which plots $\mu(E)x$ for a thin sample of manganese oxide (MnO). The oscillations become less rapid (in energy space) and smaller in amplitude as energy above the edge increases. Depending on the core hole level width, to adequately measure the structure near the edge, it is necessary to sample at approximately 1 eV or better energy spacing; above the edge the sampling can be several times larger than that. There are some advantages to sampling uniformly in k -space at intervals of 0.05–0.07 \AA^{-1} , where $k^2 = 0.2625 (E - E_0)$, E is the photon energy, and E_0 is the edge energy. Above the edge the key criterion is to sample the spectra at least twice per cycle of the highest-frequency oscillation. Oversampling is helpful to prevent aliasing of noise into the signal passband. For a more detailed discussion and rationale, see the section on “Instrument Control and Scanning Modes” and Bunker [3].

69.4 MEASUREMENT MODES

As described above, the basic mode for measuring XAFS is transmission mode, in which the fraction of X-rays that are absorbed by a sample is measured as a function of photon energy. Other modes are of considerable use however. For dilute samples, in which the element of interest is a small fraction of the total, it is often helpful to instead measure the X-ray fluorescence that is given off as a consequence of the initial absorption event. For example, exciting the K-edge of iron (at about 7112 eV) creates a vacancy (“core hole”) in the $1S$ level, which is unstable; an electron from a higher level, most often $2P$, can fall into the core hole and give up a fluorescence photon in the process, with an energy equal to the difference between the $2P$ and $1S$ level, in this case about 6400 eV. This fluorescence photon can be detected and used as a proxy for the measuring the absorption directly. The advantage of doing this is that one only gets those photons if the absorption occurred in the first place, so it increases the sensitivity of the measurement.

Another method of detection is conversion electron detection or electron yield. In the case described above, fluorescence is not the only way a core hole can deexcite: it can also relax nonradiatively, by ejecting electrons, which propagate through the material and either escape the sample or create secondary excitations. Electrons within $\approx 1000 \text{ \AA}$ of the surface can escape the sample. If the sample is surrounded by a gas, for example, He, it will ionize the gas and create a number of secondary electrons, which can be collected and amplified as in an ionization chamber. This electron yield detection confers surface sensitivity that can be very helpful experimentally. The sample must be slightly conductive and the beam intensities relatively low for this to work reliably.

Another indirect method of detection (which is very infrequently used) is optical detection of XAFS: measuring low levels of light that are produced as a consequence of the absorption event.

It is practical to measure XAFS-like data by scattering electrons or high-energy photons off a sample and analyzing the energy loss spectrum of these scattered particles. These techniques (which are beyond the scope of this chapter) are, respectively, called electron energy loss spectroscopy (EELS) and inelastic X-ray scattering (IXS).

69.5 SOURCES

69.5.1 Laboratory Sources

XAFS makes particularly stringent demands on the measuring apparatus, but it is still possible to use laboratory sources for measuring X-ray absorption spectra; this was done for many years before synchrotron radiation sources became available. Conventional X-ray tubes work by allowing electrons to accelerate under the influence of a high DC voltage and collide with a metallic anode. The decelerating electrons radiate a broad bremsstrahlung (“braking radiation”) X-ray spectrum, and they also

excite the atoms of the anode, which give off fluorescence X-rays that are characteristic of the anode material. For most experiments the narrow bandwidth fluorescence is used, but it is also possible to use the broad spectrum and pass it through a monochromator to select specific energies. This is not particularly effective for measuring XAFS because of the low throughput, but it can be, and has been, done; see, for example, [2].

Another type of lab source is a laser-induced plasma X-ray source, which can be used for stroboscopic experiments. For extremely rapid time-resolved experiments, free electron lasers are coming into use [7].

69.5.2 Synchrotron Radiation Sources

The most common X-ray sources for the measurement of X-ray absorption spectra are synchrotron radiation sources [8]. Free electron laser sources also have recently come online to facilitate experiments that require ultrafast measurements or extreme brightness. These large multiuser facilities are the technological by-product of high-energy physics facilities like Fermilab and the Large Hadron Collider, with an important difference that they accelerate electrons (rather than protons) to speeds extremely close to the speed of light with a booster synchrotron and then store them for extended periods within a ring-shaped pipe containing ultrahigh vacuum in order to minimize scattering of the electrons by residual air molecules. At specific locations around the ring strong magnets (“bend magnets”) are placed so as to bend the trajectory of the electrons, so they can circulate around the storage ring continuously for hours or days. The total current in the ring slowly degrades because of scattering from residual gas molecules in the evacuated ring, and other processes, so the current decreases exponentially, after which it is dumped and electrons are reinjected to restore the current. In some cases it is possible to operate in “top-off mode” in which the current is kept approximately constant by adding small amounts of charge to the bunches while the ring undergoes normal operations. This has significant benefits in keeping heat load on optics and ring components constant.

The electrons are not distributed uniformly around the ring; rather they are concentrated in “bunches,” the maximum number of which depends on the storage ring lattice design. The ring can be populated with a single bunch, or many bunches, depending on experimental needs. The speed of light divided by the circumference of the ring gives the single bunch repetition frequency. For example, the repetition rate for a 600 m circumference ring would be 500 kHz, and if 20 equally spaced bunches were populating the ring, the bunch repetition rate would be 10 MHz. Typical light pulse lengths are ≈ 100 ps long.

At such high speeds (close to the speed of light c) Newtonian mechanics is inadequate to describe the physics, and Einstein’s special theory of relativity is required to understand it. Synchrotron radiation is inherently a relativistic effect. The total energy of E of each electron (rest energy plus kinetic energy) amounts to billions of electron volts. For example, the electron energy is 7–8 GeV for the APS, SPRing8, and ESRF and around 1–3 GeV for most other sources. The rest energy of an electron $mc^2 \approx 0.511$ MeV; the relativistic parameter $\gamma = E/(mc^2)$ is typically of order 10^3 to 10^4 .

A principle of electrodynamics is that accelerating charged particles radiate electromagnetic waves; those moving at constant velocity do not radiate. Although the force produced by an electron moving in the field of the bend magnet does not change the electron's speed, it does change the direction of the electron's velocity, and so the electrons radiate. It is this radiation that produces a broad spectrum extending into X-ray energies. At nonrelativistic speeds the radiation pattern and spectrum would resemble that of a radio-frequency antenna. However relativistic effects radically transform the radiation spectrum and angular radiation pattern, as a consequence of the relativistic length contraction and time dilatation, shifting the spectrum up into the X-ray region, and causing the radiation pattern to be limited to angles of order $1/\gamma \approx 10^{-4}$ (radians) within the orbital plane. This concentrates the X-rays in angle so that a fan of radiation that is highly collimated in the vertical direction (i.e., localized in the plane of the ring) is emitted from the bend magnets. This is ideal for use by silicon monochromators based on Bragg diffraction.

The total radiated power is given by Lienard's generalization of the Larmor formula

$$P = \frac{\mu_0 q^2 \gamma^6}{6\pi c} \left(a^2 - \left| \frac{\vec{v} \times \vec{a}}{c} \right|^2 \right),$$

where q is the electronic charge, \vec{a} and \vec{v} are the acceleration and velocity of the electron, c is the speed of light, and μ_0 is the magnetic permeability of free space. Note the very strong dependence of the radiated power on γ .

69.5.3 Bend Magnet Radiation

The spectrum of light that is radiated from an electron in circular motion can be expressed in a general manner through a function $g_1(x) = x \int_{t=x}^{\infty} K_{5/3}(t) dt$, where K_n is the modified Bessel function of order n ; $g_1(x)$ is plotted in Figure 69.7. For many purposes a useful approximation is $g_1(x) \approx 1.72x^{0.282}e^{-0.969}$. This function allows us to calculate the spectral flux that can be expected from a bend magnet source (or wiggler—see below) for a given beam current and geometry. The number of photons/second/milli-ampere/milliradian of horizontal angular acceptance and bandwidth $\Delta\epsilon/\epsilon = 10^{-4}$ is $1.256 \times 10^6 \gamma g_1(x)$, where $x = \epsilon/\epsilon_c$, ϵ is the photon energy, and ϵ_c is a parameter called the critical energy; it depends on electron beam energy E and magnetic field B of the bend magnet (or alternatively the bend radius ρ). In practical units $\epsilon_c \approx 0.665E^2B$, $\rho \approx 3.3E/B$, and $\gamma \approx 1957E$ with ϵ_c in keV, E in GeV, B in tesla, and ρ in meters.

69.5.4 Insertion Devices: Wigglers and Undulators

Bend magnets are required to get the electrons to circulate around the ring, and they make fine sources for many purposes. After a time accelerator physicists realized that there are strong benefits to inserting specifically designed magnetic structures into the beam path in the straight sections between the bend magnets. These insertion devices

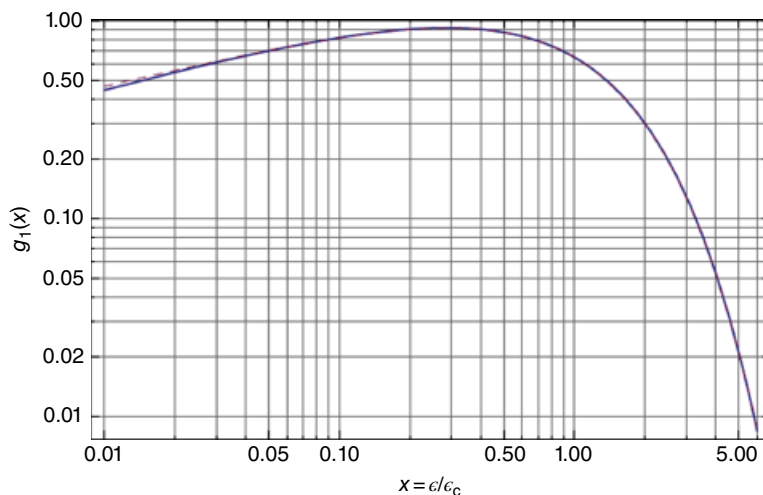


FIGURE 69.7 Energy spectrum $g_1(x)$ for bend magnets and wigglers. The solid curve is the exact function and the dashed curve is the approximation $g_1(x) \approx 1.72x^{0.282}e^{-0.969}$.

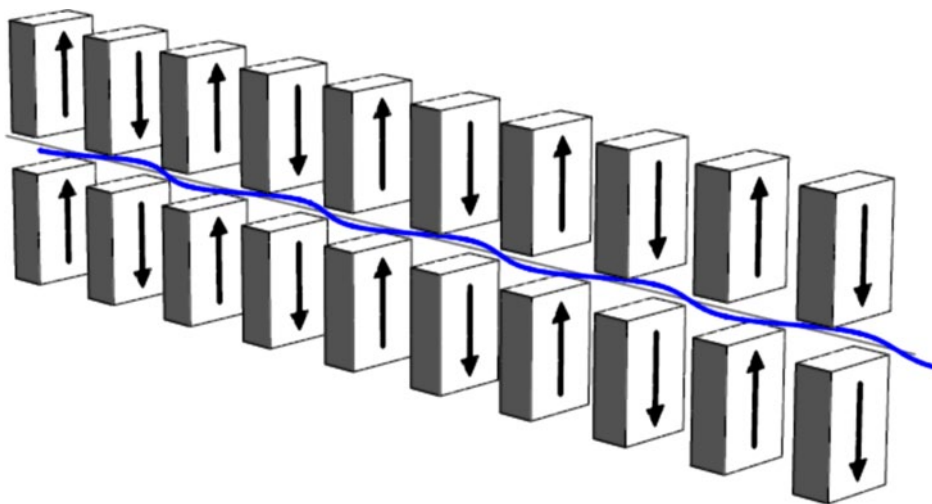


FIGURE 69.8 Schematic of an insertion device (wiggler or undulator). These devices use permanent magnets or electromagnets, either conventional or superconducting. The alternating vertical magnetic field (indicated by arrows) causes the path of the electron to undulate.

consist of periodic arrays of typically $N=50$ – 100 magnetic poles (of spatial period λ_0 , which is typically a few centimeter) as shown in Figure 69.8. If the angular deflection of the path is large compared to the angular width of the electron radiation pattern ($\approx 1/\gamma$), the X-rays emitted at each pole add up incoherently with those emitted at the other poles, so the spectrum is described by the $g_1(x)$ function with appropriate critical energy and multiplied by the number of poles. This is a wiggler.

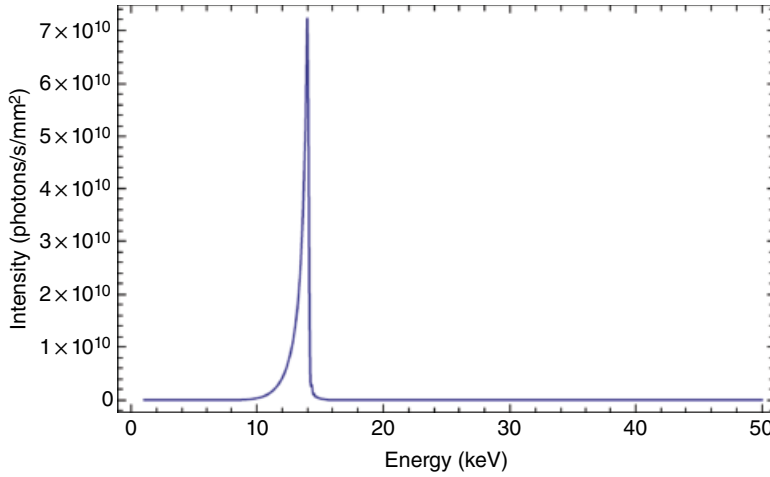


FIGURE 69.9 Computed APS type A undulator spectrum, $K=0.01$. The undulator period is 3.3 cm and the electron beam energy is 7 GeV. In this case a single peak is produced.

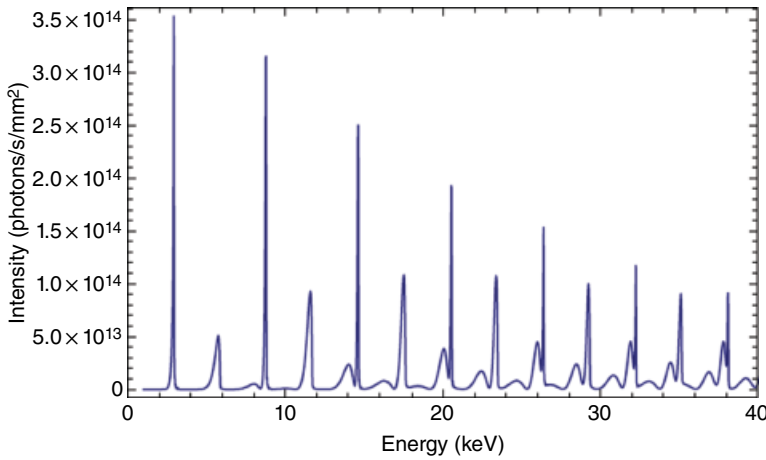


FIGURE 69.10 Computed APS type A undulator spectrum, $K=2.76$. The undulator period is 3.3 cm and the electron beam energy is 7 GeV. In this case many peaks are produced. The odd order harmonics are much stronger than the even order ones.

On the other hand, if the angular deflection $\delta_w = \lambda_0/(2\pi\rho)$ in the insertion device (where ρ is the bend radius of the trajectory) is less than the intrinsic divergence $1/\gamma$, the light emitted at each pole interferes with light emitted at the other poles. The interference effects cause the spectral peaks to be concentrated at specific energies, as shown in Figures 69.9 and 69.10, and the radiation pattern to be collimated horizontally and vertically along the axis of the insertion device within an angle $\Delta\theta \approx 1/(\gamma\sqrt{N})$. This is an undulator. The deflection is quantified using the “deflection parameter” $K = \gamma\delta_w$. The energy spectrum depends on observation angle because of the relativistic Doppler shift, with lower energies detected off-axis. The fractional width of the peaks

in energy $\Delta\epsilon/\epsilon$ is inversely proportional to the number of poles N . The wavelength of the peak radiation is $\lambda = \lambda_0 / (2\gamma^2)(1 + K^2/2 + \gamma^2\theta^2)$, where θ is the off-axis observation angle; the photon energy is then $\epsilon = hc/\lambda$. By adjusting the vertical separation between the magnets, the magnetic field strength can be systematically varied, which changes the deflection parameter, and shifts the energy of the peaks, which is used to maximize the flux that is passed subsequent X-ray optics.

69.6 BEAMLINES

Once generated, the X-rays that are produced by bend magnets or insertion devices are allowed to pass through a thin beryllium window (beryllium has very low atomic number, so it does not attenuate the beam much) and pass from the machine vacuum down into the beamline. Beamlines are complex instruments, usually tens of meters long, which through use of shielded enclosures safely convey the intense polychromatic (X-ray + UV–Vis + infrared) beam to the various X-ray optical systems, experimental enclosures, detectors, and data acquisition systems used for the measurements. The optics normally reside in shielded high-vacuum or ultrahigh vacuum environments to reduce air scattering and reduce damage by preventing ionization of gases. Cooled apertures or slits are used to define the central portion of the beam and absorb the undesired low-energy off-axis light. Many beamline components are now commercially available.

69.6.1 Instrument Control and Scanning Modes

Computer control is needed to orchestrate the various motions of the monochromator and other optics (shutters, slits, mirrors), detector readout, and sample position, laser firing, and other experimental variables. The incident flux cannot be assumed to be constant to 10^{-3} so it is essential to measure the detector signals precisely over identical fixed time intervals, so that the fluctuations in the incident intensity divide out properly. Typically motion control and data acquisition is done using dedicated VME- or VXI-based (or CAMAC-based) scalars, or GNU-/Linux-based computers, with fast real-time instrument control software such as that available in EPICS [9], Tango [10], MX [11], and others.

In step-scan mode the monochromator and other optics positions are moved (under stepper or DC servo motor control) to obtain a harmonic-free beam of the desired energy; a brief waiting period of a few hundred milliseconds is inserted to allow the vibrations of the optics to damp out; the detectors are then read out for a fixed time, or a fixed count; and the counts are recorded in a computer file. A different energy is then selected and the process is repeated to build up the sampled spectrum. In this step-scan mode, the energies are selected to provide adequate sampling of the spectrum. A pre-edge region of several hundred eV is normally measured with a coarse (≈ 10 eV) grid to measure the background trend. Over the absorption edge, where the spectrum may be changing rapidly, energy intervals on the order of 1 eV or less are commonly used.

Above the edge, the oscillations are approximately periodic in $k = \sqrt{0.263(E - E_0)}$ (in $\text{eV}\text{\AA}$ units), where E is the photon energy and E_0 is the absorption edge energy. By the Nyquist theorem, to obtain adequate sampling of the oscillations, a minimum sampling rate in k -space is needed; $\delta k \approx 0.05\text{\AA}^{-1}$ is sufficient, which implies a maximum spacing between energy points of $\delta E \approx 0.2\sqrt{E - E_0}$ at energy E above the edge. Although a uniform k -space grid above the edge is optimal, dividing the scan into ranges of different (but adequately sampled) energy subgrids is satisfactory. This is described further in [3].

An alternative approach is to use continuous scanning, by slewing the monochromator in a continuous motion, and averaging the acquired data points over known time intervals. This averages the spectrum over nearby energy values; corresponding averages must be made of the energy values themselves. This approach has the advantage of offering potentially shorter scan times, which can be used among other things to record transient phenomena. Since it is not necessary to wait for motors to move, and settling times to elapse, the duty cycle is generally better in continuous-scan mode than in step-scan mode. The time intervals for each data point must be carefully matched to the slew rate so as to get adequate spectral sampling.

“Dispersive XAFS” is done by dispersing photons of different energies at different angles using a bent crystal. The angle of incidence onto the crystal varies with position, so different wavelengths are diffracted from different parts of the bent crystal. The diffracted X-rays are arranged to pass through the sample and are detected with a spatially sensitive detector such as a photodiode array, area detector such as a CCD or pixel array detector, or (many decades ago) film. Since each small area of the bent crystal diffracts only a small range of wavelengths, this approach does not increase the overall intensity, but it does make it unnecessary to use any mechanical motions to do a scan, so it can be useful for fast phenomena.

Time-resolved experiments have become more common in recent years. These are usually pump-probe experiments in which a perturbation (e.g., laser pulse) is made to the sample, and at a fixed delay time relative to it, a fast measurement is made at fixed energy, typically with avalanche photodiode (APD) detectors. By altering the delay time, a record of the signal versus time can be built up. By altering the energy, an energy scan can be built up. The signal from a single pulse (emitted by a single electron bunch in the ring) normally is insufficient for adequate signal-to-noise (S/N) ratio, so signal averaging over many repetitions is required.

69.6.2 Double-Crystal Monochromators

A double-crystal monochromator (Fig. 69.11) selects a specific (but variable) energy band from the polychromatic beam for use in the experiment. Normally the monochromator uses Bragg diffraction from a pair of crystals (or diffraction gratings, for soft X-ray experiments). The first crystal defines the energy by setting the angle of incidence onto the crystal planes; the second crystal, parallel to the first, essentially acts like a mirror to direct the monochromatic diffracted beam parallel to its original

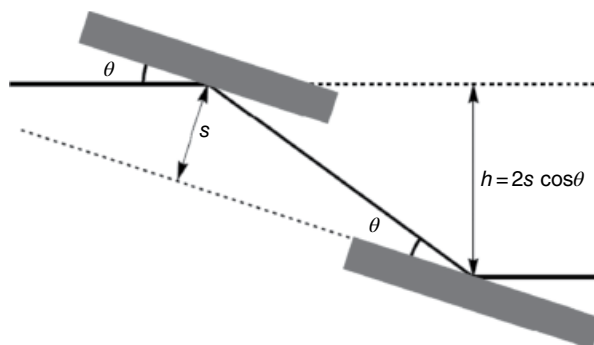


FIGURE 69.11 Schematic double-crystal monochromator. The rectangles represent crystals (typically silicon) and the solid lines show the path of the X-ray beam. The beam is displaced by a distance h , which depends on θ unless s is varied so as to compensate.

direction, down the beamline. Because the incidence angle and exit angle are equal, just as light reflects from a mirror, conventionally (but incorrectly) the diffracted beams are often referred to as “reflections.”

High-quality silicon crystals of large size are readily available from the semiconductor industry, and they have suitable properties for many purposes at hard X-ray energies. Germanium crystals are useful at high energies, and diamond crystals are of use because of their very high thermal conductivity. Silicon, germanium, and diamond all have the same crystallographic structure. Diffraction does not occur significantly for all values of hkl in the Bragg diffraction equation $n\lambda = 2d_{hkl} \sin\theta_B$, but only when hkl are all odd, or when hkl are all even and their sum is an integer multiple of 4: for example, $hkl = 111, 220, 311, 331, 400$ are all “allowed reflections.” Other crystal types will have different selection rules.

Synthetic multilayers are artificial structures consisting of alternating layers of high and low atomic number materials that act as X-ray interference coatings. Diffraction from these structures can be used for monochromators, but they do not provide sufficient energy resolution to define the beam energy in XAFS experiments. They can be used for collecting fluorescence however [12].

At a third-generation synchrotron source, a large amount of power (hundreds to thousands of watts) is typically deposited in the first crystal, and thermal management is a key aspect of the design. The angular tolerances for alignment of the monochromator crystals are quite demanding, typically on the order of seconds of arc, that is, tens of microradians. The second crystal may be sagittally bent to provide horizontal focusing.

69.6.3 Focusing Conditions

X-ray beams can be focused using reflective optics, that is, X-ray mirrors; diffractive optics such as bent crystals, asymmetrically cut crystals, and Fresnel zone plates; and refractive optics such as beryllium lenses. In comparison to properties at visible

wavelengths, at X-ray wavelengths the performance of these devices is strongly constrained because of the index of refraction of materials at X-ray energies, as described below.

Over a small region of a mirror, it will be locally flat, and the specularly reflected beam will have equal angles of incidence and reflection. The incident and reflected beams define a plane of reflection, and the mirror can be bent along an axis that is either perpendicular to that plane (meridional focusing) or parallel to the plane (sagittal focusing). For example, a horizontal mirror can be meridionally bent to provide vertical focusing or sagittally bent to provide horizontal focusing. The focusing equation for the sagittal case is $2\sin(\theta)/R_s = (1/u + 1/v)$; the focusing equation for the meridional case is $2/(R_m \sin(\theta)) = (1/u + 1/v)$. Evidently $R_s = R_m \sin^2(\theta)$, so $R_m \gg R_s$ for small angles of incidence. Here R_s and R_m are the radii of curvature for the sagittal and meridional cases, θ is the local angle of incidence onto the optic (mirror or bent crystal), and u and v are the source to optic distance and the optic to focus distance, respectively. By using a large source to optic distance u and small optic to focus distance v , the apparent size of the X-ray source (e.g., undulator) can be demagnified by the v/u , significantly reducing the size of the beam, at the cost of increased angular spread.

69.6.4 X-Ray Lenses and Mirrors

In the visible energy range, a light beam propagating in a medium with high index of refraction that is incident on an interface with a material of lower index will undergo total internal reflection, provided the angle of incidence is shallow enough. In the X-ray region the index of refraction is a complex number with a magnitude very slightly less than 1.0. In this case X-ray beams will undergo total external reflection at sufficiently shallow angles so that all materials act as X-ray mirror at shallow angles on the order of milliradians. This effect is used to make X-ray mirrors. Synchrotron X-ray mirrors are typically polished to a mean square roughness on the order of several Ångström.

The complex index of refraction \tilde{n} of materials in the X-ray regime can be written as $\tilde{n} = 1 - \delta - i\beta$, where $\delta = ne^2\lambda^2/2\pi mc^2$, $\beta = \mu\lambda/4\pi$, and μ is the X-ray absorption coefficient; λ is the X-ray wavelength, n is the number density of mobile electrons in the material, $e^2 \equiv q^2/4\pi\epsilon_0$, and q and m are the charge and mass of the electron. In pure elements this can be expressed as $\delta = N(Z/A)e^2\lambda^2/2\pi mc^2$ where ρ is the mass density, Z and A are, respectively, the atomic number and the atomic weight of the element, and N is Avogadro's number. The real and imaginary parts of $\tilde{n} = 1$, respectively, describe dispersion and absorption of the X-rays.

The closeness of \tilde{n} to 1.0 shows that the ability of materials to refract is quite limited. In recent years X-ray lenses have come into more common use at synchrotron radiation sources, particularly at high energies. In the X-ray energy region, a lens typically consists of a series of lens-shaped voids (or a more easily fabricated series of cylindrical holes) in a low atomic number material such as beryllium, which is used to minimize absorption.

Total external reflection occurs at angles $\theta < \theta_c$, where the critical angle $\theta_c = \sqrt{2\delta}$. θ_c varies inversely with X-ray energy, so that reflection of high-energy X-rays requires shallower angles of incidence onto the mirror. By choosing the angle appropriately, the mirror can be used as a low-pass filter, so that low energies are reflected and higher energies are absorbed by the mirror.

The product of energy E and θ_c is a property of the mirror surface and depends on its composition. Representative values of $E \cdot \theta_c$ (in keV-mrad) are ≈ 31 (Si), 59(Ni), 62(Pd), 67(Rh), 80(Au), and 82(Pt). θ_c depends on the nature of the coating, so it is often useful to deposit selected metals in stripes onto the X-ray mirrors. For example, it has proven useful to apply Pt and Rh coatings in stripes to an uncoated substrate made of a low expansion ceramic. By shifting the mirror sideways, different mirror coatings can be chosen without breaking vacuum. A high atomic number coating extends the range of angles that are reflected and allow use of a shorter mirror by reflecting at larger incidence angle. Absorption edges from the mirror coating over the energy range of interest generally are undesirable.

The reflectivity of mirrors (and multilayers) can be calculated using Fresnel's equations [13, 14]. For a sufficiently thick (at least tens of nanometers) mirror material or coating, the reflectivity (assuming zero roughness) may be written [15] in terms of the reduced angle $\phi = \theta/\theta_c$ as

$$R(\phi) = \frac{(\phi - A)^2 + B^2}{(\phi + A)^2 + B^2},$$

where

$$2A^2 = \left[(\phi^2 - 1)^2 + (\beta/\delta)^2 \right]^{1/2} + (\phi^2 - 1),$$

and

$$2B^2 = \left[(\phi^2 - 1)^2 + (\beta/\delta)^2 \right]^{1/2} - (\phi^2 - 1).$$

A contour plot of this theoretical reflectivity versus angle and energy is shown in Figure 69.12.

69.6.5 Harmonics

Inspection of the formula for the Bragg condition indicates that photons of an energy that is an integer multiple of the desired fundamental energy may pass through the monochromator. Some of these may be “forbidden reflections,” which are very weak, while others are “allowed reflections.” Whether a reflection is allowed or forbidden depends on crystal symmetry.

These higher harmonics must be eliminated or otherwise reckoned with, or serious systematic errors will occur in the absorption spectra. For diamond-structure

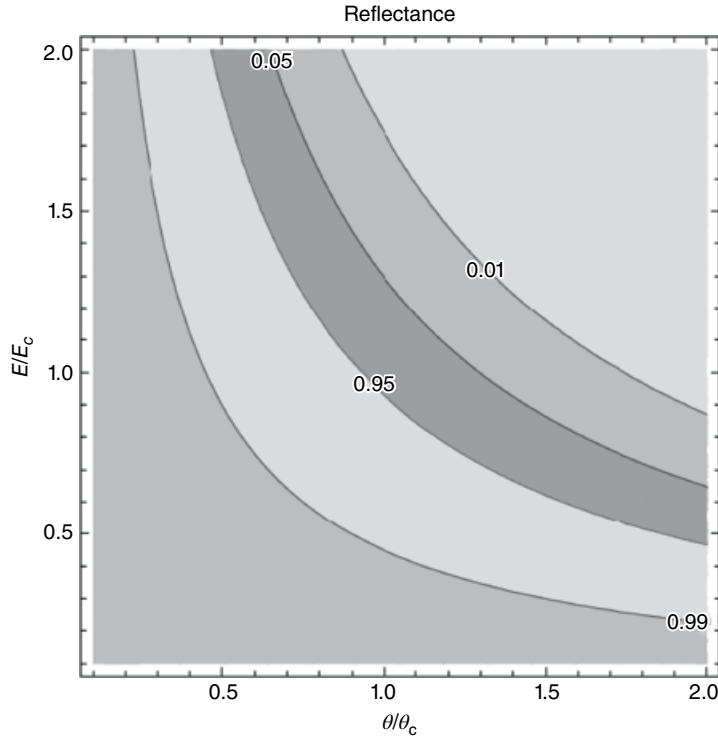


FIGURE 69.12 Contour plot of calculated mirror reflectivity versus angle for $\beta/\delta = 0.1$. Contour levels shown are 1, 5, 95, and 99% of maximum reflectivity.

monochromator crystals such as silicon and germanium (or diamond), reflections are only allowed if the indices $\langle h, k, l \rangle$ satisfy either of the following conditions: $\langle h, k, l \rangle$ are all odd integers, or $h + k + l = 4m$, where h, k, l, m are integers, because otherwise the structure factor [3] vanishes. Examples of allowed reflections are $\langle 1, 1, 1 \rangle$, $\langle 2, 2, 0 \rangle$, $\langle 3, 1, 1 \rangle$, $\langle 3, 3, 1 \rangle$, and $\langle 3, 3, 3 \rangle$. The allowed $\langle 3, 3, 3 \rangle$ reflection is three times the energy (“third harmonic”) of the $\langle 1, 1, 1 \rangle$ and must be eliminated if the $\langle 1, 1, 1 \rangle$ reflection is used in the experiments. The $\langle 2, 2, 2 \rangle$ reflection is forbidden.

Harmonics can be measured in several ways: by attenuating the beam with a series of attenuators of different thickness and known composition (which increases the proportion of harmonics, “beam hardening”) and solving a simple set of equations or using a scatterer and an energy dispersive detector (and some calibration) to measure the proportion of harmonic and fundamental.

There are several standard ways to eliminate harmonics. One method (“detuning the monochromator”) is to intentionally misalign the second crystal of the monochromator a small amount with respect to the first crystal, using a piezoelectric actuator. This reduces the amount of the fundamental energy photons that get through, typically by about half, but the higher-energy harmonics are attenuated to a much greater degree than is the fundamental.

Another common method of reducing harmonics is to use a harmonic rejection mirror, set to an angle that reflects the fundamental but does not reflect the harmonics. See “X-Ray Lenses and Mirrors.”

A third method that is not widely employed at synchrotron sources (because of their high intensities) is to use energy-sensitive detectors to selectively record photons at the fundamental energies and reject the harmonics by electronic means.

A type of device called a “beam cleaner” [16] is a medium-resolution (≈ 100 eV) bent Laue optic that can be used to select a particular harmonic (or fundamental) energy that is transmitted by the beamline’s primary high-resolution monochromator. This allows the experimenter to use higher-order harmonics for experiments, effectively extending the useful energy range of the beamline. Conversely it can be used to eliminate harmonics from the beam. Use of the medium-energy-resolution optic simplifies electromechanical tracking of the secondary monochromator with the primary monochromator, significantly reducing the cost and complexity.

69.7 DETECTORS

XAFS measurements require precise monitoring of the incident and transmitted fluxes I_0 and I . The incident intensity from a synchrotron source can fluctuate for a variety of reasons, but if the detectors are sufficiently linear and other precautions are taken, the fluctuations will divide out between the two detectors. Good linearity $\approx 10^{-4}$ is required separated for each detector. Even if the incident beam intensity I_0 were kept perfectly constant (which can be approximately arranged through use of beam intensity leveling servo controls), the I beam transmitted by the sample can vary by more than an order of magnitude, because of the energy-dependent absorption by the sample. Therefore it is not sufficient to simply make use of identical nonlinear I_0 and I detectors: it is necessary to use good quality detectors and be sure to operate them in the linear range or to otherwise compensate for their nonlinearities on a detector by detector basis. A good reference on detectors is the book by Knoll [17].

In fluorescence detection a large area detector is placed near the sample, typically at 90° to the beam in the horizontal plane, to collect the fluorescence. The X-rays produced by bend magnets, and planar wigglers and undulators, are normally polarized in the horizontal (orbital) plane of the ring. The elastic scattering, which causes an undesirable background, is minimum along the direction of the X-ray polarization vector.

Background is caused by scattered X-rays from the sample or undesired fluorescence from elements other than the one of interest. If the number of detected background photons were strictly constant, that background could simply be subtracted out, but because the signal; consists of photons, their number fluctuates, contributing noise. Let signal S be the number of the desired fluorescence photons that are detected in a counting interval, and let B be the number of background photons. The noise fluctuations vary as $\sqrt{S+B}$, so the S/N ratio is $S/\sqrt{(S+B)}$. It is convenient to define

$N_{\text{eff}} = S^2/(S+B) = S/(1+B/S)$, the effective number of counts, as the number of signal photons that would give the same S/N ratio in the absence of background. This shows that the presence of background reduces the effective counts by the ratio $1/(1+B/S) \approx S/B$ for large background-to-signal ratio. A background-to-signal ratio of 100 effectively increases the necessary counting time per point to reach a specified S/N ratio by a factor of 100. Therefore it is essential to reduce the amount of background when measuring dilute species.

The detectors for fluorescence are selected by the experimenter for a (sample-dependent) compromise between large area (or solid angle), good energy resolution/background rejection, and high maximum count rate. The most common detectors for fluorescence detection are multielement solid-state detectors (silicon or germanium), including silicon drift detectors (SDDs); large area fluorescence ionization chambers with Z-1 filters and slits; positive-intrinsic-negative (PIN) diodes/PIPS detectors; and scintillators with photomultipliers. Each of these can be supplemented with diffractive analyzers (synthetic multilayers, or Bragg or Laue geometry diffractive optics) in order to reduce background.

69.7.1 Ionization Chambers and PIN Diodes

The most common detectors to use in transmission mode XAFS experiments are ionization chambers, which consist of parallel conducting plates, internally supported by insulators, in a conducting box with X-ray transparent windows on each end, and containing a suitably chosen “fill gas,” for example, N_2 , Ar . Ionization chambers do not provide energy resolution—basically they measure the total amount of energy that is deposited in the chamber by absorption of photons. It takes approximately 30 eV on average to create an electron–hole pair in fill gases.

A voltage (typically several hundred volts) is applied between the plates so that when X-rays pass through the windows and between the conducting plates, partially ionizing the gas, the electric field pushes the negative electrons toward the anode and the positive ions toward the cathode. The resulting currents (typically 10–100 nA) are converted to voltage signals using a current amplifier (transimpedance or transconductance amplifier) and are measured by the data acquisition system. This often is done using a voltage to frequency converter (usually 100 kHz/V), which generates pulses at a frequency that is proportional to the input voltage. The pulses are then counted with a multichannel scaler for a fixed time on the order of 1 s. Alternatively, A/D converters can be used to digitize the signal. The time constants over which the input currents are smoothed/integrated by the amplifier should be the same for the I_0 and I channels; otherwise variations in flux of the incident beam will not accurately divide out.

It is essential to operate ionization chambers in what is called the “plateau region.” If the voltage applied to the ionization chamber is insufficient to separate the electron–ion pairs from each other, they will recombine, and the ionization event not be recorded, giving nonlinear response. For this reason the experimenter should verify that the

voltage is sufficient to obtain linear response. However if the voltage is too large, liberated electrons will be accelerated to sufficient energy before colliding with another gas atom that they can ionize that atom and create secondaries, so that one obtains a pulse of current whose total charge is proportional to the energy of the absorbed photon. This is a proportional counter. Increasing the voltage further yields a Geiger counter. These have their uses but have limited count rates, so normally for XAFS measurements a linear ionization chamber response is required.

PIN diodes and PIPS detectors (Canberra Inc.), when operated in current mode, essentially behave like solid-state ionization chambers, except they normally absorb nearly all of the X-rays incident upon them, and are therefore not used as semitransparent intensity monitors in the way ionization chambers are. PIN diodes can be used as intensity monitors by detecting scatter or fluorescence from thin foils inserted in the beam. They do not require high-voltage supplies to create sufficient electric field to separate the electrons from their corresponding positively charged “holes.” Additionally they can be operated in pulse-counting mode as described in the next section. PIN diodes are convenient for monitoring X-ray intensities when space is limited. They are sensitive to near-infrared and visible light so they must be shielded from it when using them for X-ray detection.

69.7.2 Solid-State Detectors, SDDs, and APDs

Solid-state (semiconductor) detectors, such as silicon or germanium detectors, and SDDs are “energy dispersive,” that is, they are normally used to provide a measure of energy resolution of the absorbed photons. A photon that is absorbed in the semiconductor material (Si or Ge) creates electron–hole pairs, which, as in an ionization chamber, are separated by application of an electric field, often associated with a high applied voltage, and collected by electrodes. This electric field would result in a current sufficient to damage the device, if the semiconductor were kept at room temperature, so the detector elements normally are cooled with liquid nitrogen or Peltier cooler, which also reduces electronic noise.

The charge pulses are preamplified and then passed through a shaping amplifier, which returns voltage pulses with heights that are proportional to the total charge of each pulse. These pulses have energies that are proportional to the energy of the absorbed photon, and they can then be selected with discriminators or sorted into bins with a multichannel analyzer (MCA) to provide an energy spectrum.

Alternatively, a more modern approach (e.g., X-ray Instrumentation Associates Digital X-ray Processor) uses a very high-speed analog-to-digital converter (flash ADC) to sample and digitize the pulses and proprietary algorithms running on field-programmable gate arrays (FPGAs) to perform the quantitative pulse analysis. These devices effectively combine the functionalities of a high-voltage supply, preamplifier, shaping amplifier, and MCA into a single compact unit per detector channel. Devices similar to this are available from several vendors.

APDs are fast detectors that operate effectively like solid-state Geiger counters. They offer little to no energy discrimination but are useful for ultrafast experiments.

69.8 SAMPLE PREPARATION AND DETECTION MODES

69.8.1 Transmission Mode

As described above, in transmission mode the quantity that is measured is the fraction of X-rays that are transmitted through a sample at a given energy. To do this the sample must be thin enough to allow sufficient X-rays to penetrate the sample so that adequate photon statistics are obtained, while it must be thick enough to provide adequate absorption contrast as a function of energy. Although theoretically the optimal S/N ratio is obtained when the edge step $\Delta\mu x \approx 2.5$, where $\Delta\mu$ is the difference in absorption coefficient above and below the edge and x is the sample thickness, in practice systematic errors such as “thickness effects” become worse for thicker samples, and experience has shown that in most cases restricting $\Delta\mu x \approx 1$ gives more reliable results. If the edge step is less than about 0.1, it may be better to consider fluorescence measurements. Thickness effects generally reduce the apparent amplitude of structure in the spectra. These effects are described in more detail in Bunker [3].

Samples for transmission mode can be polycrystalline, crystalline, or amorphous films, coatings, and powders on an X-ray thin support or bound in a compressed pellet. It is important that the sample is homogeneous and of uniform thickness (meaning the integrated absorption through the sample is constant from point to point over the illuminated area of the sample). A variation in thickness introduces a nonlinear dependence of the apparent absorption $(\mu x)_{\text{eff}}$ on the absorption coefficient μ :

$$(\mu x)_{\text{eff}} = \mu \bar{x} - \mu^2 \sigma^2 / 2 + \dots$$

where σ^2 is the variance of the thickness distribution [3].

69.8.2 Fluorescence Mode

A schematic fluorescence mode experiment is shown in Figure 69.13. In this case the sample typically is placed at 45° to the incident beam, and the fluorescence detector is placed at 90° in the horizontal plane, to minimize elastic scatter. If there are effective ways to eliminate scatter, it can be beneficial to collect a greater solid angle by using multiple, or larger, detectors.

The measured fluorescence signal is proportional to

$$\mu_i(E) \csc(\theta_{\text{in}}) (1 - \exp(-\alpha\tau) / \alpha),$$

where $\mu_i(E)$ are, respectively, the absorption coefficients of the element of interest and the total sample absorption coefficient at energy E , τ is the sample thickness, $\alpha = \mu_t(E_p)$

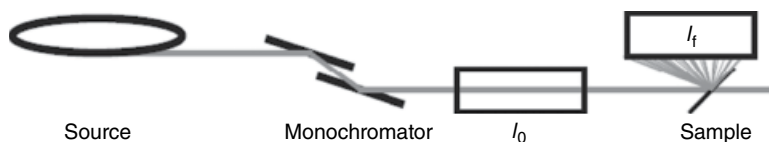


FIGURE 69.13 Schematic fluorescence mode X-ray absorption measuring apparatus (not to scale).

$\csc(\theta_{\text{in}}) + \mu_{\text{T}}(E) \csc(\theta_{\text{out}})$, $\mu_i(E)$, and θ_{in} and θ_{out} are the incidence and exit angles measured relative to the surface. This expression in principle should be integrated over the various exit angles subtended by the detector, but this simple form is sufficient for our purposes.

For thin samples, that is, $\alpha\tau \ll 1$, this becomes $\mu_i(E) \csc(\theta_{\text{in}})\tau$ as expected: the measured signal is proportional to the absorption coefficient times the projected sample thickness. In this case the measured fluorescence is proportional to the absorption coefficient.

For samples that are thick, if the element of interest does not contribute significantly to the total absorption coefficient μ_{T} , then the measured fluorescence depends linearly on the absorption coefficient μ_i but with an energy-dependent prefactor that depends on the composition of the sample matrix. Generally this can be calculated and compensated for.

For samples that are thick, if the element of interest *does* contribute significantly to the total absorption coefficient μ_{T} , the measured fluorescence depends nonlinearly on the absorption coefficient μ_i . In this case measured spectra can be seriously distorted if no other measures are taken. The nature of these distortions typically is that fine structure in the measured spectra is suppressed. This effect is called self-absorption or over-absorption, because the origin of the effect is a variation in effective penetration depth that depends on the absorption coefficient of the element of interest. As the true absorption $\mu_i(E)$ increases, the measured fluorescence does not increase in proportion, and features may be suppressed.

Algorithms have been developed [3] to computationally correct for these effects if the composition of the sample is known. Alternatively, by choosing $\theta_{\text{in}} \approx 90^\circ$ and $\theta_{\text{out}} \approx 0^\circ$, these effects can be significantly reduced, at the cost of reduced solid angle subtended by the detector.

69.8.3 HALO

A mnemonic, “HALO,” has proved helpful in remembering to take measures that will reduce errors in XAFS measurements. It stands for harmonics, alignment, linearity, and offsets.

Harmonics must be eliminated from the beam; sample alignment must be checked, so that nothing intercepts the beam between the I_0 detector and I (or I_f) detector except

for the homogeneous sample and homogeneous windows; linearity of detectors must be maintained (e.g., by checking ion chambers are used in their plateau region) or corrected, if nonlinear (e.g., by performing dead time corrections in pulse-counting detectors); and offsets (dark currents/amplifier offsets) must be measured and subtracted out before dividing by reference signals.

It should be mentioned that errors will be introduced if the I_0 detector (and in transmission the I detector) is placed too close to the sample, allowing fluorescence from the sample to get into the detector(s). This additional signal slightly corrupts the measured signal(s) and should be avoided.

69.8.4 Sample Geometry and Background Rejection

Several strategies are used to reject scattered background and unwanted extraneous fluorescence before they reach the detector. These include use of beam polarization; grazing incidence; attenuators, filters, and slits; and diffractive analyzers.

69.8.4.1 Use of Polarization Normally at synchrotron radiation sources (bend magnets, wigglers, planar undulators), the X-rays from the source are linearly polarized with the electric field vector in the horizontal plane; also the polarization is perpendicular to the direction of propagation. The scattering is minimum along the polarization direction, so fluorescence detectors are often placed in that location relative to the sample, that is, centered in the horizontal plane and perpendicular to the direction of travel of the X-ray beam. This significantly reduces background from elastic scatter.

69.8.4.2 Grazing Incidence At sufficiently small angles of incidence (typically milliradians) onto a flat surface, total external reflection from the sample will occur (see “X-Ray Lenses and Mirrors”). This has the effect of enhancing the X-ray field and confining it to a region very close to the surface of the sample. This confers a surface sensitivity to fluorescence experiments. It is also possible to determine the absorption coefficient and provide depth-dependent information by carefully measuring the reflected beam intensity as a function of angle and energy.

Even when the angle of incidence is too large for total external reflection to occur, a shallow (“grazing” or “glancing”) angle of incidence enhances the surface sensitivity for purely geometrical reasons, which can be beneficial, particularly if elastic scatter or fluorescence from the substrate produces undesired background.

69.8.4.3 Electron Yield for Reducing Background Use of electron yield detection can eliminate problems associated with self-absorption effects in fluorescence and can reduce the effects of fluorescence from the sample substrate. As described above, most XAFS experiments are carried out in transmission mode or fluorescence mode. A third mode is to measure (directly or indirectly) the electrons that are ejected from the

surface of the sample from nonradiative deexcitation of the core hole that is produced by the initial absorption event. If the samples can be placed in a vacuum (as is typically done for soft X-rays), the electrons can be collected and integrated to produce a signal current. Alternatively, the sample can be surrounded by helium, and the “conversion electrons” that are produced by collisions of the electrons that are ejected from the sample with the helium atoms can be collected by applying an electric field. The advantage of this method is that it confers strong surface sensitivity, because the electrons within the sample have path lengths only on the order of 1000 \AA , so only those atoms that are near the surface are visible by this method. It should be mentioned that this method can be difficult when using high-intensity undulator beams, because of sample charging and other effects.

69.8.4.4 X-Ray (Z-1) Filters and Attenuators In fluorescence detection of dilute species, if the fluorescence energy is significantly below the absorption edge of the element of interest, it may be possible to make a foil of containing a different element (typically one atomic number lower in the periodic table) that has its own absorption edge between the fluorescence energy of interest and the energy of excitation. When this is possible, it may be possible to selectively attenuate the elastically scattered background while only moderately absorbing the desired fluorescence. Of course this “Z-1 filter” itself will fluoresce, but through careful use of slits much of this refluorescence can be blocked from entering the detector. Significant improvements in S/N ratio can be obtained with an optimized system.

If there is a large amount of unwanted fluorescence at much lower energies than the desired fluorescence, for example, fluorescence from an element of lower atomic number, it may be beneficial to place an attenuator such as aluminum foils between the sample and the detector. The absorption coefficient of most elements varies as $1/E^3$ over energy regions in which there are no absorption edges, so the absorption coefficient is approximately eight times larger at half the energy. Choosing an attenuator of optimal thickness can significantly improve the S/N ratio. The optimal thickness can be calculated or measured experimentally by optimizing the effective counts as a function of attenuator thickness.

69.8.4.5 Diffractive Analyzers Since the desired fluorescence is found at a well-defined energy, it may be possible to accept a useful fraction of the fluorescence with a diffractive analyzer, provided the spot size on the sample is sufficiently small. The difficulty is that the fluorescence radiation is emitted in all directions into 4π solid angle, and it is difficult for an optic to collect a large fraction of that. In order to meet the Bragg condition across a curved crystal, the correct shape is a logarithmic spiral that is specific to the fluorescence energy and the crystals and crystal cut that are used. To approximate that precise shape, the crystals must either be bent or a series of crystals may be used to approximate a bent surface. The incident and diffracted X-rays may be on the same side of the crystal (Bragg geometry) or the diffracted X-rays may pass

through the crystal and emerge from the other side (Laue geometry). Laue geometry can be advantageous for improving the collection efficiency because the X-rays can be incident at a small angle with respect to the local surface normal. The angle tolerances for perfect crystals in Bragg geometry are very small, requiring high precision. In Laue geometry, under strongly bent conditions and suitable asymmetric cut crystals (in which the diffracting planes are parallel to the crystal surface), the perfect crystals have a much broader angular acceptance, which improves the throughput. Diffractive analyzers (some of which are commercially available) have been made with silicon (Bragg and Laue) [18–20], LiF (Bragg), pyrolytic graphite (Bragg) [21], and graded synthetic multilayers (Bragg) [12].

69.8.5 Oriented Samples

In single crystals or otherwise oriented samples, the absorption coefficient depends on the relative orientation of the X-ray polarization vector and the crystal axes. The absorption coefficient is well approximated as a second rank tensor (in the dipole approximation). It is usually assumed that these complications average out to near zero when the sample is in polycrystalline or solution form, and they automatically vanish for crystals of cubic symmetry (the absorption is isotropic). However if the sample is a powder and the grains have a nonrandom orientation (“preferred orientation” or “texture”), systematic errors can result. Magic angle spinning can be used to average out such undesired effects for crystalline samples [3].

In this case the sample is spun around the sample normal direction \hat{n} , which makes an angle ($\arccos(1/\sqrt{3}) \approx 54.74^\circ$) (“magic angle”) relative to the X-ray polarization direction (Fig. 69.14). This averages out the off-diagonal terms of the tensor and makes the resulting averaged absorption equivalent to the isotropic case. There should be an integer number of full revolutions of the sample rotation (or at least a large number of rotations) within a signal integration period.

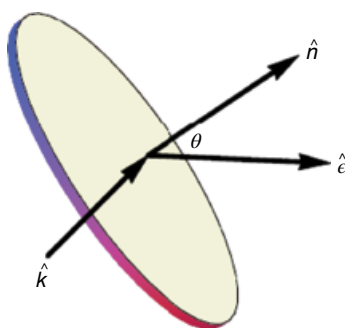


FIGURE 69.14 Magic angle spinning geometry. \hat{n} is the sample normal direction, \hat{e} is the electric polarization direction, and \hat{k} is the direction of the incident beam. θ is set to the magic angle 54.7° .

69.9 ABSOLUTE MEASUREMENTS

In routine XAFS measurements it is normal to measure the absorption of windows and air path integrated with the data and then to remove it through numerical background subtraction using cubic splines or other means. Similarly the energy dependence of detector efficiency contributes an additive background to the transmission mode XAFS spectra, and an estimate of that effect can be computationally generated and removed. In fluorescence and electron yield mode, this same effect multiplies the data by a slowly varying energy-dependent function, which will introduce a small systematic error in the amplitudes if uncorrected. Of course these background effects can be measured directly and corrected for but at the cost of valuable additional beam time. In routine experiments the numerical approach is conventionally used.

However for more accurate or absolute measurements, these effects and others must be accounted for. An example is measuring or correcting for the signals from scattered and fluorescence X-rays that are emitted by the sample and that may be absorbed in the detectors. If the detectors are too close to the sample, these effects can be significant; they can be reduced through use of suitable masks and evacuated pipes (“flight tubes”) to increase the distance between the sample and the detectors.

Other confounding effects can stem from temperature and pressure variations within ionization chambers, which can affect the gas density in a time-dependent manner. A 3°C variation in the ionization chamber temperature can cause a 1% variation in detector output. These environmental influences can be controlled in obvious ways.

Examples of high-accuracy absolute measurements of X-ray absorption are given in Glover and Chantler [22] and subsequent work. Absolute measurements of the X-ray absorption coefficient also require precise characterization of the sample uniformity and thickness.

REFERENCES

1. E.A. Stern and S.M. Heald “Basic Principles and Applications of EXAFS”, in E.E. Koch (ed), *Handbook on Synchrotron Radiation*, North-Holland Publishing Company, Amsterdam/New York/Oxford, 955–1014, (1983).
2. D.C. Koningsberger and R. Prins *X-Ray Absorption: Principles, Applications, Techniques of EXAFS, SEXAFS, and XANES*, John Wiley & Sons, New York, (1988).
3. G.B. Bunker *Introduction to XAFS: A Practical Guide to X-Ray Absorption Fine Structure Spectroscopy*, Cambridge University Press, Cambridge, UK, (2010).
4. C.T. Chantler, K. Olsen, R.A. Dragoset, J. Chang, A.R. Kishore, S.A. Kotochigova, and D.S. Zucker, X-Ray Form Factor, Attenuation and Scattering Tables (version 2.1) (2005). [online] <http://physics.nist.gov/ffast> (accessed on December 2, 2015).
5. Lawrence Berkeley Labs, Center for X-Ray Optics, X-Ray Data Booklet. [online] <http://xdb.lbl.gov/> (accessed on December 2, 2015).

6. [online] <http://www.csrr.iit.edu/periodic-table.html> (accessed on November 4, 2015).
7. [online] <http://www.lightsources.org/fels> (accessed on November 4, 2015).
8. [online] <http://www.lightsources.org> (accessed on November 4, 2015).
9. [online] <http://www.aps.anl.gov/epics/> (accessed on November 4, 2015).
10. [online] <http://www.tango-controls.org/> (accessed on November 4, 2015).
11. [online] <http://mx.iit.edu/> (accessed on November 4, 2015).
12. K. Zhang, G. Rosenbaum, R. Liu, C. Liu, C. Carmeli, G. Bunker, and D. Fischer "Development of multilayer analyzer array detectors for X-ray fluorescence at the third generation synchrotron source eighth international conference on synchrotron radiation instrumentation" *AIP Conf. Proc.*, 705, 957–960, (2004).
13. D.J. Griffiths *Introduction to Electrodynamics*, Third Edition, Prentice Hall, Saddle River, NJ, (1999).
14. J. Als-Nielsen and D. McMorrow *Elements of Modern X-Ray Physics*, John Wiley & Sons, Ltd, Chichester, UK, (2011).
15. T. Matsushita and H. Hashizume "X-Ray Monochromators", in E.E. Koch (ed), *Handbook on Synchrotron Radiation*, Vol. 1, North-Holland Publishing Company, Amsterdam/New York/Oxford, 261, (1983).
16. C. Karanfil, L.D. Chapman, G.B. Bunker, C.U. Segre, and N.E. Leyarovska "A 'beam cleaner' for harmonic selection/rejection" *Rev. Sci. Instrum.*, 73, 3, 1505, (2002).
17. G.F. Knoll *Radiation Detection and Measurement*, John Wiley & Sons, Inc., Hoboken, NJ, (2002).
18. B.W. Adams and K. Attenkofer "An active-optic X-ray fluorescence analyzer with high energy resolution, large solid angle coverage, and a large tuning range" *Rev. Sci. Instrum.*, 79, 023102-1–12, (2008).
19. Z. Zhong, L. Chapman, B. Bunker, G. Bunker, R. Fischetti, and C. Segre "A bent Laue analyzer for fluorescence EXAFS detection" *J. Synchrotron Radiat.*, 6, 212, (1999).
20. C. Karanfil, G. Bunker, M. Newville, C.U. Segre, and D. Chapman "Quantitative performance measurements of bent crystal Laue analyzers for X-ray fluorescence spectroscopy" *J. Synchrotron Radiat.*, 19, 375380, (2012).
21. D.M. Pease, M. Daniel, J.I. Budnick, T. Rhodes, M. Hammes, D.M. Potrepka, K. Sills, C. Nelson, S.M. Heald, D.I. Brewe, and A. Frenkel "Log spiral of revolution highly oriented pyrolytic graphite monochromator for fluorescence X-ray absorption edge fine structure" *Rev. Sci. Instrum.*, 71, 3267, (2000).
22. J.L. Glover and C.T. Chantler "The analysis of X-ray absorption fine structure: beam-line independent interpretation" *Meas. Sci. Technol.*, 18, 29162920, (2007).

NUCLEAR MAGNETIC RESONANCE (NMR) SPECTROSCOPY

KENNETH R. METZ

Chemistry Department, Merkert Chemistry Center, Boston College, Chestnut Hill, MA, USA

70.1 INTRODUCTION

Nuclear magnetic resonance (NMR) spectroscopy is arguably the single most powerful instrumental technique used in modern chemistry and biochemistry, and it also plays significant roles in physics, materials science, physiology, and medicine. The amount of information available from NMR studies is enormous and ranges from the structures of molecules to the details of molecular interactions and the dynamics of molecular motions. It is nondestructive and noninvasive, a crucial asset for many investigations. Virtually any sample type (solid/liquid/gas, or animal/vegetable/mineral) is amenable to study by this method. The primary limitation of NMR is its relatively low sensitivity compared to other forms of spectroscopy, which usually mandates the use of large samples or long data acquisition times, if not both.

Although it is fair to characterize NMR as a mature technique, the field continues to evolve dynamically with frequent, regular advances in both hardware and methodology. Nearly 70 years of continuous, aggressive development has provided investigators with a staggering range of NMR techniques and experiment types. At first glance, so much technology can seem bewildering and intimidating. However, the great majority of NMR studies employ a relatively small set of pulse experiments. This chapter presents the basic principles and some experimental aspects of NMR.

70.2 HISTORICAL REVIEW

At the heart of NMR is nuclear spin, the existence of which was confirmed by the 1924 atomic beam experiments of Stern and Gerlach [1]. Magnetic resonance with atomic beams was actually achieved by Rabi in 1938 [2]. (Nobel Prizes in Physics for that work were awarded to Stern in 1943 and to Rabi in 1944.) The next logical step was to observe NMR in condensed matter. Initial attempts at ^7Li NMR by Gorter and Broer [3] in 1942 were unsuccessful, mainly due to bad luck, before World War II interrupted research in the field. Aided by wartime developments in radio technology and knowing that NMR represented a still-unplucked “plum,” several major physics groups were poised to tackle the problem when the war ended. As a result, in 1945, two American groups succeeded in observing ^1H NMR signals in water and paraffin for the first time [4, 5]. The importance of that work resulted in physics Nobel Prizes in 1952 for the lead investigators of both groups (Felix Bloch at Stanford and Edward Purcell at Harvard). Physicists initially hoped that NMR methods would allow the measurement of exquisitely accurate values for magnetic moments and fundamental physical constants like the gyromagnetic ratio (γ), but Knight [6], followed by Proctor and Yu [7], soon discovered that the experimental values depended on which chemical compound was studied. This was termed the “chemical shift,” and it suddenly made NMR highly interesting to chemists for studies of molecular structures. The 1950s and the early 1960s witnessed enormous development in NMR technology, including the introduction of the first commercial spectrometers, routine line narrowing by sample spinning, magnetic field stabilization (which revealed spin–spin coupling), decoupling methods, and “magic angle” sample spinning for studies of solid samples. In 1966, Richard Ernst and Weston Anderson, working at Varian Associates, took advantage of the growing availability of digital computers to introduce pulsed Fourier transform NMR (FTNMR) [8]. By the mid-1970s, spectrometers could be purchased with dedicated minicomputers to perform the data acquisitions and calculations required for FTNMR. During that same decade, high-field superconducting magnets became available, which displaced the large electromagnets that dominated early NMR. The signal/noise improvements that high-field spectrometers provided stimulated the extension of NMR to nuclides beyond ^1H , ^{13}C , ^{19}F , and ^{31}P . In fact, virtually every element in the periodic table became susceptible to study by NMR, opening the door to important new investigations in physical, inorganic, biological, and organometallic chemistry. The dynamic decade of the 1970s witnessed the beginnings of another pivotal development in NMR: multidimensional experiments. Much of the work in two-dimensional (2D) NMR occurred in the research groups of Richard Ernst (Zürich) and Ray Freeman (Oxford), with Ernst receiving the 1991 Nobel Prize in Chemistry for that and his many other contributions to NMR. Multidimensional methods proved to be powerful tools for structure assignments in small molecules, and they revolutionized the determination of three-dimensional (3D) solution-state structures of macromolecules such as proteins. The 2002 Nobel Prize in Chemistry was shared by Kurt

Wüthrich for the work of his group in that area. Biomedical applications of NMR have driven a great deal of development since the mid-1970s. That began with techniques for noninvasive ^1H , ^{13}C , and ^{31}P measurements of metabolite concentrations in perfused tissues and live animals. It soon came to include mapping the spatial distributions of water and fat ^1H NMR signals in humans and animals, which, under the name “magnetic resonance imaging” (MRI), plays a crucial role in modern medicine. In 2003, Paul Lauterbur and Peter Mansfield shared the Nobel Prize in Physiology or Medicine for their early work in that area. Clearly, NMR has been a mainstream technique for over half a century and has contributed to major progress in a very wide range of applications, yielding Nobel Prizes for eight investigators!

70.3 BASIC PRINCIPLES OF SPIN MAGNETIZATION

NMR spectroscopy operates by manipulating nuclear spins. Nuclear spin is quantized, with a nuclear spin quantum number generally represented by the symbol I . A great many different isotopes (or “nuclides”) are endowed by nature with nuclear spins and are thus amenable to study by NMR. That includes at least one nuclide for nearly every element in the periodic table. However, NMR observability is by no means universal. Specifically, nuclides having even atomic numbers *and* even mass numbers possess no nuclear spin (i.e., $I=0$) and cannot be studied by NMR. For example, ^{12}C with its atomic number of 6 and atomic mass of 12 yields no NMR signal. Nuclides that combine an odd atomic number with even atomic mass possess integer spin values, such as $I=1$ for ^2H and ^{14}N and $I=3$ for ^{10}B . Finally, an odd atomic mass (regardless of atomic number) produces half-integer spins, such as ^1H , ^{13}C , and ^{31}P with $I=1/2$, ^{23}Na and ^{11}B with $I=3/2$, and ^{17}O with $I=5/2$. Table 70.1 lists some of the more important of the available NMR nuclides, along with their I values and other NMR parameters.

Fortunately, some very common nuclides like ^1H and ^{31}P are NMR active, and that has certainly contributed to the tremendous utility and explosive growth of NMR over the years. At the same time, it is tempting to lament the NMR inactivity of other nuclides such as ^{12}C and ^{16}O , the most abundant isotopes of those important elements. Although it is still possible to perform NMR studies with carbon and oxygen, the rare nuclides ^{13}C and ^{17}O must be used, resulting in relatively noisy spectra and long data acquisition times. If ^{12}C and ^{16}O did possess nuclear spin, however, spin coupling would greatly complicate routine ^1H NMR spectroscopy for most samples. So, in that respect, one can view the absence of nuclear spin in ^{12}C and ^{16}O as a “glass half full.”

The nonzero spin of a nucleus endows it with a magnetic moment μ that behaves like a tiny, spinning bar magnet (Fig. 70.1). The magnitude of μ for a particular nucleus is given by

$$\mu = \gamma \left(\frac{h}{2\pi} \right) [I(I+1)]^{1/2}, \quad (70.1)$$

TABLE 70.1 NMR Properties of Selected Nuclides

Nuclide	% Abundance	Spin I	Gyromagnetic Ratio γ (MHz T ⁻¹)	Larmor Freq. ν (MHz) at 11.744 T ^a	Relative Sensitivity ^b	Relative Receptivity ^c
¹ H	99.985	1/2	42.576	500.01	1.00	1.00
² H	0.0156	1	6.5355	76.753	9.65×10^{-3}	1.50×10^{-6}
⁷ Li	92.58	3/2	16.546	194.32	2.93×10^{-1}	2.72×10^{-1}
¹⁰ B	19.58	3	4.5754	53.733	1.99×10^{-2}	3.89×10^{-3}
¹¹ B	80.42	3/2	13.660	160.42	1.65×10^{-1}	1.33×10^{-1}
¹³ C	1.108	1/2	10.705	125.72	1.59×10^{-2}	1.76×10^{-4}
¹⁴ N	99.63	1	3.0755	36.119	1.01×10^{-3}	1.00×10^{-3}
¹⁵ N	0.37	1/2	-4.3142	50.666	1.04×10^{-3}	3.85×10^{-6}
¹⁷ O	0.037	5/2	-5.7719	67.785	2.91×10^{-2}	1.08×10^{-5}
¹⁹ F	100	1/2	40.054	470.39	8.33×10^{-1}	8.33×10^{-1}
²³ Na	100	3/2	11.262	132.26	9.25×10^{-2}	9.25×10^{-2}
²⁷ Al	100	5/2	11.103	130.39	2.07×10^{-1}	2.07×10^{-1}
²⁹ Si	4.70	1/2	-8.4577	99.327	7.84×10^{-2}	3.68×10^{-4}
³¹ P	100	1/2	17.235	202.41	6.63×10^{-2}	6.63×10^{-2}
³⁹ K	93.08	3/2	1.9869	23.334	5.08×10^{-4}	4.73×10^{-4}
⁵⁹ Co	100	7/2	10.054	118.07	2.77×10^{-1}	2.77×10^{-1}
¹⁰⁹ Ag	48.48	1/2	-1.9813	23.268	1.01×10^{-4}	4.89×10^{-5}
¹¹³ Cd	12.26	1/2	-9.4427	110.90	1.09×10^{-2}	1.34×10^{-3}
¹³³ Cs	100	7/2	5.5846	65.586	4.74×10^{-2}	4.74×10^{-2}
¹⁹⁹ Hg	16.84	1/2	7.5902	89.139	5.67×10^{-3}	9.54×10^{-4}
²⁰³ Tl	29.52	1/2	24.332	285.76	1.87×10^{-1}	5.51×10^{-2}
²⁰⁵ Tl	70.48	1/2	24.570	288.55	1.92×10^{-1}	1.35×10^{-1}

^aLarmor (NMR resonance) frequency calculated from Equation 70.2. Signs are irrelevant in this context and have been ignored.

^bNMR sensitivity at constant magnetic field relative to an equal number of ¹H nuclei. Calculated from $\gamma^3 \cdot I(I+1)$, ignoring signs.

^cNMR receptivity, including effects of natural abundance, calculated from Equation 70.6.

**FIGURE 70.1** Magnetic moment μ of a nucleus with nonzero spin.

where h is Planck's constant and γ is the gyromagnetic (or "magnetogyric") ratio, a fundamental constant for each nuclide. Table 70.1 lists the γ value for each nuclide. The standard units are $\text{rad s}^{-1} \text{T}^{-1}$ (where T represents teslas), but for the table, radians have been converted to Hertz to permit the use of somewhat more convenient units of MHz T^{-1} .

In an NMR experiment, the nucleus is placed in an external magnetic field, causing the μ vector to precess about the field direction in analogy to the precession of a gyroscope in a gravitational field. The precession frequency is given by the so-called Larmor equation,

$$\begin{aligned}\omega_0 &= \gamma B_0 \text{ (in rad s}^{-1}\text{), or} \\ \nu_0 &= \frac{\gamma B_0}{2\pi} \text{ (in Hz),}\end{aligned}\tag{70.2}$$

where B_0 represents the magnetic field strength or, more correctly, its flux density (in teslas). Typical laboratory magnets produce fields that result in Larmor frequencies in the radiofrequency (RF) range for most nuclides, as shown in Table 70.1. Thus, for a B_0 of 11.744 T, the Larmor frequency of ^1H is 500.0 MHz, as required by the ^1H gyromagnetic ratio of $+42.5770 \text{ MHz T}^{-1}$.

In addition to causing nuclear precession, the external magnetic field forces the nucleus to adopt a specific orientation. A spin-1/2 nucleus orients either parallel or antiparallel to the direction of the field, as illustrated in Figure 70.2. The two states differ in the sign of μ and have slightly different energies U :

$$U = -\mu B_0.\tag{70.3}$$

A small majority of spins adopt the lower-energy parallel orientation, in turn causing the entire sample to develop a small bulk magnetization vector μ_{bulk} in the direction of the external magnetic field. Because the energy difference between the two states is tiny, only a slight excess of spins (on the order of 0.001–0.0001% for ^1H) adopt the parallel orientation to the field, and the bulk magnetization vector of the sample is miniscule. It is this vector that produces the NMR signal, which explains why NMR signals are generally so weak.

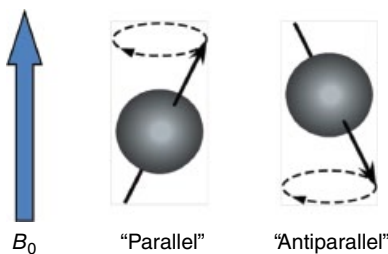


FIGURE 70.2 Possible orientations of spin-1/2 nuclei in an external magnetic field.

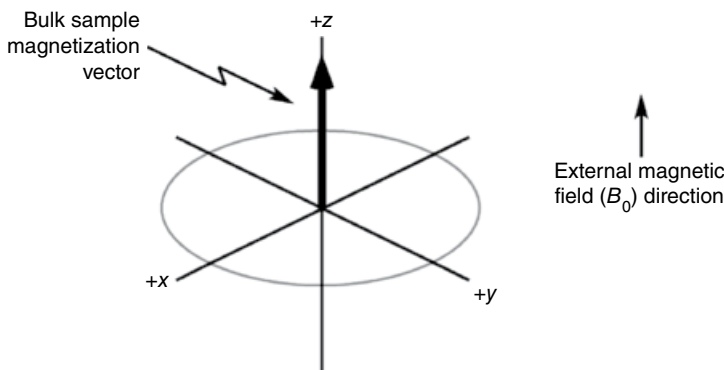


FIGURE 70.3 Representation of an NMR sample's bulk magnetization vector in the "laboratory" frame of reference.

The sample's bulk magnetization vector points in the same direction as the external magnetic field, as shown in Figure 70.3. Two important conventions apply to the Cartesian axis system in the figure. First, the main, external magnetic field vector is always assumed to point in the direction of the $+z$ -axis (which, of course, is why the bulk sample magnetization also develops there). Its orientation along the z -axis is not meant to imply that the magnetic field itself must actually be vertical in the laboratory. Although the superconducting magnets used in commercial high-resolution NMR spectrometers happen to produce vertical B_0 fields, it is perfectly possible to perform NMR with the main field oriented in any physical direction. There is a second important convention: if the Cartesian axis system is time invariant (i.e., stationary), it is defined as the "laboratory reference frame." A useful alternative is discussed later in this chapter.

70.4 EXCITING THE NMR SIGNAL

Once the sample inside the magnet has developed its bulk magnetization vector, the NMR excitation can be performed. Usually it is done by applying a pulse of RF current to a wire coil that surrounds the NMR sample centered in the external B_0 magnetic field. That creates a second external magnetic field, B_1 , along the laboratory reference frame's $+x$ -axis to tip the sample magnetization vector away from the $+z$ -axis. Once it departs the axis, the magnetization vector precesses about the z direction at the Larmor frequency ν_0 just like the individual nuclear spin vectors from which it arises. If the sign of the gyromagnetic ratio γ is positive, the most common case, the precession occurs as shown in Figure 70.4. Negative gyromagnetic ratios cause precession in the opposite direction.

While the bulk sample magnetization vector is tipping away from the $+z$ -axis, it simultaneously precesses as discussed earlier. In other words, the tip of the vector undergoes "nutations," tracing an oscillatory spiral path in the laboratory reference

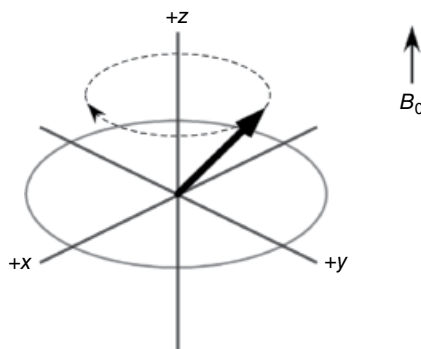


FIGURE 70.4 Precession of the bulk sample magnetization vector about the z -axis in the laboratory reference frame. The direction of precession shown assumes the gyromagnetic ratio has a positive value.

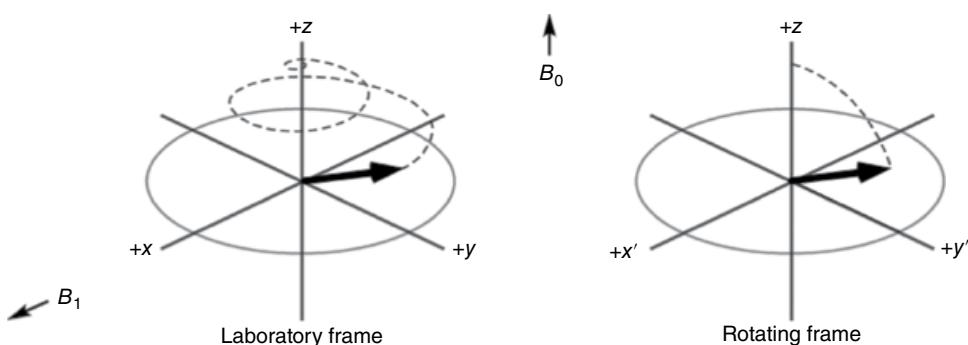


FIGURE 70.5 The motion of a sample's bulk magnetization vector in laboratory and rotating reference frames. If the rotating frame rotates at the precession frequency of the vector, the oscillatory motion disappears. (That condition is called “on resonance.”) Using the rotating frame greatly simplifies discussions of NMR experiments.

frame. This complex motion makes it difficult to visualize many NMR experiments, so it is common to abandon the laboratory reference frame and use a frame that rotates about the z -axis at frequency ν_0 . Conceptually, this is a bit like jumping on a carousel: a motion that appears spiral to an observer standing next to the carousel will lack its rotational component when observed by someone who stands on the carousel. This new frame of reference is called (not surprisingly) the rotating reference frame. The difference between vector motions in the two frames is illustrated in Figure 70.5. The precessional motion stops only if the frame's rotation frequency matches the precession frequency of the vector exactly, a condition known as being “on resonance.” If the sample produces multiple vectors with different frequencies, the rotating frame cannot match all of them simultaneously and, at most, only one will show no oscillation. Note that a prime is often added to x and y to denote the rotating frame. (It can be added to z as well, but there is little point since the z -axes coincide in both frames.)

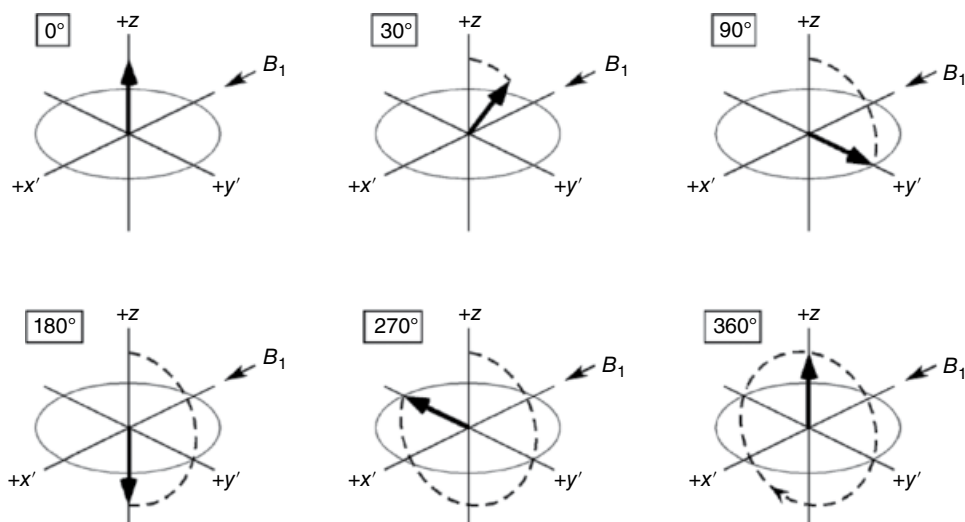


FIGURE 70.6 Nutation (or “tip”) angles produced by various B_1 fields applied along the $+x'$ -axis in the rotating frame.

Since the bulk sample magnetization precesses at frequency ν_0 , an *oscillating* B_1 field at that frequency must be applied along $+x$ in the laboratory frame to tip the vector. In the rotating frame, the field is still applied along $+x'$ but it appears to be *static* and not oscillating. The vector can then be considered to precess about the new $+x'$ magnetic field for the length of time that field remains on. Once the field is turned off, the vector is left somewhere in the $y'z$ plane at an angle θ relative to $+z$. This is referred to as the “nutation angle” or often just the “tip angle.” By analogy with Equation 70.2, the vector’s precessional frequency ν_1 about the B_1 magnetic field is

$$\nu_1 = \frac{\gamma B_1}{2\pi} \text{ (in Hz)}. \quad (70.4)$$

In practice, the B_1 field is normally applied as a pulse of RF energy of some finite duration t_{pulse} , making it easy to calculate the resulting nutation angle:

$$\theta = \nu_1 (t_{\text{pulse}}) (360^\circ) = \gamma B_1 (t_{\text{pulse}}) (360^\circ). \quad (70.5)$$

Longer pulses and greater RF field strengths clearly produce greater tip angles. Examples of nutation are shown in Figure 70.6, from which it is clear that the pulse width or B_1 can be chosen to leave the sample magnetization vector at any desired nutation angle. Commercial NMR spectrometers use B_1 fields that are orders of magnitude smaller than their B_0 fields, so pulse widths of 1–100 μs are usually needed to produce a 90° tip.

NMR spectra usually contain multiple signals distributed over a range of precession frequencies. So, it is vital to be able to tip vectors over that, or preferably a greater,

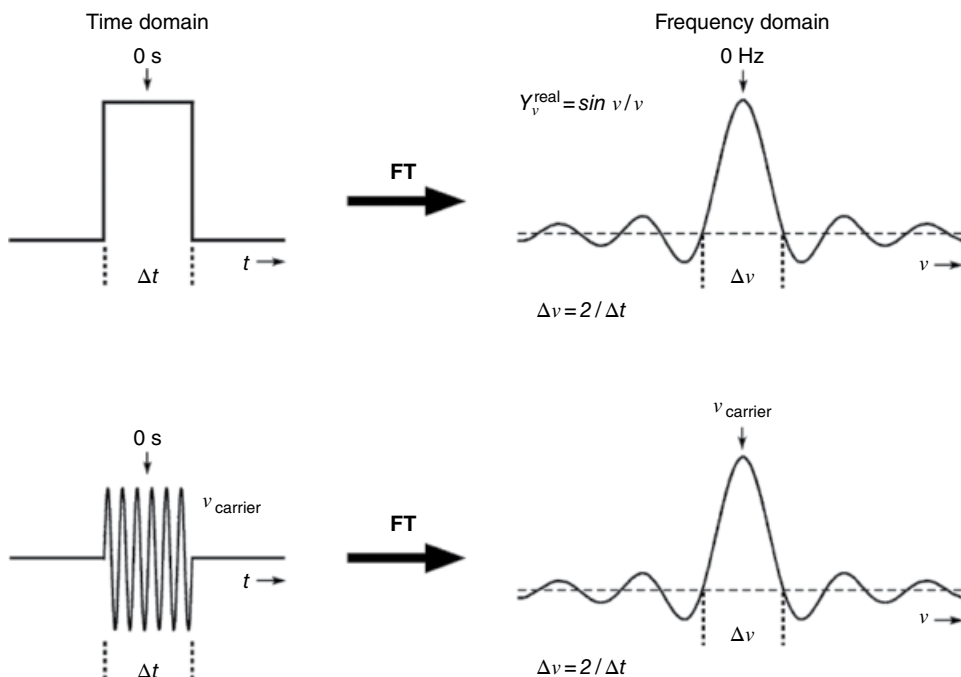


FIGURE 70.7 Fourier pairs showing the relationship between time-domain excitation pulses and the frequencies they excite. The greatest RF power is produced at the “carrier” frequency within the time-domain pulse, but substantial (though not uniform) power is applied at a wide range of other frequencies.

range of frequencies. That is why short RF *pulses* are used for excitation. According to Fourier theory, applying a time-domain pulse excites a range of frequencies simultaneously. As illustrated in Figure 70.7, the Fourier transform of a simple rectangular time-domain pulse waveform produces a frequency-domain $(\sin v)/v$ amplitude function centered at 0 Hz. When the pulse is a rectangular envelope containing a signal oscillating at frequency ν_{carrier} , then the frequency-domain excitation profile is still a $(\sin v)/v$ function, but it is centered at ν_{carrier} instead of at 0 Hz. An inverse relationship exists between the pulse width Δt and the frequency range $\Delta \nu$ that is excited (i.e., $\Delta \nu = 2/\Delta t$). It is generally best to use the shortest practical pulse since that will excite the widest range of frequencies. As an example, suppose a 500 MHz NMR spectrum is expected to exhibit peaks over a frequency range $\Delta \nu$ of 2000 Hz. In that case, the maximum practical RF pulse width is $\Delta t = 2/2000 \text{ Hz} = 1 \text{ ms}$. Provided the 500 MHz carrier frequency is exactly centered in the NMR spectrum, the spectrum will just fit into this 2000 Hz window and all the peaks will be excited. However, from the shape of the $(\sin v)/v$ function in the figure, it is clear that they will not be excited *equally*! If reasonably uniform excitation is required, the entire spectrum should occupy only the middle 5–10% of the excitation bandwidth $\Delta \nu$. In the case of the 2000 Hz spectrum, $\Delta \nu$

must then be $\geq 20,000\text{ Hz}$, necessitating an RF pulse width of $\leq 100\mu\text{s}$ (or preferably $\leq 50\mu\text{s}$). Lest you think that the best way to run all NMR experiments is simply to use extremely short RF pulses all the time, recall that the pulse must also cause a substantial nutation and that large nutation angles result from *long* RF pulses (see Eq. 70.5). Consequently, a compromise pulse width must be employed that (1) is long enough to produce the necessary nutation and (2) short enough to do so more or less uniformly over a suitably wide range of frequencies. To reconcile these competing needs, commercial NMR spectrometers incorporate 50–1000 W RF power amplifiers so that quite short pulses can still produce significant nutations.

In special cases, the inverse relationship between Δt and $\Delta\nu$ can be exploited to excite only a narrow range of frequencies. The most common application of that in NMR is for suppressing intense solvent peaks that would otherwise obscure small analyte signals. A long RF pulse centered at the solvent peak frequency will tend to saturate the peak without perturbing the rest of the spectrum too much. (Saturation results in a weak peak, as discussed in the succeeding text.) An alternative approach is to use long, shaped pulses to selectively excite only the parts of the spectrum one wishes to observe while leaving the interfering solvent peaks mostly unexcited.

70.5 DETECTING THE NMR SIGNAL

Once the sample spins have been excited, the resulting NMR signal must be detected. The same physical coil of wire that is used as the RF transmitter during excitation is also used as the RF receiver during signal detection. Conceptually, however, it is easiest to consider the transmitter and receiver coils to be separate and aligned along the laboratory-frame x and y axes, respectively, as illustrated in Figure 70.8. When the transmitter

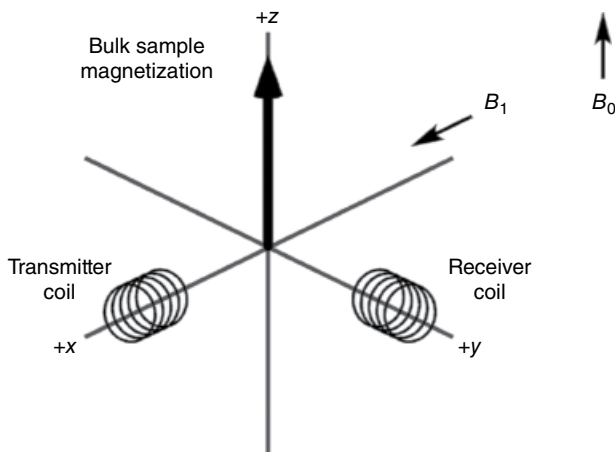


FIGURE 70.8 Conceptual positions of the NMR RF transmitter and receiver coils in the laboratory reference frame.

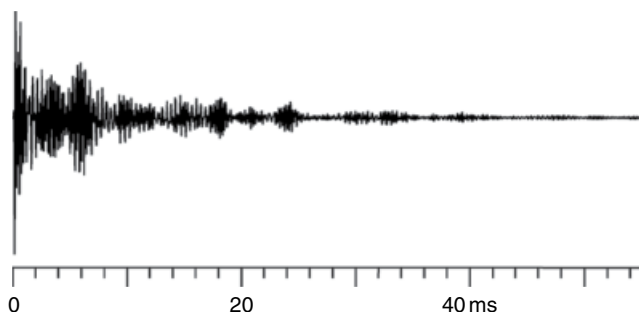


FIGURE 70.9 An experimental ^{13}C NMR free induction decay obtained at 75 MHz from a sample of rat liver in a 7.05 T magnetic field. Nearly all the signals arise from tissue lipids. The oscillatory nature of the FID is evident, as is constructive and destructive interference among the many different frequencies it contains. The progressive decrease in signal intensity is a result of relaxation processes that destroy the xy magnetization.

coil's B_1 field is turned on, the sample magnetization vector nutates continuously until the B_1 field is turned off, at which time the vector is left θ° away from $+z$ (Eq. 70.5). In the laboratory frame, the vector then precesses about the z direction, creating a small magnetic component in the xy plane and a weak oscillating magnetic field in the receiver coil. That oscillating field induces an oscillating current in the coil that is converted to voltage, amplified, and ultimately digitized as a time-domain signal called the “free induction decay” (FID). An example is shown in Figure 70.9. The FID constitutes the fundamental data set in an FTNMR measurement. As shown in the figure, its overall amplitude decreases with time. That results partly from destructive interference among its various component frequencies but is mostly caused by relaxation as the signal's xy components disappear and magnetization is restored along the z -axis.

Since the NMR receiver coil is considered to lie on the $+y$ -axis, it is clear that the most intense FID (and spectrum) will be obtained by applying a 90° excitation pulse to tip the initial z magnetization fully into the xy plane. If multiple FIDs are averaged to form the spectrum, then relaxation effects usually become important and smaller tip angles work best, as discussed later. Figure 70.6 also predicts that a 180° excitation pulse should produce no signal, since the magnetization vector is left aligned along $-z$ with no component in the xy plane. In ideal cases, that is quite correct, and it is often nearly true in practice.

The intensity of the received NMR signal is governed by more than the excitation tip angle. Some nuclides inherently produce stronger signals than others. For equal numbers of nuclei at a constant magnetic field, a nuclide's NMR sensitivity is proportional to $\gamma^3 \cdot I(I+1)$ in the absence of complicating relaxation effects. The signal intensity obtained from a normal, unenriched sample also depends on the natural abundance A of the nuclide. For example, a sample of benzene (C_6H_6) would be expected to produce a stronger NMR signal for ^1H than for ^{13}C simply because nearly all the hydrogen atoms are NMR-active ^1H ($A=99.985\%$) while few of the carbon atoms are ^{13}C

($A=1.108\%$). Scaling the sensitivity by the nuclide's natural abundance yields the NMR "receptivity":

$$\text{receptivity} \propto A \gamma^3 I(I+1), \quad (70.6)$$

an informative parameter that expresses the overall ease of observing the nuclide in an NMR experiment. Since ^1H has the highest receptivity of all the common nuclides, values are usually tabulated relative to it, as in Table 70.1. Owing to relaxation effects, the relative receptivity parameter is useful mainly for comparing nuclides having the same spin quantum number. Among spin-1/2 nuclides, ^1H and ^{19}F are the most observable. It may be surprising to learn that ^{205}Tl is even more receptive than widely studied ^{31}P , resulting in a substantial literature on ^{205}Tl (and ^{203}Tl) NMR despite the rather high toxicity of thallium compounds. With a relative receptivity of only 0.00018, ^{13}C is not an especially attractive NMR nuclide. However, the prevalence of carbon in chemical compounds makes ^{13}C NMR disproportionately important.

70.6 COMPUTING THE NMR SPECTRUM

Once the FID has been digitized and saved in the computer, it is subjected to discrete Fourier transformation (FT) to form the final frequency-domain NMR spectrum. If the spectrometer is "on resonance" (i.e., if its rotating-frame frequency exactly equals the Larmor frequency), then the FID does not oscillate but simply decays with time. For most samples, the on-resonance time-domain decay is monoexponential, and the FID is just the simple function $y=M_0\cdot\exp(-t/\tau)$, where M_0 is the initial amplitude of the signal at time $t=0$. The parameter τ is called the "relaxation time" and $1/\tau$ is the relaxation rate for the process. When such a function is Fourier transformed, a "Lorentzian" frequency-domain line shape is produced, as shown in Figure 70.10. Samples dissolved in liquid solvents normally produce NMR signals that decay monoexponentially, so they produce Lorentzian peaks. FT yields both a real spectrum and a corresponding imaginary spectrum. As case (a) illustrates, the real FT of an on-resonance exponential decay forms a well-behaved, symmetrical "absorption mode" Lorentzian line, while the imaginary FT produces a complicated "dispersion mode" line shape. The latter is difficult to interpret, so only the absorption mode spectrum is usually observed. In practice, the frequency-domain spectrum obtained immediately following FT is rarely either pure absorption or pure dispersion mode. However, it is simple mathematically to process the spectrum to produce the pure modes. That process, called "phasing," can be performed automatically by the computer or interactively under operator control.

What if the FID is acquired "off resonance" (i.e., at a rotating frame frequency that differs from the precession frequency)? That is extremely common and causes the FID to oscillate at the frequency difference while simultaneously decaying exponentially. Figure 70.10 illustrates the effect (case b). Luckily, FT still produces a Lorentzian line,

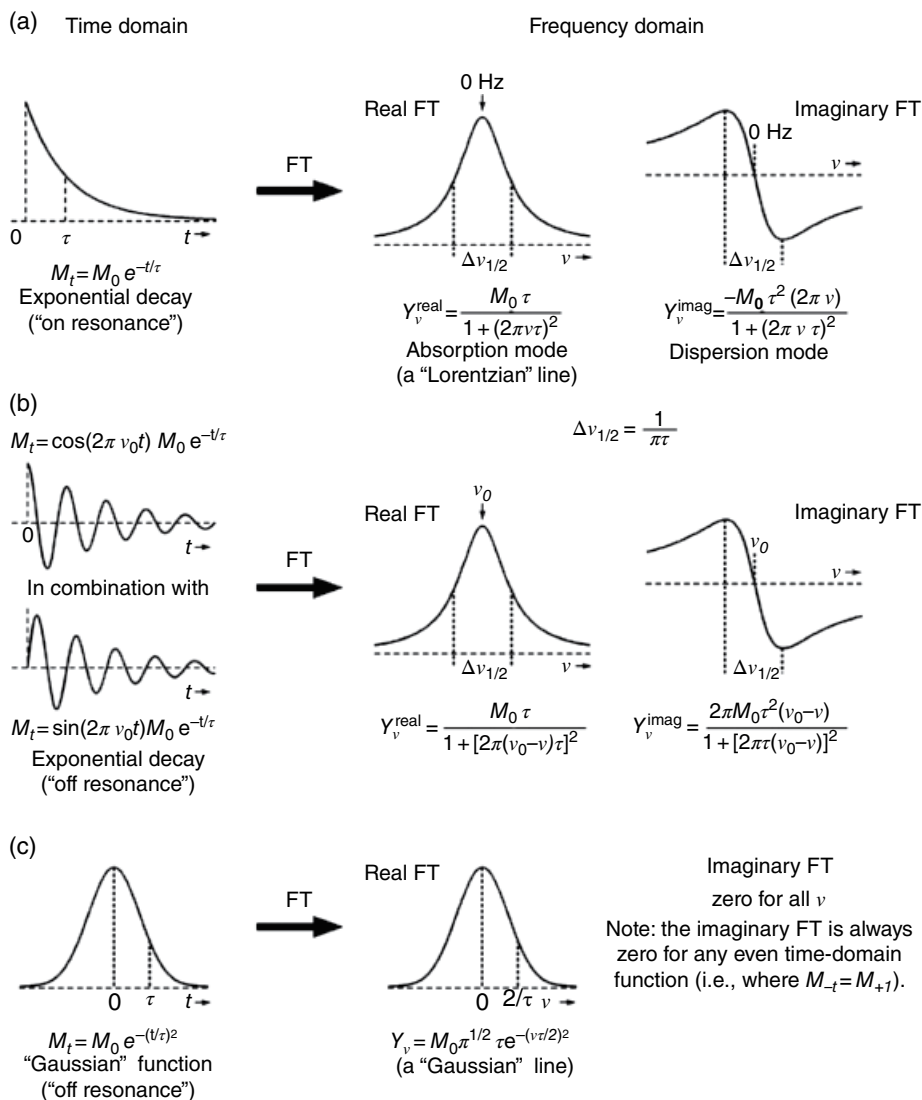


FIGURE 70.10 Fourier pairs relating to NMR line shapes. (a) Fourier transforming a simple monoexponential time-domain decay produces a Lorentzian line having real and imaginary parts as shown. Usually, only the real (absorption mode) part is displayed in spectra. (b) An oscillating monoexponential decay function also produces a Lorentzian line, but one centered at the frequency of the oscillation. (c) Fourier transforming a Gaussian time-domain decay function yields a Gaussian NMR line.

but its frequency is shifted away from zero. The horizontal scale in an NMR spectrum is usually labeled "chemical shift," but in fact it is a frequency scale and the various peaks appear at their frequency differences relative to the spectrometer's rotating frame. Since instrumental factors can shift all the frequencies simultaneously, an

internal standard compound such as tetramethylsilane (TMS) is usually added to the sample, and its peak position is arbitrarily defined as zero on the scale. Other NMR peak positions are defined by their differences relative to that reference frequency.

Unlike solutions, solid NMR samples generally produce FIDs that have maximum amplitude at time 0 and then exhibit Gaussian decay (i.e., as a function of $\exp(-t/\tau)^2$ instead of $\exp(-t/\tau)$). Interestingly, the Fourier transform of a Gaussian function yields another Gaussian function (case (c) in the figure). So, NMR spectra of solid samples typically contain Gaussian, and not Lorentzian, peak shapes.

So far we have mostly assumed the presence of only a single resonance line in the spectrum. NMR would not be very useful if its applications were restricted to samples with single lines! Exciting a multiple-spin system produces an FID that is a sum (i.e., interferogram) of individual oscillatory functions. As shown in Figure 70.9, the appearance can be quite complicated. Fortunately, the FT of a sum of functions is equal to the sum of the FTs of each separate function:

$$\text{FT}\{f_1(t) + f_2(t) + f_3(t) + \dots\} = \text{FT}\{f_1(t)\} + \text{FT}\{f_2(t)\} + \text{FT}\{f_3(t)\} + \dots \quad (70.7)$$

This is called the Fourier addition theorem [9]. Each component of the intimidating FID transforms independently and contributes a separate, easily discernible peak to the frequency-domain spectrum. For ideal liquid samples, all the peaks will have Lorentzian shapes, but that does not mean that all the line *widths* are necessarily the same. As shown by the Fourier pairs in Figure 70.10, the width of a Lorentzian peak is an inverse function of its decay time constant τ . So, if the various spins in the sample produce time-domain signals that decay at different rates, their spectral line widths will vary. That can easily be caused by differing relaxation rates, as discussed later.

70.7 NMR INSTRUMENTATION

Despite the impressive appearance and expense of a typical NMR spectrometer, the hardware design concept is relatively straightforward (Fig. 70.11). It begins with a high-quality RF transmitter (usually a computer-controlled digital synthesizer). The transmitter is “on” continuously. Its signal of, typically, a few tenths of a volt is applied to the input of a computer-controlled RF switch. That device creates the RF pulse by turning on, passing RF energy for the desired number of microseconds, and then turning off again. The resulting small RF pulse then enters an RF power amplifier and emerges with an amplitude of, typically, tens of volts. (Recall that high RF power is needed so the sample’s bulk magnetization can be nutated even when using a very short pulse.) The amplified RF pulse then passes through series crossed diodes that have essentially no effect on it. The diodes are needed because RF power amplifiers tend to produce unwanted noise when no pulse is present, and since the diodes block signals of less than about 0.5 V, they prevent the amplifier noise from contaminating the tiny NMR signal when it is detected later. After passing through the diodes, the RF pulse is applied to the

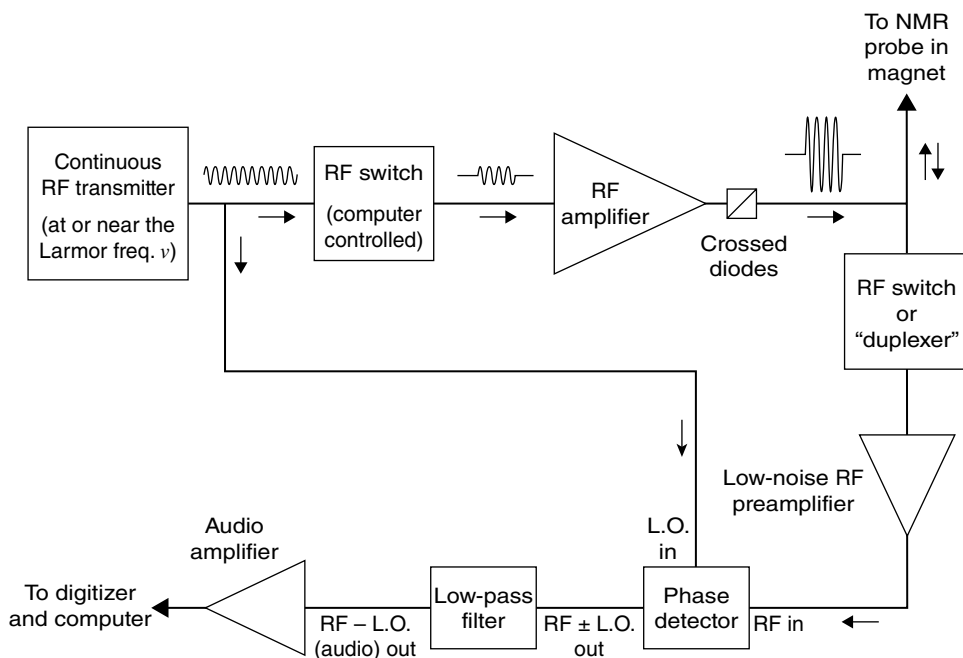


FIGURE 70.11 Block diagram of a pulsed Fourier transform NMR spectrometer operating at the Larmor frequency ν . This conceptual diagram shows the major features but excludes some practical details, such as additional tuned filters at various stages that help to suppress noise.

NMR probe, where it excites the sample. At that point, the excitation part of the experiment has been accomplished and it only remains to detect the sample's resulting NMR signal. That tiny (microvolts) signal returns from the probe and proceeds downward in the diagram to an active RF switch or, equivalently, a passive "duplexer." That device is not fundamental to the principle of the measurement, but it plays a crucial practical role: it is turned "off" during the high-power RF pulse to prevent the pulse from reaching the extremely sensitive RF preamplifier and destroying it. Once the pulse ends, this switch is closed again, providing a path for the sample's NMR signal to reach the low-noise preamplifier. The role of the preamplifier is to amplify the NMR signal voltage, typically by 1000-fold or more, while contributing little additional noise itself. The boosted NMR signal is then applied to the input of a "phase detector." This device is "borrowed" from superheterodyne radio technology. It combines the NMR RF signal ($A_1 \cos(2\pi\nu_1 t)$) with a local oscillator or "L.O." signal ($A_2 \cos(2\pi\nu_2 t)$) that comes directly from the original transmitter. The output of the phase detector can be considered to be the product of the two input signals ($A_1 A_2 \cos(2\pi\nu_1 t) \cdot \cos(2\pi\nu_2 t)$), which, due to a mathematical identity, is equivalent to the sum and difference of their frequencies:

$$\text{output signal} = \left\{ \frac{1}{2} A_1 A_2 \cos[2\pi(\nu_1 + \nu_2)t] \right\} + \left\{ \frac{1}{2} A_1 A_2 \cos[2\pi(\nu_1 - \nu_2)t] \right\}.$$

For example, if the transmitter produces 500.000 MHz and the actual NMR signal has a Larmor frequency of 500.003 MHz, the output of the phase detector will be an interferogram containing signals of 1000.003 MHz (the sum) and 0.003 MHz (the difference). By then passing this through a low-pass audio filter, the high-frequency component is blocked and only the 0.003 MHz (or 3 kHz) signal reaches the remaining amplifier, digitizer, and computer. Why do this instead of digitizing the NMR signal directly? The reason is a practical one. NMR spectra nearly always occupy only a tiny fraction of the whole frequency range. Normal ^1H NMR spectra, for example, are fully contained within a 20 ppm range of frequencies, so in our example, all the spectral information of interest would occupy about a 0.010 MHz range of frequencies near 500 MHz (e.g., between 500.000 and 500.010 MHz). By removing the frequencies below 500 MHz, the phase detector/low-pass filter effectively expands just the small range that contains the NMR information. This low-frequency (audio) signal is easy and cheap to digitize. A possible alternative would be to eliminate the phase detector/low-pass filter and just use an extremely high-speed digitizer operating at over 1000 megasamples per second to characterize all frequencies from 0 to 500 MHz or more. However, since the experiment excites no observable NMR signals below the ^1H range near 500 MHz, no interesting or useful data exist below that frequency and the capabilities of the expensive fast digitizer would be wasted. The phase detector approach is far more practical.

NMR probes contain tuned circuits to sense the precession of the sample's tiny bulk magnetization vector. There is a common misconception that the NMR signal consists of RF photons emitted or "released" by relaxing molecules after they have been excited, in analogy with spectroscopies at higher frequencies such as in the ultraviolet, visible, and infrared regions of the electromagnetic spectrum. By contrast, the NMR signal is detected as an oscillating voltage that the bulk magnetization vector induces in the receiver coil that surrounds the sample. The coil essentially acts as an RF antenna to detect the sample's oscillating magnetic field. Since that field is exceedingly small, the coil is incorporated into a resonance circuit tuned to oscillate at or near the Larmor frequency of the sample (Fig. 70.12). The probe circuit also contains a matching capacitor that matches the extremely high input impedance of the parallel-tuned resonance circuit to the relatively low (e.g., $50\ \Omega$) impedance of the rest of the spectrometer electronics. Impedance mismatches cause RF signals to reflect, much as water waves in a swimming pool reflect from the walls of the pool. Consequently, without the matching capacitor, little RF energy from the power amplifier would enter the resonance circuit and the sample would not be excited. A mismatch would also prevent the small NMR signal from escaping the circuit to reach the preamplifier. The physical shape of the transmitter/receiver coil varies depending on the configuration of the NMR probe. The key is that the B_1 field produced by the coil must be orthogonal to the B_0 field produced by the main magnetic field. For magnets that produce horizontal fields, such as low-field permanent magnets, a vertical solenoid makes an excellent RF coil. For superconducting magnets, which normally produce vertical B_0 fields, a "saddle" RF coil design is normally used since it produces a horizontal B_1 field.

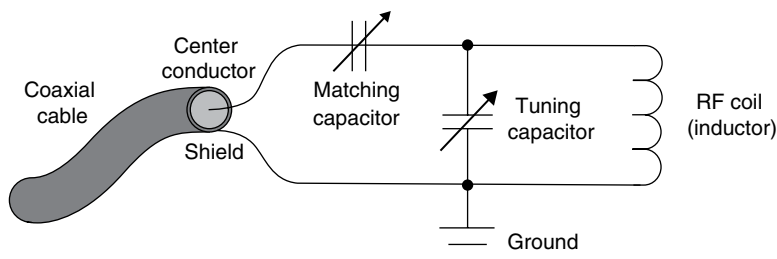


FIGURE 70.12 A parallel-tuned resonance circuit for NMR. The RF coil surrounds the sample and detects the oscillating magnetic field produced by the sample's net magnetization vector. A tuning capacitor connected in parallel with the coil forms a resonance circuit with an oscillation frequency at or near the Larmor frequency of the sample. A separate matching capacitor connected in series matches the input impedance of the parallel-tuned circuit to the, typically, $50\ \Omega$ impedance of other spectrometer components such as the coaxial cable. Without this matching capacitor, the impedance mismatch would prevent RF energy from entering the resonance circuit to excite the sample, and the tiny NMR signal would not escape the circuit to be amplified.

The magnetic field B_0 is crucial to the NMR experiment, and it is important to ensure that field is both as intense and as uniform as practical. It is these requirements that make the magnet the most expensive part of a typical NMR spectrometer. Modern superconducting magnets are designed as shown in Figure 70.13. The actual magnet is only a small part of the overall assembly, being enclosed within a Dewar of liquid helium at 4.55 K, which, in turn, is inside a large outer Dewar of liquid nitrogen at 77.4 K. The magnet must be cooled to liquid helium temperatures since a superconducting wire alloy is used to make the magnet. Unfortunately, liquid helium is expensive and sometimes difficult to get at all, so the helium Dewar is immersed in the liquid nitrogen so that small heat leaks will tend to boil away the relatively inexpensive nitrogen instead of the helium. During initial installation, the magnet is energized but then the power supply is removed and the current continues indefinitely inside the wire. The only time reenergization should be required is in the case of a magnet “quench,” where the wire suddenly becomes resistive and the current drops precipitously, usually boiling off the cryogen in a spectacular, and occasionally dangerous, fashion. A quench can damage a magnet permanently, but designs have improved over the years and they now usually survive quenches successfully.

All magnets are prone to drift slightly. Consequently, NMR samples are usually prepared in deuterated solvents, and a resonance circuit built into the probe continuously monitors the ^2H NMR signal to detect any change in its resonance frequency. Circuitry then compensates for the drift to avoid smearing the NMR signals. This is referred to as “field-frequency” locking.

While locking is generally a robust technique, the deuterated solvent must be chosen based on the analyte's chemical compatibility and solubility. Expense is another consideration, as fully deuterated solvents can easily cost \$1 per gram, especially if

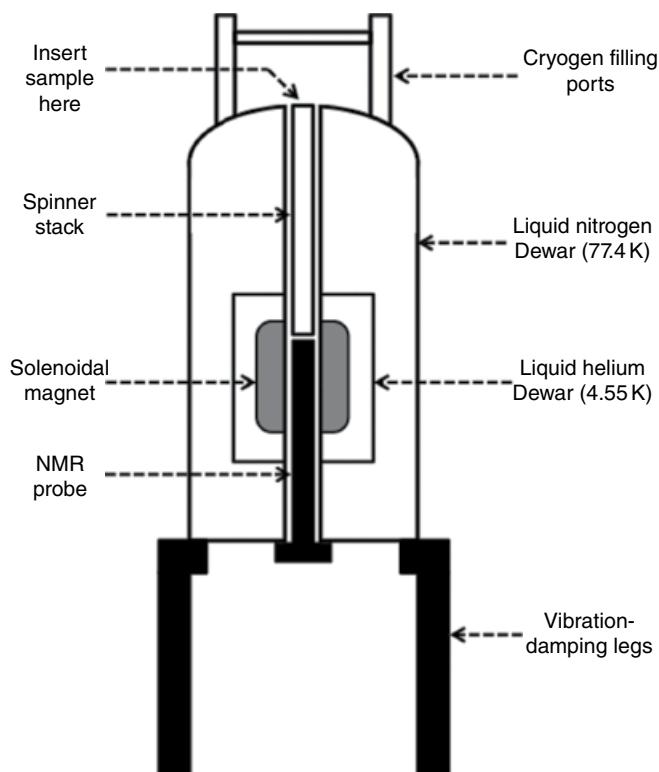


FIGURE 70.13 Cross-sectional diagram of an NMR magnet. Most of the volume is consumed by the liquid nitrogen Dewar. The liquid helium Dewar is contained inside, with the actual superconducting magnet located inside that. Filling ports for the cryogenes are located on top of the magnet. The NMR probe inserts from the bottom, and the sample is placed in the top. The entire assembly is supported by vibration-damping legs to prevent building vibrations from causing side bands in the spectra.

high deuterium enrichment (e.g., 99.9%) is needed. The most common solvents are deuterium oxide (D_2O) for polar analytes, chloroform-d (i.e., $CDCl_3$) for nonpolar analytes, and perdeuterated dimethyl sulfoxide ($DMSO-d_6$), which is a surprisingly good solvent both polar and nonpolar analytes. Since absolutely complete deuteration is not possible, even the best available solvents exhibit small peaks due to residual protonated molecules. Table 70.2 lists several common deuterated solvents, along with the positions and multiplicities of their 1H peaks due to incomplete deuteration. In D_2O , it is unlikely that any given water molecule contains more than one 1H , so its residual protonated water peak is referred to as an “HOD” peak. Table 70.2 also lists the solvents’ ^{13}C chemical shifts and multiplicities. Unlike the typically small residual 1H peaks, the ^{13}C peaks are relatively intense since samples contain mostly solvent and the solvent contains full natural-abundance ^{13}C , just like the dissolved analytes. ^{13}C -depleted solvents can be purchased for special work, although they are quite

TABLE 70.2 NMR Properties of Common Deuterated Solvents^a

Solvent	Formula	Cost	¹ H Chem. Shift	¹ H Multiplicity	¹³ C Chem. Shift	¹³ C Multiplicity
Acetone-d ₆	(CD ₃) ₂ CO	\$\$\$	2.06	5	29.9 206.7	7 1
Benzene-d ₆	C ₆ D ₆	\$\$\$	7.16	1	128.4	3
Chloroform-d	CDCl ₃	\$	7.24	1	77.2	3
Deuterium oxide	D ₂ O	\$\$	4.80	1	—	—
Dimethyl sulfoxide-d ₆	(CD ₃) ₂ SO	\$\$\$	2.50	5	39.5	7
Methylene chloride-d ₂	CD ₂ Cl ₂	\$\$	5.32	3	54.0	5

^aChemical shifts are in ppm relative to TMS or TSP at 0 ppm. Multiplicities indicate the number of peaks present in the solvent's resonance.

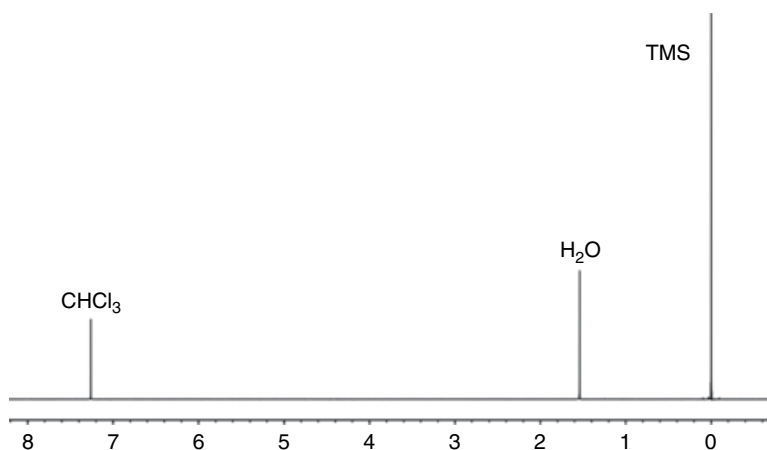


FIGURE 70.14 Proton NMR spectrum of deuteriochloroform (CDCl₃) solvent containing 0.1% TMS. Parameters: 599.7 MHz; 25°C; 45° tip; 1 average; 20 Hz spin rate. Residual protonated chloroform is visible near 7.24 ppm, and a small quantity of contaminating moisture produces the peak near 1.5 ppm.

expensive. One additional problem associated with NMR solvents is the presence of moisture. Even relatively nonpolar solvents like CDCl₃ can contain enough H₂O to produce a significant peak, as illustrated in Figure 70.14. To make matters worse, the chemical shift of the water peak can vary enormously depending on the amount of hydrogen bonding present. As shown in Table 70.2, the residual HOD peak in D₂O appears near 4.8 ppm, but the peak shifts far upfield (1–3 ppm) in solvents that interfere with hydrogen bonding. That can confound structure assignments and may require running a solvent blank to determine exactly which peaks belong to the analyte. Even that can fail, however, since the analyte itself may be the main source of the moisture.

Extraordinarily high-quality NMR magnets are now available commercially. Not only are their magnetic fields intense but, nearly as important, their fields are extremely homogeneous (i.e., uniform). That homogeneity is crucial to achieving the best spectral line shapes and resolution. The reason is easy to understand by making a small modification to the Larmor equation:

$$\Delta\nu_0 = \gamma \Delta B_0, \quad (70.8)$$

where ΔB_0 represents a range of magnetic fields over the sample volume and $\Delta\nu_0$ is the resulting range of resonance frequencies. So, variations in the magnetic field tend to smear the NMR lines over a range of frequencies, producing deleterious broadening. For optimum results, the field should vary by no more than about one part per billion over the sample volume. That would correspond, for example, to field-induced line broadening of 0.6 Hz on a 600 MHz spectrometer. Such superb homogeneity cannot be obtained from the basic magnet alone, but the residual field nonuniformities can be corrected by using “shim coils” built into the inner bore of the magnet. Current is passed through those coils to create additional small fields that “shape” B_0 and make it as uniform as possible. Every sample should be “shimmed” individually before its spectrum is acquired, in part because the sample itself can induce field inhomogeneities, particularly at interfaces such as the upper surface where the sample liquid meets air above it. Older spectrometers required the investigator to adjust the currents through the shim coils manually while monitoring the FID to produce the slowest possible signal decay and, therefore, narrowest lines (see Fig. 70.10). Experienced operators usually achieved a good shim in several minutes. With modern instruments, the shimming process is fully automated, and the software usually produces an excellent shim in only a minute or two without operator intervention. In addition to shimming, sample spinning is used to obtain narrow peaks. Spinning the tube about its long (z) axis tends to average field inhomogeneity in the transverse (x and y) directions. A spinning rate of about 20 Hz is usually sufficient, and the improvement obtained can be dramatic (Fig. 70.15). Small spinning side bands are often be observable at the spinning rate or integer multiples of it, as shown in the figure. These can easily be distinguished from other, more meaningful small peaks because (1) they appear on every peak in the spectrum, (2) their positions change if the spinning rate is varied, and (3) they disappear altogether if the spinning is stopped. To keep the spinning side band amplitudes low, it is important to use both the best practical shim and the highest quality NMR sample tubes. Tubes are marketed by several manufacturers and must conform to rigid specifications for roundness, concentricity, and camber (Fig. 70.16). Low-quality tubes may suffice for undemanding applications at low magnetic fields, for example, in nonspinning studies at 60 MHz, but high-field magnets require quality sample tubes that can be expensive. For most spectrometers, 5 mm outer diameter sample tubes are used, but “wide-bore” instruments allow 10 and even 20 mm tubes to fit. Achieving an excellent shim is often more challenging with large-diameter tubes since the homogeneity must be optimized over a

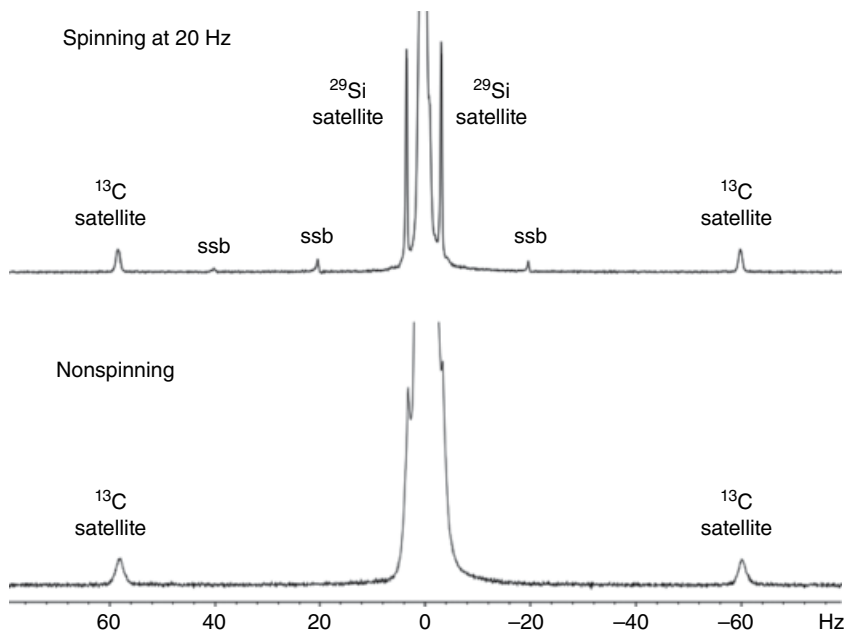


FIGURE 70.15 Expanded proton spectra of 0.1% (v/v) tetramethylsilane (TMS) in CDCl_3 . Acquisition parameters: 599.7 MHz; 25°C; 45° tip; 27.2 s total cycle time; 32 averages. The displayed spectral width is about 0.27 ppm. In addition to satellite peaks due to ^{29}Si - and ^{13}C -containing molecules, the spinning spectrum shows multiple spinning side bands (ssb) that are separated from the central peak by the spinning rate (20 Hz) or integer multiples of it. Turning off the spinning (bottom spectrum) eliminates the spinning side bands but degrades the NMR line widths.

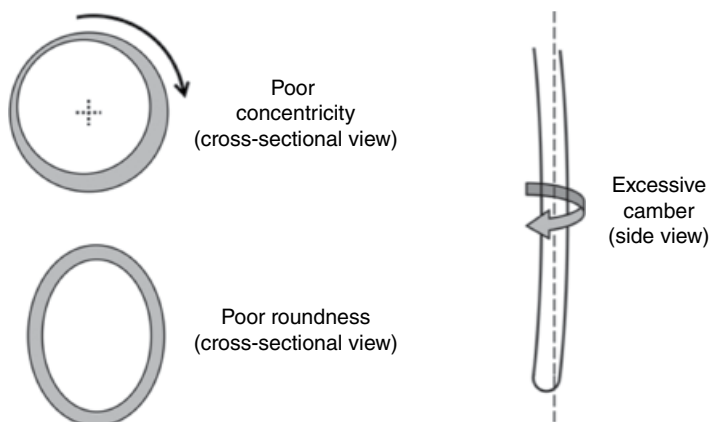


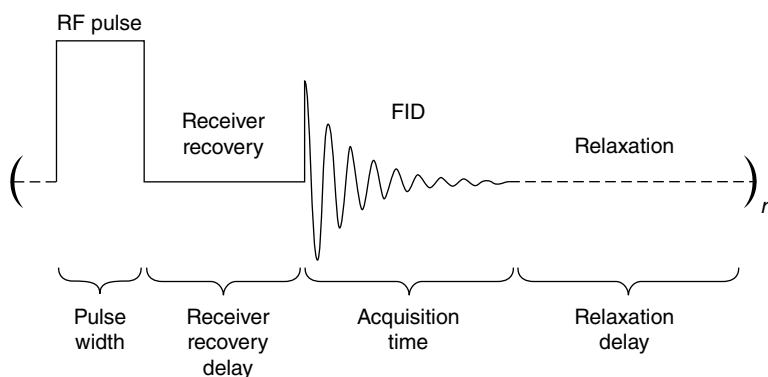
FIGURE 70.16 NMR sample tube characteristics. Concentricity indicates how well the inner and outer surfaces of the tube wall are centered relative to each other. Roundness is a measure of how closely the tube's cross section conforms to a perfect circle. Camber reflects the tube's deviation from straightness. The concentricity, roundness, and camber shown in this figure are all dismal. Spinning a sample inside such a tube would cause a periodic modulation of the NMR signal, resulting in intense spinning side bands that flank every peak in the spectrum. Excellent sample tube quality is important for all NMR applications but is especially crucial for studies in high magnetic fields. The best tubes marketed for use at or above 600 MHz can cost over \$30 each.

greater volume. Although sample spinning and a quality shim are the best ways to obtain good NMR line shapes and resolution, major improvements are also possible after the data have already been collected by using the “reference deconvolution” technique [10, 11]. The essence of the technique is to isolate an intense, well-resolved single peak in the spectrum (often the solvent) and then to deconvolve its shape from the rest of the spectrum. Since B_0 inhomogeneity affects the shapes of all the peaks in the same manner, deconvolving should ideally remove the effects and leave perfectly Lorentzian and narrow lines throughout the spectrum. The method can work remarkably well and, in many cases, without degrading the signal/noise ratio very greatly (or even at all!).

70.8 THE BASIC PULSED FTNMR EXPERIMENT

The great majority of NMR studies are performed on liquids or dissolved samples using a simple “1-pulse” sequence (also known as “pulse and collect”). The sequence is illustrated in Figure 70.17. It begins with the application of a short (microseconds) rectangular pulse of RF energy to the sample, which tips the bulk magnetization vector away from the z -axis. Unfortunately, the RF pulse also saturates the sensitive preamplifier in the spectrometer’s receiver, which responds by emitting wild signal excursions for, typically, 1–20 μ s. To avoid acquiring that meaningless signal, a short receiver recovery delay is inserted immediately following the pulse. Then the computer’s digitizer is turned on and the FID is acquired as a time series of points. Finally, a relaxation delay is used to give the sample time to relax before the next pulse is applied.

The sequence is usually repeated multiple (n) times, and the resulting n FIDs are averaged to improve the signal/noise ratio. Signal averaging is a routine operation in



(The relative time intervals have been altered for clarity)

FIGURE 70.17 The “1-pulse” NMR sequence. The bulk sample magnetization vector is tipped away from the z -axis by a short (microseconds) rectangular pulse of RF energy. Following a brief recovery delay, the FID is acquired. The spin system is then allowed to relax during the relaxation delay. The sequence is repeated n times, recording the n FIDs and averaging them.

NMR since the basic signal is nearly always weak and noisy. When n independent FIDs are summed, the signal adds linearly as $n \cdot S_1$, where S_1 is the amplitude of the signal obtained in a single scan. The noise also adds, but only in proportion to $\sqrt{n} \cdot N_1$, where N_1 is the time-domain noise amplitude in a single scan. Consequently, the ratio of signal/noise improves in proportion to the square root of the number of averaged scans:

$$\left(\frac{S}{N}\right)_n = \frac{n \cdot S_1}{\sqrt{n} \cdot N_1} = \sqrt{n} \cdot \left(\frac{S}{N}\right)_1. \quad (70.9)$$

Averaging 4 FIDs doubles the S/N . Clearly, time averaging can easily reach a point of diminishing returns. For example, if 5 s are needed to acquire a single FID, doubling the S/N requires 20 s, which is usually very acceptable. However, to improve the S/N by 10-fold, 100 scans and 500 s (8.3 min) are needed. It will be necessary to average 10,000 scans to improve the S/N by 100-fold, requiring 50,000 s (833 min)! That would be prohibitive in all but the most dire cases. Alternatives to signal averaging could include isotopic enrichment, increasing the sample concentration, or using a larger-diameter sample tube to place more sample in the NMR coil (if the probe allows). A larger tube increases the S/N , but generally not in direct proportion to the increased sample volume since (1) NMR coils lose efficiency as their diameters increase and (2) the sample itself becomes a more significant source of noise.

Once the averaged FID is obtained, it is Fourier transformed to produce the one-dimensional (1D) NMR spectrum. That spectrum must normally be “phased” to produce pure absorption-mode peaks (the real FT of Fig. 70.10) instead of dispersive peaks (the imaginary FT of Fig. 70.10) or some mixture of the two modes. Phasing is largely automatic on modern spectrometers, although a little manual touch-up is sometimes needed.

Prior to acquiring the NMR spectrum as in the earlier text, it is often necessary to determine the RF pulse width that corresponds to a 90° (or other) tip angle. In that case, a set of spectra is acquired by using a different, incremented pulse width for each. Following FT and phasing, it is possible to determine the amount of nutation caused by each of the pulse widths (Fig. 70.18). It is common practice to search for the pulse width that yields a 360° nutation, since its “null” signal is superior to that produced by a 180° pulse if the sample extends outside the RF coil where the B_1 field is very inhomogeneous. Once the 360° pulse width has been found, dividing it by four yields the 90° pulse width.

70.9 CHARACTERISTICS OF NMR SPECTRA

NMR spectra are often discussed by using special terms, as illustrated in Figure 70.19. The rationale for some of these is purely historical and not rooted in modern practice, but the terms continue to be used anyway. One important modern convention is that the frequency increases from right to left, just as in other types of spectroscopy.

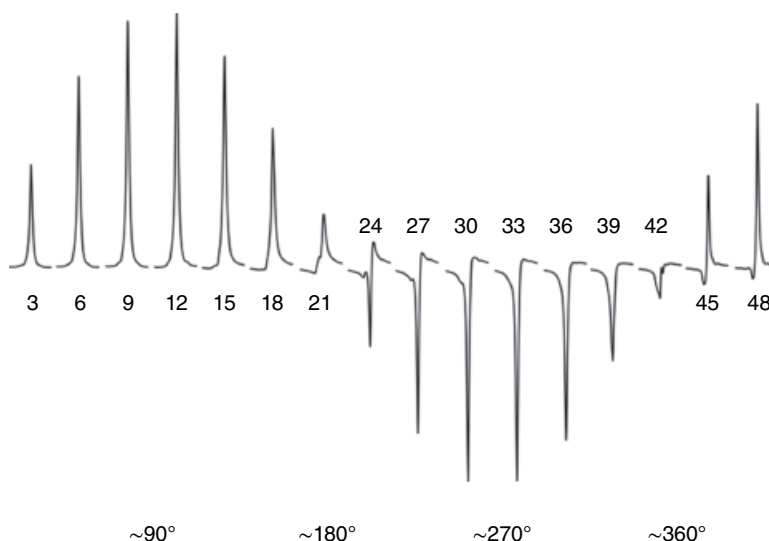


FIGURE 70.18 Proton RF pulse width calibration data for the H_2O peak in a CDCl_3 solvent. Acquisition parameters: 599.7 MHz; 25°C; 42 s total cycle time; 2 averages; 20 Hz spin rate. Each peak is labeled with the pulse width (in μs) used to produce it. Approximate tip angles are indicated along the bottom of the figure. A 43 μs pulse produces close to a 360° tip, meaning the 90° pulse width is about $43/4 \approx 10.8 \mu\text{s}$.

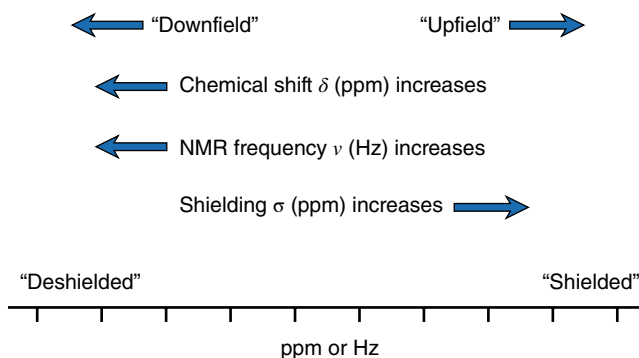


FIGURE 70.19 Terminology conventions in NMR spectroscopy. All common chemical shift references (TMS, TSP, and DSS) produce signals at 0 ppm, significantly upfield from nearly all other ^1H and ^{13}C NMR lines.

70.9.1 The Chemical Shift

The most important parameter in NMR spectroscopy is the chemical shift. The horizontal scale of the spectrum actually represents frequency but is almost universally labeled chemical shift with units of parts per million (ppm). Zero on the scale is assigned based, ideally, on a standard compound actually dissolved in the sample. For ^1H and ^{13}C NMR, three different compounds are widely used as standards (Fig. 70.20).

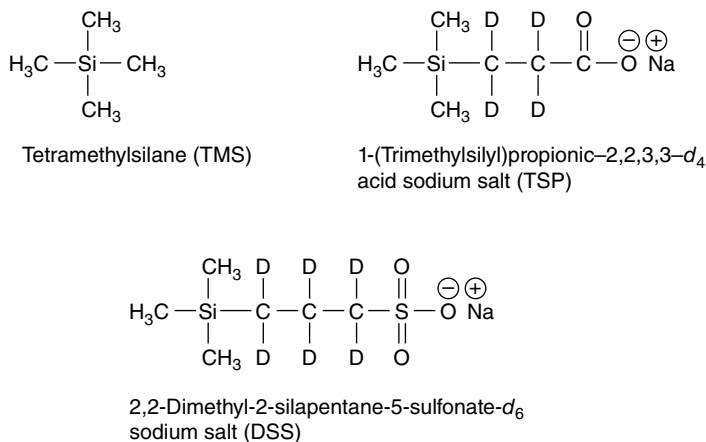


FIGURE 70.20 Chemical structures of the three most common chemical shift reference compounds for ^1H and ^{13}C NMR. TMS is the ultimate reference standard, but its poor solubility in polar solvents requires the use of the ionic salts TSP or DSS instead. All three compounds exhibit a ^1H reference peak at 0.00 ppm.

TMS is the ultimate reference. It is quite nonpolar and dissolves readily in relatively nonpolar NMR solvents such as CDCl_3 and $\text{DMSO}-d_6$. It also has the advantage of a low boiling point (27°C), which facilitates recovering the analyte from the sample without contamination. For polar solvents, low solubility precludes the use of TMS, and ionic salts are employed instead. TSP is the most common, though DSS is also widely used. All three compounds produce ^1H peaks at virtually the same chemical shift and may be assigned values of 0.00 ppm for practical purposes. The pure compounds can be purchased commercially, but deuterated solvents can easily be purchased that already contain 0.1 or 1% of them. When an internal standard cannot be used, due perhaps to a fear of contaminating an important sample, it is possible to fall back on the solvent as a secondary chemical shift reference. Table 70.2 gives the ^1H and ^{13}C chemical shifts of typical solvent peaks relative to TMS, TSP, or DSS. Another alternative is to place the standard inside a glass capillary inserted into the sample, but that is less desirable since (1) the shim is usually degraded and (2) magnetic susceptibility differences between the sample and capillary can easily shift the reference peak by 0.05 ppm or more.

The chemical shift of a nuclide in a molecule can be influenced by several factors (Fig. 70.21). Electrons shield the nuclei from the main B_0 field according to a modified Larmor equation:

$$\nu_0 = \gamma (1 - \sigma) B_0, \quad (70.10)$$

where σ is called the shielding constant. High electron density about a nucleus makes its shielding constant large and that, in turn, produces an upfield shift to low frequency and a small chemical shift value. Any electronegative atom X present in the molecule

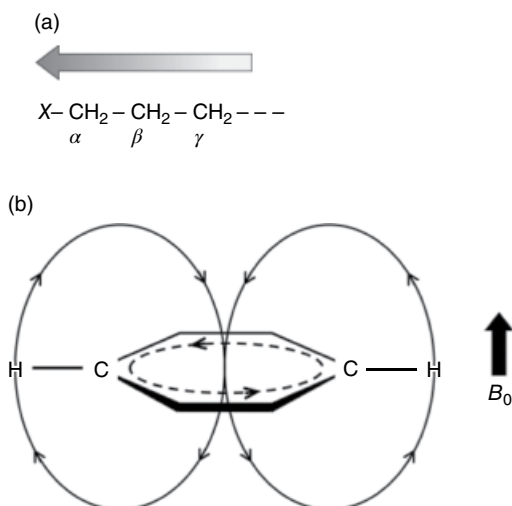


FIGURE 70.21 Some causes of chemical shifts. (a) When relatively electronegative elements X , such as fluorine or oxygen, are covalently bonded in the molecule, they withdraw electron density from nearby atoms. These deshielded atoms produce resonance lines that are shifted downfield (to high frequencies) in the NMR spectrum. (b) Side view of the planar ring of an aromatic compound, such as benzene (C_6H_6), that contains delocalized electrons in π molecular orbitals. In the presence of an external magnetic field B_0 , these electrons circulate (dotted line), creating a small opposing magnetic field in the center of the ring. By tracing the solid lines of flux, it can be seen that this field actually reinforces B_0 at the positions of the molecule's hydrogen atoms in the periphery. To satisfy the Larmor equation, the resonance frequency must increase in response to this slightly higher field, producing a downfield shift for the attached 1H .

tends to withdraw electron density via the inductive effect, leading to downfield (high-frequency) shifts for atoms one or two bonds away. The effect is evident in the chemical shifts of, for example, alcohols *versus* aliphatic compounds shown in Table 70.3. In special cases, more subtle chemical shift effects can occur. For example, an iodine atom, which might be expected to withdraw a little electron density from a molecule, often does just the opposite due to its high polarizability and causes an upfield, instead of downfield, shift. Anionic substances in solution often show relatively large upfield chemical shifts due to shielding by their extra electron(s). For example, the 1H chemical shift of the four equivalent protons of borohydride ion (BH_4^-) coincides almost exactly with that of TMS. Cations can show just the opposite effect. Elegant studies of the distribution of electric charge along hydrocarbon chains have been performed by measuring changes in NMR shielding [12].

Chemical shifts are normally expressed in ppm from the reference line of TMS, TSP, etc. An important ramification is that the separation *in ppm* between two peaks with different chemical shifts should be constant even when measured with spectrometers operating at different magnetic fields. However, the separation *in Hz* will differ. For example, suppose the chemical shift difference between two proton peaks is

TABLE 70.3 Typical Chemical Shift Ranges for ^1H and ^{13}C Relative to TMS

Functional Group	Approximate Chemical Shift (ppm)	
	^1H	^{13}C
CH (aliphatic, alicyclic)	0.0–2.0	0–55
CH (alkyne)	2.0–3.0	60–90
XCCH (α -monosubstituted aliphatic)	2.0–5.0	25–90
X_2CCH (α,α -disubstituted aliphatic)	2.4–7.0	25–90
COH (alcohols, water)	0.5–6.3 (4–6 typ.)	60–80
CNH_2 (amine)	1.7–5.0	25–55
$-\text{CONH}_2$ (amide)	5.0–8.7	150–185
CH (alkene)	4.3–7.5	110–150
CH (aromatic, heteroaromatic)	6.0–9.1	90–160
CH (aldehyde)	8.9–10.2	175–220
COOH (carboxylic acid)	10.0–13.2 (when dimerized)	150–185

2.0 ppm. In a B_0 field of 9.4 T, corresponding to 400 MHz for ^1H , there is 400 Hz per ppm so the separation between the peaks is 2×400 Hz or 800 Hz. In a field of 14.1 T, the proton resonance frequency is 600 MHz and 2 ppm corresponds to 2×600 Hz or 1200 Hz. So, a higher magnetic field puts more “distance” between the peaks, in Hz, even though the separation is still 2 ppm. At higher fields, the spectrum will contain more space for peaks, decreasing the likelihood of peak crowding and overlap. This is one major advantage of using strong magnets in NMR.

It is clear from Table 70.3 that aromatic protons are shifted rather far downfield, even in the absence of any significantly electronegative atoms. For example, the ^1H chemical shift of the six equivalent protons in benzene is about 7.2 ppm, compared to simple aliphatic protons that normally appear in the 0–2 ppm region. A different mechanism is responsible for this effect: ring currents. Aromatic compounds contain loosely bound electrons in π molecular orbitals that are free to circulate in response to the B_0 field. This electron circulation sets up an extra magnetic field that detracts slightly from the main B_0 field in the center of the aromatic ring but reinforces it in peripheral regions outside the ring. Thus, protons attached around the ring experience a slightly increased B_0 field and resonate at a higher frequency. Elegant confirmation of this mechanism is found in the chemical shifts for the aromatic compound [18]-annulene, which contains protons both on the inside and outside of the ring (Fig. 70.22). At -60°C , distinct [18]-annulene proton peaks are observed at -3.0 and $+9.3$ ppm [13]. The -3.0 ppm peak is due to the interior protons that experience the reduced magnetic field, and the peak at $+9.3$ ppm arises from outer protons that are exposed to a stronger field and resonate at a higher frequency. Other large, cyclic aromatic compounds such as porphyrins often exhibit similar effects. The presence or absence of substantial ring current is generally regarded as a reliable test for aromaticity in newly synthesized compounds [14].

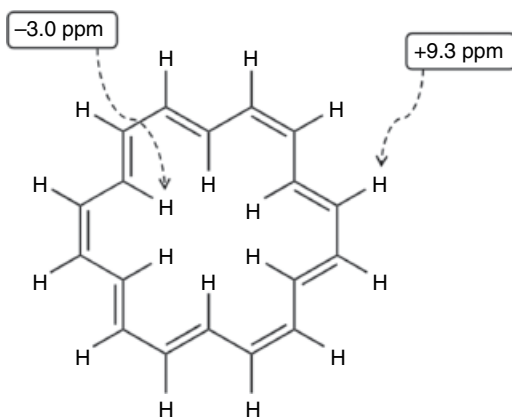


FIGURE 70.22 The chemical structure of [18]-annulene. The six protons in the center of the aromatic ring inhabit a region of decreased magnetic field due to electron ring currents and produce a ^1H peak at -3.0 ppm relative to TMS. The 12 exterior protons experience an enhanced field and produce a peak at $+9.3\text{ ppm}$ near the traditional aromatic chemical shift range of the proton NMR spectrum.

In many alcohols and carboxylic acids, hydrogen bonding can profoundly affect the chemical shifts of $-\text{OH}$ peaks. Strong hydrogen bonding shifts the peaks downfield, while reduced hydrogen bonding results in an upfield shift. This is reflected in the wide range of values shown in Table 70.3 for $-\text{OH}$ species and in the upfield shift for water in CDCl_3 (Fig. 70.14). Compounds that form especially strong hydrogen bonds, such as carboxylic acid intermolecular dimers, exhibit large shifts to low field, as shown by their roughly $10\text{--}13\text{ ppm}$ chemical shift. An even more impressive case occurs in compounds like enols that can form especially stable, intramolecular hydrogen bonds in six-membered rings. One such compound, the enol tautomer of 2,4-pentanedione (Fig. 70.23) exhibits an $-\text{OH}$ peak shifted nearly to $+15.3\text{ ppm}$, beyond the range for any “normal” proton. NMR peaks caused by $-\text{OH}$ protons must be interpreted with care since, depending on the solvent and molecular structure, they can appear at nearly any position in the spectrum.

Table 70.3 includes chemical shift information for ^{13}C along with ^1H . The shift trends for ^{13}C roughly parallel those of ^1H , but the range is over ten times bigger. Since ^{13}C is surrounded by more electrons than is ^1H , there is greater potential for changing the electron density near ^{13}C and that leads to a wider chemical shift range. The effect is even bigger for heavier nuclides. For example, ^{59}Co (atomic number 27) has a known chemical shift range of over $18,000\text{ ppm}$! Its shift is so sensitive that separate peaks appear for the (+) and (–) enantiomers of tris(ethylenediamine)cobalt(III) ion in solution when they are merely ion paired with optically active tartaric acid [15]. Nearly every element in the periodic table has at least one NMR-active nuclide, and virtually all of those have been studied at one time or another. Extremely useful chemical shift compilations have been published for the less common NMR nuclides [16, 17].

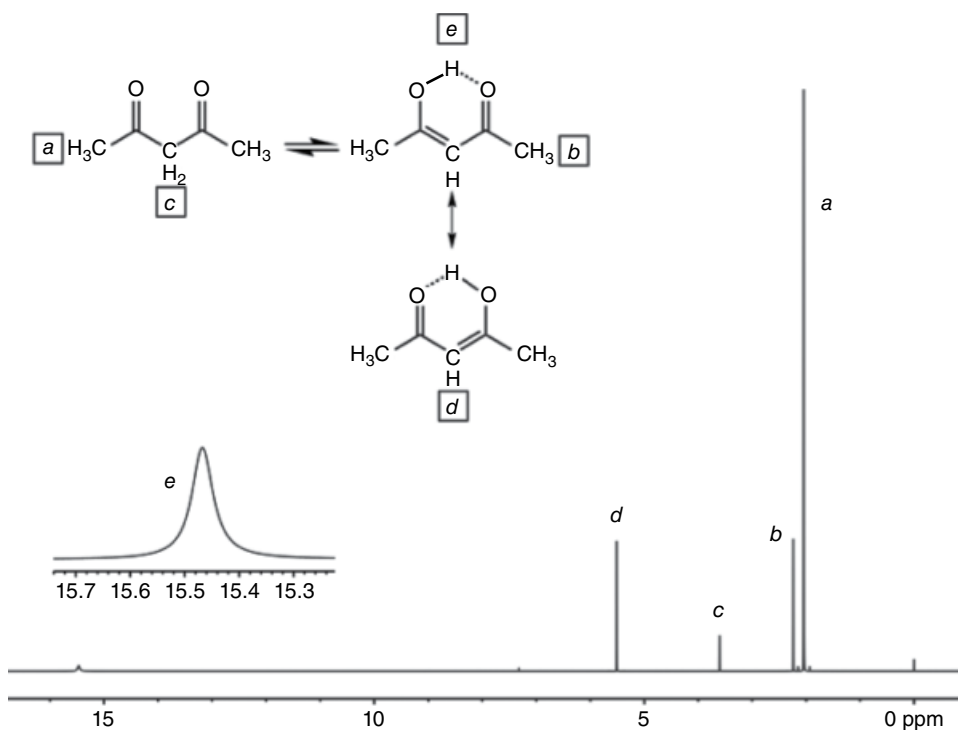


FIGURE 70.23 Proton NMR spectrum of 10% (v/v) 2,4-pentanedione in CDCl_3 containing 0.1% TMS. Acquisition parameters: 599.7 MHz; 25°C; 30° tip; 30.0 s total cycle time; 8 averages; nonspinning. Tautomerization yields an equilibrium mixture of the keto form with a lesser quantity of the enol. Exchange between the two forms is slow on the “NMR time scale,” allowing distinct resonance peaks to be observed for each (assignments shown). Resonance stabilization of the 6-member ring in the enol form produces an unusually strong hydrogen bond and a corresponding large downfield shift for the OH proton resonance. The tiny peak near 7.24 ppm arises from CHCl_3 contamination in the solvent.

Many references provide chemical shift data for standard nuclides like ^1H and ^{13}C . Early compilations [18, 19] remain extremely useful, but more recent monographs contain updated information [20–22]. In addition, the extensive $^1\text{H}/^{13}\text{C}$ spectral libraries published by Aldrich [23] can be quite helpful. Even if the specific compound of interest is not included, the library will often provide spectra for related compounds that can assist in making the assignments.

70.9.2 Spin–Spin Coupling

Spin coupling between various protons, carbons, and other NMR-active nuclides can be exploited for assigning spectral lines to individual nuclei and for determining molecular structures by NMR. Coupling in molecules occurs through chemical bonds via the Fermi contact mechanism. The concept is easy to grasp. When a nuclear spin 1

in a molecule adopts its quantized orientation relative to B_0 , its tiny magnetic moment polarizes the adjacent bonding electrons. Those, in turn, set up a weak magnetic field in the vicinity of another nearby nucleus 2, causing its resonance frequency to shift slightly. In a different molecule, spin 1 may adopt a different quantized orientation, which changes the shielding on nucleus 2. The overall effect in large ensembles of spins is to split the single resonance line of nucleus 2 into multiple peaks that reflect the possible orientations of nuclear spin 1. That information is conveyed by the electrons in the chemical bonds and is called scalar coupling. The effect is most likely to be strong when the two nuclei are in close proximity and tends to grow weaker with the number of intervening chemical bonds.

The magnitude of the scalar coupling is indicated by the spacing, in Hz, between the peaks in the coupled multiplets, and is designated by the letter J . The strength of scalar coupling is controlled by the bonding details of molecules and *has no dependence on the external B_0 field*. Consequently, J values are always expressed in Hz and never in ppm. If measured using a high-field NMR system, a J value will be identical to that found by using lower-field instruments.

In the absence of complicating effects, if two nuclei are coupled such that one exhibits a doublet of peaks, the other will also be a doublet. The two peaks in both doublets will be separated by J Hz, so it is often easy to determine which nuclei are coupled simply by inspecting the spectrum. Coupling between two directly bonded nuclides (such as C—H) is called 1-bond coupling and is symbolized by $^1J_{\text{CH}}$, where the superscript indicates the number of chemical bonds that separate the coupled nuclei. Coupling between two vicinal protons bonded as H—C—C—H would be $^3J_{\text{HH}}$ since there are three intervening bonds between the two protons. As a general trend, $^1J > ^2J > ^3J$, and so on since the Fermi contact mechanism loses efficiency with the successive addition of more bonds.

When spin-1/2 nuclei couple, the appearance of the resulting coupling patterns can be predicted nicely by Pascal's triangle in most cases (Fig. 70.24.) For example, the ^1H NMR spectrum of TMS (Fig. 70.15) displays doublets due to spin coupling of the observed methyl protons with directly bonded ^{13}C . Since only 1.1% of all carbon is ^{13}C , the doublet appears as the weak “ ^{13}C satellites” indicated in the figure. $^1J_{\text{HC}}$ in this case is 118.0 Hz. Nearer to the center of the TMS ^1H line, two clear peaks are visible due to ^1H – ^{29}Si coupling. The coupling constant is comparatively small ($^2J_{\text{HSi}} = 6.57$ Hz) since it results from 2-bond coupling. A careful examination of the spectrum also reveals that neither the ^{29}Si nor the ^{13}C doublet is centered exactly on the main peak. That is because the ^1H chemical shifts change slightly when ^{13}C is substituted for ^{12}C and ^{29}Si is substituted for ^{28}Si .

The simple coupling patterns predicted in Figure 70.24 are ideal approximations. In practice, spectra conform well to them provided $\Delta\nu \gg J$, where $\Delta\nu$ is the difference in Hz between the chemical shifts of the coupled partners. When the chemical shift difference between two coupled nuclei begins to approach the coupling constant, so-called second-order effects usually occur, distorting the expected peak intensities and spacings and even

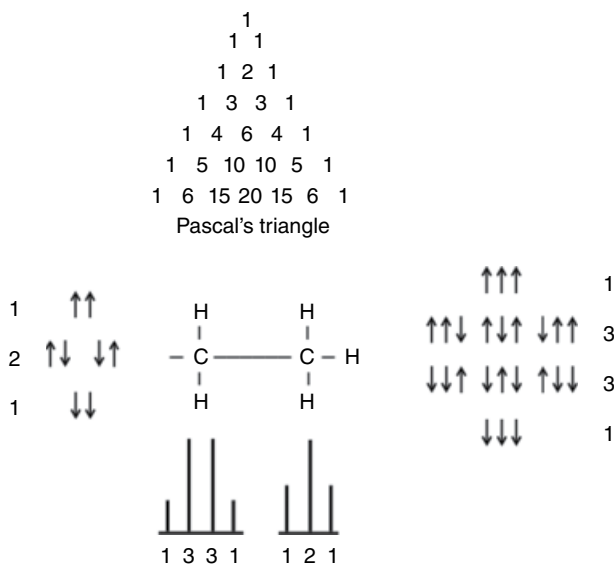


FIGURE 70.24 Predicting the number and intensities of lines in NMR multiplets due to spin–spin coupling between spin-1/2 nuclides. The CH_2 group is split by the three methyl H's, which can take on four different overall spin energies. The middle two energies are three times more likely. So, the CH_2 protons are split into four peaks with intensity ratios of 1:3:3:1. This corresponds to the fourth row of Pascal's triangle. A similar argument accounts for the triplet of 1:2:1 intensity ratios for the CH_3 protons.

forming extra peaks that are not predicted by first-order theory. An example of small second-order coupling effects is in Figure 70.25, which shows spectra of ethyl *trans*-crotonate acquired at 60 and 600 MHz. The inset in (a) expands the spectrum of vinyl proton e, which is a beautiful, though slightly overlapped, doublet of quartets due to its spin coupling with the three protons of the nearby methyl group and the one other vinyl proton. The spectrum exhibits good symmetry and nearly ideal relative peak amplitudes. By contrast, the downfield quartet in the corresponding 60 MHz spectrum (b) shows markedly reduced intensity due to second-order coupling effects since the chemical shift difference, in Hz, is more similar to J at 60 MHz than it is at 600 MHz. The doublet of quartets at 5.8 ppm that corresponds to proton d also shows intensity distortions for the same reason. The spectra of Figure 70.25 remind us of one more lesson: using a high-field NMR spectrometer provides more “empty” space between the multiplets in a spectrum, making it possible to study more complicated molecules without risking significant peak overlap. Obviously that could be a bigger problem in the 60 MHz spectrum of spectrum (b), despite the fact that all the coupling constants in Hz are the same in both spectra.

Spin coupling patterns are a bit more complicated when the coupled nuclides are not spin 1/2. The general rule for predicting the number of lines (assuming first-order coupling) is

$$\text{number of lines} = 2nI + 1, \quad (70.11)$$

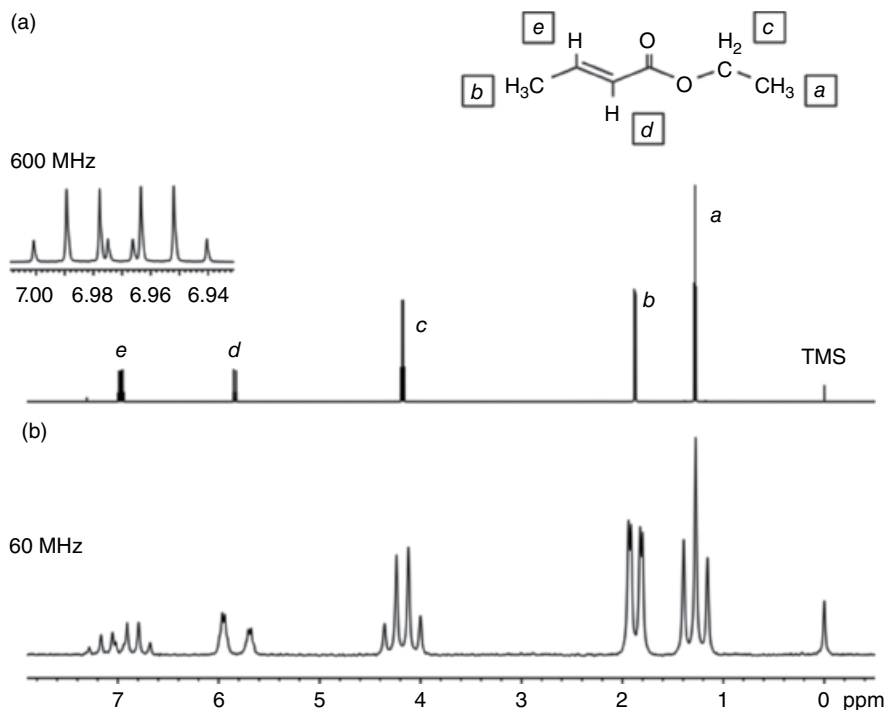


FIGURE 70.25 Proton NMR spectra of 10% (v/v) ethyl *trans*-crotonate in CDCl_3 with TMS. Similar acquisition parameters were used for both spectra except that spectrum (a) was measured at 600 MHz while spectrum (b) was acquired at 60 MHz. The inset shows an expansion of the 600 MHz spectrum for the doublet of quartets assigned to proton *e* in the structure. Note the spacing between peaks within each spin-coupled multiplet is the same in Hz, regardless of magnetic field strength, but it differs greatly in ppm.

where n is the number of nuclei to which the observed nuclide is coupled and I is the spin quantum number of those nuclei. As a simple example, consider spin-1 nuclides like deuterium. They can take on three different quantized spin states: +1, 0, and -1, all with equal probabilities. Consequently, the ^{13}C NMR peak for CDCl_3 is a 1 : 1 : 1 triplet due to coupling of the ^{13}C with the attached deuterium atom. According to the equation, $(2 \cdot 1 \cdot 1) + 1 = 3$ peaks are expected for the ^{13}C . More complicated coupling can give rise to beautiful, elegant spectra in many cases. As illustrated in Figure 70.26, the ^1H spectrum of borohydride ion (BH_4^-) exhibits two distinct sets of peaks, a prominent set of four accompanied by a less intense set of seven. Boron exists as approximately 80% ^{11}B and 20% ^{10}B . ^{11}B is spin 3/2, so according to Equation 70.11, $(2 \cdot 1 \cdot 3/2) + 1 = 4$ proton lines should result from the spin coupling. ^{10}B is spin 3, so the equation predicts $(2 \cdot 1 \cdot 3) + 1 = 7$ lines, precisely as found. Careful measurements of the peak areas confirm the four larger peaks combine to produce the expected 80/20 ratio relative to the sum of the seven smaller peaks. The inset at the top of Figure 70.26 shows the ^{11}B NMR spectrum of the same sample, and it contains the 1 : 4 : 6 : 4 : 1 quintet predicted

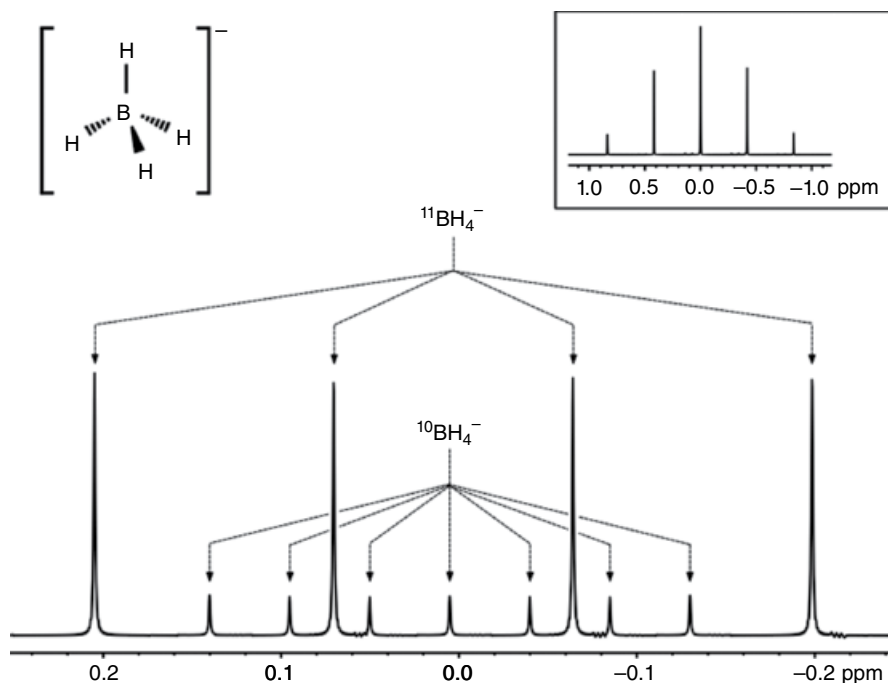


FIGURE 70.26 Proton NMR spectrum of 10% (w/v) sodium borohydride (NaBH_4) dissolved in D_2O . ^1H acquisition parameters: 599.7 MHz; 25°C; 30° tip; 26.7 s total cycle time; 4 averages. Chemical shifts are relative to the HOD peak at 4.80 ppm. The borohydride ion is tetrahedral, as shown. Boron exists as 19.58% ^{10}B ($I=3$) and 80.42% ^{11}B ($I=3/2$). Consequently, ^{10}B -containing borohydride splits the proton spectrum into seven lines, and ^{11}B -borohydride exhibits four stronger lines. The boxed inset shows the ^{11}B NMR spectrum of the same sample, with its central line set arbitrarily to 0.0 ppm. The spectrum is a 1:4:6:4:1 quintet due to boron coupling with the four equivalent protons, with $^1J_{\text{H-}^{11}\text{B}} = 80.62$ Hz. ^{11}B acquisition parameters: 192.4 MHz; 25°C; 30° tip; 12.4 s total cycle time; 4 averages.

by Pascal's triangle for coupling to the four spin-1/2 protons in the ion. The coupling is completely first order since the chemical shift difference between ^1H and ^{11}B is gigantic (407.3 MHz) compared to the $^1J_{\text{HB}}$ value of 80.62 Hz. The ^{10}B NMR spectrum is not shown, but it resembles the ^{11}B spectrum except that the signal/noise ratio is reduced (due to the lower natural abundance and gyromagnetic ratio of ^{10}B), and the splitting between the five lines is only 27.00 Hz, equivalent to the splitting for the ^{10}B species in the ^1H spectrum. Careful measurements reveal that the center of the $^{11}\text{BH}_4^-$ quartet is at +0.00325 ppm, while it is at +0.00500 ppm for the $^{10}\text{BH}_4^-$ septet. This same phenomenon was encountered before in Figure 70.15 and results because the chemical shift changes slightly when one isotope is substituted for another.

Spin-spin coupling can be a useful tool for understanding spectra and determining chemical structures. However, it can also be an impediment in some cases. Early in the history of NMR, spin “decoupling” techniques were developed to eliminate spin

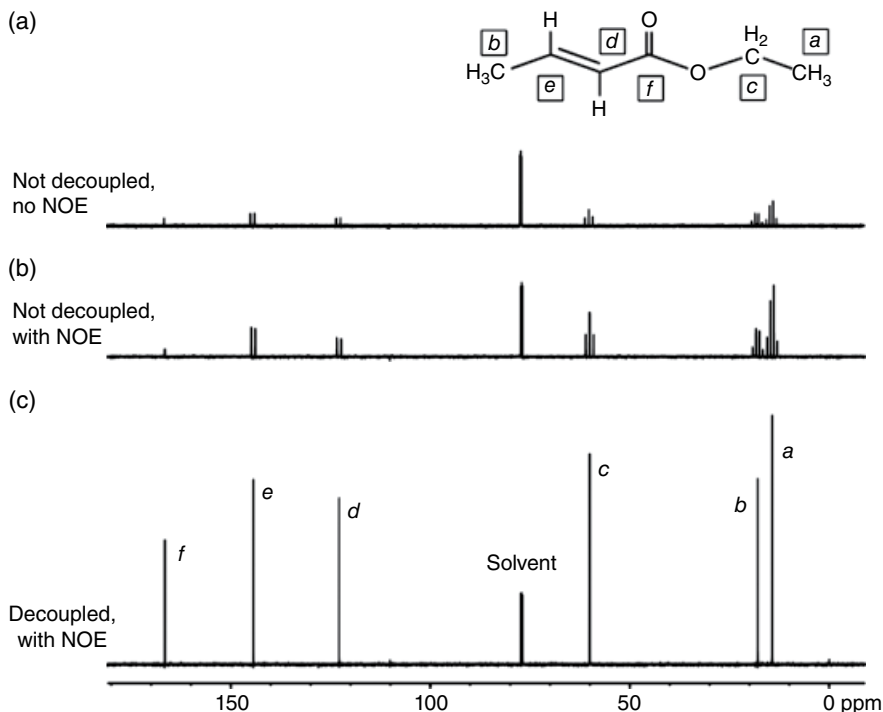


FIGURE 70.27 ^{13}C NMR spectra of 10% (v/v) ethyl *trans*-crotonate in CDCl_3 with 0.1% TMS. Acquisition parameters for all spectra: 150.8 MHz; 25°C; 45° tip; 0.87 s acquisition time and 2.0 s relaxation delay (giving 2.87 s total cycle time); 32 averages; 20 Hz spin rate. (a) With the proton decoupler turned off; (b) with the proton decoupler turned off during FID acquisition but on during the relaxation delay; (c) with the proton decoupler turned on continuously throughout all periods in the experiment.

coupling, either selectively for certain multiplets or throughout the entire spectrum in the case of broadband decoupling. A nice, routine application for broadband decoupling is in the acquisition of ^{13}C NMR spectra. In carbon-containing compounds, the carbon atoms are usually bonded to hydrogens. So, ^{13}C NMR spectra will exhibit spin coupling with ^1H . That can be a very useful peak assignment tool since it indicates whether the carbon is part of a methyl, methylene, etc., group. However, the signal/noise ratio in ^{13}C spectra is often poor and can be improved by decoupling the protons, thus collapsing the multiplets and concentrating all the available signals into a single peak. In addition, if the decoupling ^1H RF field is left on during a substantial part of the experiment, an extra ^{13}C signal/noise improvement of up to almost threefold can be gained due to the so-called nuclear Overhauser enhancement (NOE). Decoupling and NOE effects are illustrated in Figure 70.27. The ^{13}C spectrum (a) was acquired with the proton decoupler turned completely off, and it exhibits full coupling with the ^1H . From the multiplicities, each carbon type can be determined clearly. For example, *a* and *b* are both quartets and must be methyl carbons, while *f* appears to be a singlet and must not

contain an attached proton. In spectrum (b), the decoupler was turned on during the relaxation delay of the experiment but was turned off just before the FID was acquired so as not to interfere with the coupling. The resulting ^{13}C spectrum still exhibits multiplets, but, due to the NOE, the peaks from carbons bearing attached protons are substantially more intense than in spectrum (a). (The solvent peak did not grow since it contains ^2H instead of ^1H .) In spectrum (c), the proton decoupler was turned on continuously throughout the entire experiment to gain the NOE and to collapse the multiplets into singlets. The improvement in signal/noise is obvious, and it is now trivial to count the total number of carbon atoms in each molecule simply by counting the number of peaks in the ^{13}C spectrum. Because the NOE preferentially enhances the signals of carbons bearing attached protons, it distorts the relative peak areas in the ^{13}C spectrum. Consequently, there is usually little to be gained by integrating the ^{13}C peaks in the (forlorn) hope of determining the relative numbers of them in the molecule.

Proton decoupling in a ^{13}C spectrum is an example of *heteronuclear* decoupling. RF pulses are applied at the proton resonance frequency at the same time the spectrometer's receiver is acquiring the ^{13}C FID. Since the resonance frequencies for the two nuclides typically differ by hundreds of MHz, it is easy to build electronic circuitry that blocks even high-power ^1H RF pulses and prevents them from "bleeding" into the ^{13}C receiver channel. Now suppose one wishes to acquire a proton NMR spectrum where all the protons are fully decoupled from each other. That would be an example of *homonuclear* broadband decoupling. Such a spectrum would contain only a single peak at each chemical shift without the complexities and peak overlap that the spin-coupled multiplets often produce. The practical problem is that RF pulses applied in the same frequency range as the one being detected will interfere grossly with the detector electronics and produce a useless spectrum. So, broadband homonuclear decoupling has been a long-standing technological challenge in NMR, and attempts to solve the problem have only met with limited success. Fortunately, a very recent approach now promises much better results [24].

70.10 NMR RELAXATION EFFECTS

Discussions so far have emphasized exciting NMR spin systems and acquiring and analyzing the resulting spectra. Little has been said about how the excited sample relaxes back to its original state, yet that can profoundly affect the spectrum and NMR spectroscopists ignore it at their peril.

70.10.1 Spin-Lattice Relaxation

First consider the microscopic details of how bulk $+z$ magnetization first develops in a spin-1/2 sample. At the instant the sample is inserted into the magnet, the individual nuclear spin vectors align either parallel or antiparallel to the direction of B_0 . A given

nucleus is equally likely to adopt either orientation, so equal numbers of nuclei end up in each state and their individual vectors cancel. Consequently, the bulk sample is initially unmagnetized. As the sample “soaks” for a time inside the B_0 field, a small population of antiparallel nuclei flip to the lower-energy, parallel orientation, producing the tiny bulk sample magnetization along $+z$. These flips to the low-energy state release heat, an exothermic process that increases (very slightly) the kinetic energy of molecules inside the sample. Magnetizing an NMR sample is therefore an *enthalpic* process.

The growth of sample magnetization along the z -axis is characterized by the sample’s “spin–lattice” (or “longitudinal”) relaxation time, universally designated by the symbol T_1 . When that magnetization is perturbed in an NMR experiment, the T_1 value determines how long it takes the sample to reestablish its magnetization along the $+z$ -axis. Following an RF pulse, the z component of the sample magnetization increases exponentially according to the equation:

$$M_\tau = M_\infty \cdot (1 - A \cdot e^{-\tau/T_1}), \quad (70.12)$$

where M_τ is the amplitude of the z magnetization at time τ after the pulse, M_∞ is the fully relaxed z magnetization amplitude (i.e., at $\tau = \infty$), T_1 is the spin–lattice relaxation time, and A is a constant that depends on the initial nutation angle. If a 90° RF pulse is applied, $A = 1$ so that M_τ varies from 0 to $+M_\infty$. Applying an initial 180° pulse causes M_0 to be $-M_\infty$, making $A = 2$ in the equation. The time course for the recovery of magnetization along the z -axis in these two cases is illustrated in Figure 70.28.

T_1 turns out to be a very important practical parameter. Most NMR spectra are measured by repeating an RF pulse sequence many times and averaging the resulting signals.

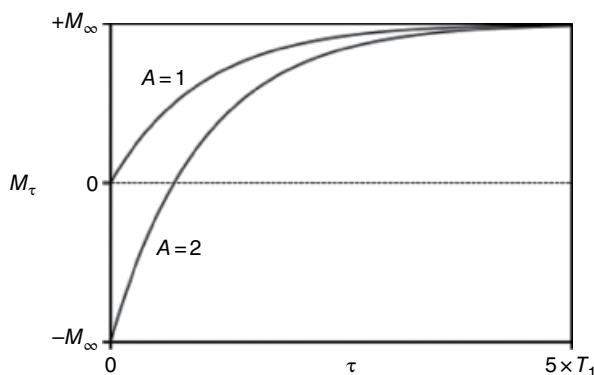


FIGURE 70.28 Recovery of z magnetization following an RF pulse. A perfect 90° pulse makes the z magnetization zero at $\tau = 0$, and it returns to its initial value according to the $A = 1$ curve. (That curve also illustrates how bulk sample magnetization forms when the sample is first inserted into the NMR magnet.) A 180° pulse inverts the initial magnetization, which recovers as shown by the $A = 2$ curve. In both cases, z magnetization is almost completely reestablished after a time period of $5 \times T_1$, where T_1 is the spin–lattice relaxation time.

If the sample's $+z$ magnetization does not recover fully between pulses, then the magnetization becomes partly “saturated” and only a fraction of the possible NMR signal will be obtained. Thus, T_1 controls how long one must wait before repeating pulses. Samples having long T_1 values can force the investigator to use long delays, leading to inefficient data collection and poor signal-to-noise ratios in the spectra. To make matters worse, the various types of nuclei in the sample may relax at different rates, causing some of them to saturate more than others. That distorts the relative peak areas in the NMR spectrum, leading to errors in determining the relative numbers of, say, methyl *versus* methylene protons and in quantifying analyte concentrations. Quantitative applications of NMR require an acute appreciation of the sample's T_1 values!

70.10.2 Spin–Spin Relaxation

After an RF pulse nutates a sample's bulk magnetization vector away from the $+z$ -axis, the vector has a magnetic component in the xy plane that is detected as the FID. The amplitude of the FID is known to decrease with time (Fig. 70.9), corresponding to a decrease in the xy component of the magnetization vector. This time-dependent loss of xy magnetization is caused by “spin–spin” (or “transverse”) relaxation and, in most samples, is represented well by the equation:

$$M_\tau = M_0 \cdot e^{-\tau/T_2}, \quad (70.13)$$

where M_0 is the magnitude of the xy magnetization component immediately following the pulse (i.e., at $\tau=0$). T_2 is the time constant for the exponential decay and is called the spin–spin or transverse relaxation time. The time course for the loss of xy magnetization is illustrated in Figure 70.29. Since magnetization in the xy plane is the result of partial phasing of individual precessing nuclei in the sample, spin–spin relaxation occurs when those phases become more random. It is complete when full

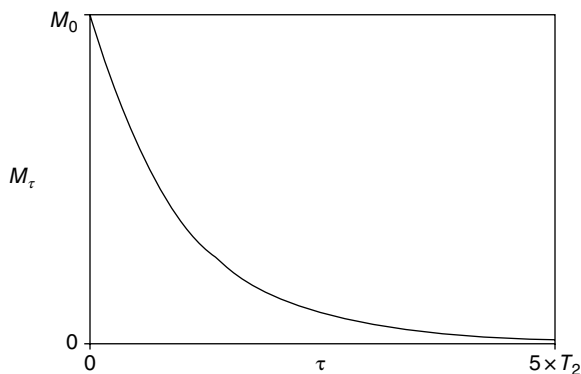


FIGURE 70.29 Loss of xy magnetization following an RF pulse. Virtually no transverse magnetization remains after a time period of $5 \times T_2$, where T_2 is the spin–spin relaxation time.

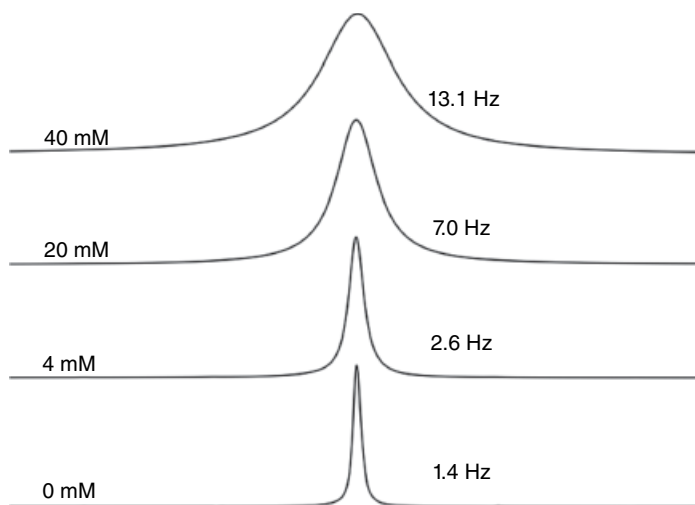


FIGURE 70.30 The effect of paramagnetic ions on proton NMR line widths of HOD. Acquisition parameters: 599.7 MHz; 25°C; 30° tip; 10.5 s total cycle time; 4 averages; 20 Hz spin rate. The samples contained approximately 20% (v/v) HOD in D₂O, along with 0–40 mM CuSO₄. A spectral region 100 Hz wide is shown in each case. Full widths at half maximum (FWHM) for the lines varied from 1.4 to 13.1 Hz. The $T_2^* = 1/(\pi \cdot \text{FWHM})$ so, for the 4 mM solution, T_2^* was 0.12 s. Its actual T_2 was measured to be 0.30 s, so most of the width was due to magnetic field inhomogeneity.

phase randomization is achieved. This process involves no energy change, but it does increase the sample's spin entropy. Therefore, spin–spin relaxation is an *entropic* process, in contrast to *enthalpic* spin–lattice relaxation.

The mathematical form of Equation 70.13 determines the line shapes in NMR spectra. As illustrated in Figure 70.10, exponential decay in the time domain produces Lorentzian lines. Moreover, the T_2 value for a time-domain NMR signal determines the minimum width of its Lorentzian peak: $\Delta\nu_{1/2} \geq 1/(\pi \cdot T_2)$, where $\Delta\nu_{1/2}$ is the full width of the peak at half-maximum (FWHM) height, in Hz. That has practical importance: fast relaxation (small T_2) produces broad NMR lines (large $\Delta\nu_{1/2}$), resulting in peak overlap, poor resolution, and perhaps reduced signal/noise ratio. Figure 70.30 shows proton NMR lines for aqueous solutions of copper(II) ion, a paramagnetic species that hastens relaxation. The line widths clearly increase with the ion concentration. The reason the sample's T_2 only sets the minimum width for the line is because static magnetic field inhomogeneity also broadens all the spectral lines, so only part of the width of any given experimental peak is due to its T_2 . The width that results from the combination of T_2 and magnetic field inhomogeneity is so widely encountered that it is given a special symbol: T_2^* (pronounced “tee too star”). This is an “effective” spin–spin relaxation time. In other words, T_2^* is the T_2 value that *would be* required to produce the experimentally observed line width *if* the B_0 field were perfectly homogeneous. When a raw NMR line width is used to calculate the relaxation time, it is always T_2^* and not T_2 that

is being measured. By using a well-shimmed magnet, the true T_2 value can often be estimated with rough accuracy from the line width alone, as shown in Figure 70.30. Mathematically, the relationships between T_2 , T_2^* , and their corresponding rates R ($=1/T$) are expressed by

$$\frac{1}{T_2^*} = \frac{1}{T_2} + \frac{1}{T_{\text{field inhomog}}} = R_2^* = R_2 + R_{\text{field inhomog}}.$$

Spin–lattice and spin–spin relaxation times can be measured accurately by using pulse experiments. The most common sequence for measuring T_1 is inversion recovery, but saturation recovery is also used. For T_2 , the Carr–Purcell–Meiboom–Gill (CPMG) sequence is usually employed. Such measurements are beyond the scope of this work but may easily be found in standard NMR references. All commercial NMR spectrometers come equipped with the required pulse sequences.

NMR spectra are almost always measured by acquiring multiple FIDs, which are then averaged to produce the final result. That requires repeating the pulse sequence multiple (n) times, in cycles of (RF pulse—acquisition—relaxation delay) $_n$. The total time between RF pulses is called the “cycle time.” It is tempting to acquire NMR spectra by using the smallest possible cycle time so the greatest number of FIDs can be acquired in a given period. By averaging more FIDs, the signal/noise ratio should increase. That line of reasoning is seductive, but it carries a hidden flaw: if the sample is not allowed to relax completely between pulses, then each pulse produces only a fraction of the full signal. A 90° RF pulse might produce a big NMR signal the first time it is applied, but if the pulse is repeated before the spin system has time to relax completely, the second FID will be less intense than the first. The two approaches, using a short cycle time so many FIDs can be acquired *versus* using a long cycle time that permits full relaxation, are mutually exclusive. There must be some intermediate cycle time that produces the best overall result. That optimum compromise is expressed by the equivalent equations:

$$\cos \alpha = e^{-(T_{\text{cycle}}/T_1)} \quad \alpha = \cos^{-1} \left\{ e^{-(T_{\text{cycle}}/T_1)} \right\} \quad \frac{T_{\text{cycle}}}{T_1} = -\ln(\cos \alpha).$$

In the equations, α is the so-called Ernst angle, T_{cycle} is the total time between RF pulses, and T_1 is the sample’s spin–lattice relaxation time. The Ernst angle is the tip angle that maximizes the signal/noise ratio in the spectrum for any given ratio of cycle time to T_1 . In practice, tip angles of 30°, 45°, and 60° are most widely used, and the equations show the corresponding optimum cycle times to be $0.14 \times T_1$ (for 30°), $0.35 \times T_1$ (for 45°), and $0.69 \times T_1$ (for 60°). Using a bigger tip angle produces more signal per FID, but it tends to saturate the spin system more and therefore requires a longer delay to permit relaxation. Note that these combinations maximize the signal/noise ratio but they also saturate the signals to some extent and will therefore distort the relative peak areas if the various spins have different T_1 values.

70.10.3 Quantitative Analysis by NMR

Providing spin saturation is avoided, the area under an NMR peak is directly proportional to the number of nuclei that produce the peak. There is no need to scale the area to account for differences in detector sensitivity, making NMR an obvious method for quantitative analysis. There are two problems, however. First, compared to other instrumental methods, NMR is not very sensitive so large samples and/or high analyte concentrations must be used. The second problem is relaxation. Measurements must be performed in a way that does not distort the relative peak areas in the spectrum. That automatically precludes running under conditions that yield the optimum signal/noise ratio per unit time. The only practical method is to choose a cycle time much longer than T_1 , usually at least $3 \times T_1$ but often as much as $5 \times T_1$. From Equation 70.12 with $A = 1$, it is easy to show that using a cycle time of $3 \times T_1$ causes a 4.98% error in the peak area while a cycle time of $5 \times T_1$ produces only a 0.67% error. The latter is generally acceptable and is widely used for the best work. Having to wait $5 \times T_1$ between pulses implies a slow experiment, and that can certainly be the case. However, it is often possible to speed relaxation by adding a paramagnetic substance. Paramagnetics not only can reduce the T_1 values, but, in samples with nuclei that relax at very different rates, the paramagnetic can help equalize the T_1 's as it simultaneously reduces them. Although the advantages are clear, paramagnetic compounds also tend to broaden the NMR peaks (Fig. 70.30) so the temptation to use too much must be resisted.

An example of tissue phospholipid quantitation by ^{31}P NMR [25, 26] is shown in Figure 70.31. The T_1 's for the various compounds were long (seconds), especially for the tributylphosphate internal standard, so a paramagnetic was added to decrease them and the differences between them. Since lipid analytes are generally nonpolar, the solvent (chloroform) was also necessarily nonpolar, making it impossible to dissolve simple ions like copper(II). Fortunately, highly paramagnetic chromium(III) can be purchased as the relatively nonpolar acetylacetonate complex ($\text{Cr}(\text{acac})_3$). By dissolving that complex in the sample, the ^{31}P T_1 values were reduced enough to allow accurate quantitation of phospholipids from 0.5 to 1 g of tissue in a 2 h measurement. With the 7.06T magnetic field used, considerable peak overlap was unavoidable. However, accurate peak areas were still obtained by fitting theoretical Lorentzian peaks to the experimental spectrum.

70.11 DYNAMIC PHENOMENA IN NMR

NMR is a powerful tool for analyzing molecular structures, but it has another important application: it can reveal crucial information about molecular dynamics. Molecules in solution undergo rapid, continuous motion, and those motions greatly simplify NMR spectra by averaging the chemical shifts, which depend on the molecule's orientation relative to the B_0 direction. Most solution spectra contain narrow, discrete lines

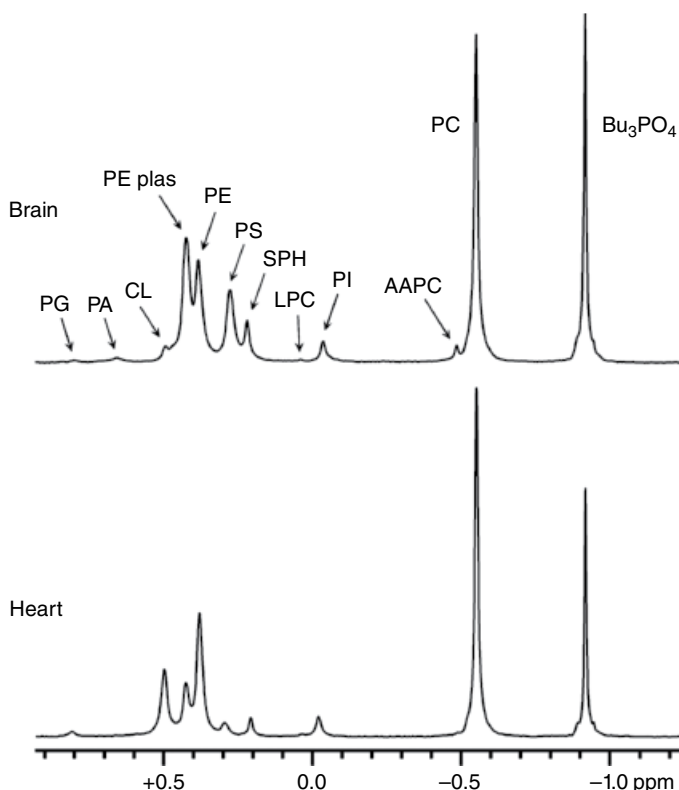


FIGURE 70.31 ^{31}P NMR spectra used to quantify the phospholipids in extracts of rat brain and heart. Tributylphosphate was added as an internal standard, and the acetylacetonate complex of chromium(III) was included to reduce the T_1 values of the compounds. A 7.06 T magnet was used, giving a ^{31}P resonance frequency of 121.6 MHz. The peak labels designate various classes of phospholipids, which are dominated by phosphatidylcholine (PC).

due to that motional averaging and are said to be in the “fast exchange limit.” In some cases, the motion is slow enough or can intentionally be made slow enough to prevent complete averaging, thus allowing NMR to be used to study the dynamics of processes. Whether such effects can be observed depends on their rates relative to the “NMR time scale.” Suppose a molecule can adopt either of two conformations that exhibit two ^1H NMR lines separated by, say, 1 ppm. If we acquire the spectrum of this compound with a 300 MHz spectrometer, the two lines will be separated by 300 Hz. If the compound is able to switch conformations faster than roughly 300 times per second, the two distinct lines will tend to average and merge into one. On a 600 MHz spectrometer, the exchange rate between the two conformations must occur twice as fast to average the same lines, which are now separated by 600 Hz. So, it is clear that the “NMR time scale” is not fixed but depends on the conditions of the experiment. Despite that, NMR can provide insights into molecular dynamics that are challenging to obtain in other ways.

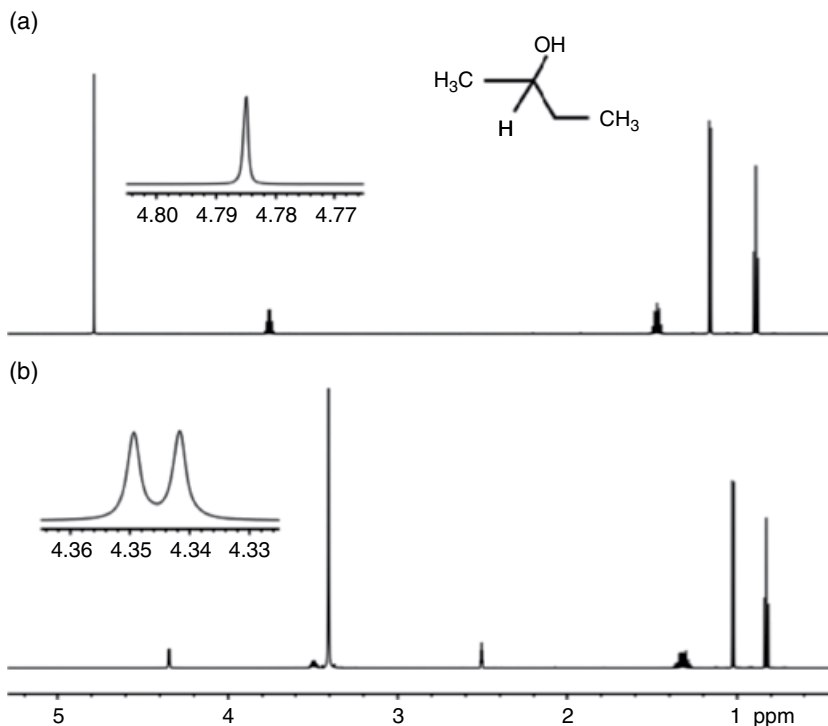


FIGURE 70.32 Proton NMR spectrum of approximately 5% (v/v) 2-butanol dissolved in (a) D_2O and (b) $DMSO-d_6$. Acquisition parameters: 599.7 MHz; 25°C; 30° tip; 10.5 s total cycle time; 8 averages; 20 Hz spin rate. The chemical shift reference was internal TSP (for D_2O) or TMS (for $CDCl_3$). In spectrum (a), the butanol $-OH$ proton exchanged rapidly with the solvent and merged with the HOD peak to form a singlet near 4.785 ppm. In $DMSO-d_6$ (b), the exchange was slowed sufficiently to produce a separate butanol $-OH$ peak near 4.35 ppm. The water peak appeared near 3.4 ppm, and residual protonated DMSO produced the multiplet at 2.5 ppm.

If the molecular exchange rate is slow on the NMR time scale, distinct peaks will be observed for each form of the molecule. This is the “slow exchange limit.” An example of that has already been encountered in Figure 70.23, which contains the spectrum of 2,4-pentanedione, a keto-molecule that can tautomerize to form an enol. In solution both tautomers exist in equilibrium simultaneously and, at room temperature, the exchange rate from one form to the other is slow enough to produce very distinct NMR peaks. By integrating peaks associated with each form, the equilibrium constant can be calculated.

The exchange rate between chemical forms can sometimes be manipulated to either produce or eliminate multiple peaks in the spectrum. An example of using the solvent to do that is shown in Figure 70.32. The 1H spectrum (a) is that of 2-butanol dissolved in D_2O . Hydrogen bonding between the alcohol and solvent promotes relatively fast exchange of the alcohol $-OH$ proton with the D_2O to produce a single, averaged singlet near 4.8 ppm. Dissolving the alcohol in $DMSO-d_6$ slows the exchange rate

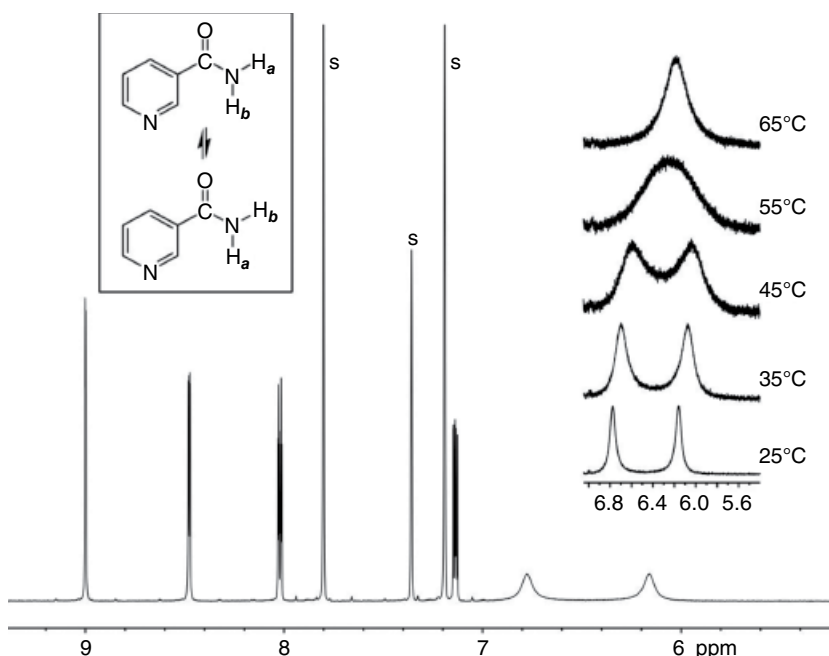


FIGURE 70.33 Proton NMR spectra of nicotinamide in nitrobenzene- d_5 . Acquisition parameters: 599.7 MHz; 30° tip; 5.5 s total cycle time; 8 averages; 20 Hz spin rate. The temperature was 25°C except where noted. Peaks labeled “s” arise from residual ^1H in the solvent, which was used due to its high boiling point. Delocalization of lone-pair electrons from the nitrogen atom into the carbonyl group endows the C—N bond with partial double-bond character, inhibiting free rotation. Consequently, the two amido protons are not equivalent and produce separate peaks. Raising the temperature increases the rotation rate about the bond and averages the amido proton chemical shifts. The two peaks just merge at 54°C .

enormously, allowing a separate, distinct alcohol —OH peak to be observed (b). Diminished hydrogen bonding also shifts the peak upfield from its position in spectrum (a). With slow ^1H exchange, the alcohol —OH proton remains attached to the molecule long enough to manifest spin coupling with the lone proton on the adjacent carbon atom, yielding a doublet that proves it to be a secondary alcohol. This technique can usually be used to determine unambiguously whether an unknown alcohol is primary, secondary, or tertiary.

The choice of solvent is not the only approach for manipulating the dynamics of molecular exchange processes. Figure 70.33 shows how temperature can be used to the same end. In nicotinamide, as in most amide compounds, partial double-bond character in the C—N bond slows its rotation and makes the two amido hydrogen atoms inequivalent. Each hydrogen produces a separate broad resonance peak at a different chemical shift. (Broadness in NMR peaks usually suggests the presence of some dynamic process.) As shown in the figure, the molecule can be considered to exist in two distinct states where the positions of those hydrogens are interchanged. By raising the sample

temperature, the rotational rate about the C—N bond increases and leads to averaging on this time scale, causing the two spectral peaks to converge and broaden. The temperature at which they just merge into a single, flat-topped peak is called the coalescence temperature (54°C, in this case). Continued increases in temperature narrow the peak. At the coalescence temperature, the lifetime $\tau_{\text{coalescence}}$ for each of the two states can be calculated from

$$\tau_{\text{coalescence}} = \frac{1}{k_{\text{coalescence}}} = \frac{\sqrt{2}}{\pi \cdot \Delta\nu}, \quad (70.14)$$

where $k_{\text{coalescence}}$ is the pseudo-first-order rate constant (in Hz) for exchange between the two states and $\Delta\nu$ is the chemical shift difference (in Hz) between the two peaks in the slow exchange limit (i.e., at very low temperatures). By using the chemical shift difference (369.5 Hz) at 25°C as an estimate of $\Delta\nu$ for nicotinamide, Equation 70.14 shows the lifetime to be 1.22 ms at 54°C, with a corresponding exchange rate of 821 s⁻¹. Since the true $\Delta\nu$ is undoubtedly larger than the estimate, perhaps by severalfold, these values should be considered as the upper limit for the lifetime and lower limit for the exchange rate. Even so, they are informative. Although the exchange phenomenon in this particular example involves peaks with different chemical shifts, exactly the same kind of averaging can occur when the peaks arise from scalar coupling instead. In other words, molecular motions can average both chemical shift and scalar coupling information.

It is possible to calculate the shapes of NMR spectral lines in exchanging systems for any combination of $\Delta\nu$ and lifetime τ . (See Ref. 27 for a particularly clear discussion.) Fitting the theoretical shapes to experimental spectra allows the exchange rate constant to be found at temperatures that differ from the coalescence temperature. Variable-temperature NMR is widely used to find thermodynamic parameters for molecular inversion and other dynamic processes [28]. For intramolecular exchange processes such as rotation, the thermodynamic activation parameters can be calculated from the relationship

$$k_r = \kappa \frac{kT}{h} \cdot e^{\left(-\Delta G^\ddagger / RT\right)} = \kappa \frac{kT}{h} \cdot e^{\left(-\Delta H^\ddagger / RT\right)} \cdot e^{\left(-\Delta S^\ddagger / RT\right)}, \quad (70.15)$$

where k_r is the exchange rate constant ($=1/\tau$), κ is a constant that depends on the nature of the process, k is Boltzmann's constant, T is absolute temperature, h is Planck's constant, R is the gas-law constant, ΔG^\ddagger is the free energy of activation, ΔH^\ddagger is the enthalpy of activation, and ΔS^\ddagger is the entropy of activation. The value of κ depends on the particular system, but in practice, it is often approximated as 1.

A practical complication in variable-temperature NMR studies is that, while commercial spectrometers are reasonably good at maintaining a constant, uniform sample temperature, they often do not report it very accurately. The actual and displayed temperature can easily differ by several degrees. That is not necessarily the

fault of the manufacturer. The actual sample temperature depends on several variables such as the gas flow rate through the NMR probe, the volume of sample in the NMR tube, and whether broadband decoupling is being used. In the latter case, the RF pulses from the decoupler can actually heat the sample by $\geq 10^{\circ}\text{C}$, especially when using a solvent with a high dielectric constant. So, it is difficult to produce a temperature calibration that is reliable under all conditions. Consequently, careful variable-temperature work usually requires one or more extra NMR experiments to measure the actual sample temperature. These are not difficult to perform and rely on known effects of temperature on the chemical shifts for various compounds. Among the most widely used compounds for ^1H NMR are methanol (MP = -97.8°C ; BP = 64.7°C) for work at low temperatures and ethylene glycol (MP = -13°C ; BP = 197.6°C) for high-temperature studies. The ^1H NMR spectra of both of those compounds contain only two peaks, and the chemical shift difference between them depends almost linearly on temperature [29]. The compound tris(trimethylsilyl)methane has been proposed as an “NMR thermometer” for measurements in ^{13}C NMR [30]. It is normally unwise, if not impossible, to add the temperature calibrant directly to the sample. Instead, the usual procedure is to place the compound in a glass capillary, insert the capillary into the NMR sample, allow the temperature to equilibrate, and then run a quick spectrum to obtain the chemical shift data. The capillary is likely to degrade the shim, but it will still be good enough to permit accurate temperature measurements. Another option is to fill a separate NMR tube with the calibrant, insert it into the spectrometer in place of the original sample, allow its temperature to equilibrate, and then measure the necessary chemical shift(s). A final possibility is to insert a thermistor into the center of the sample. That can work well, but the wire leads from the thermistor will prevent the tube from spinning, which might affect the sample temperature slightly. RF pulses should never be applied while the thermistor is in place since its leads can act as antennas, concentrating the energy and causing the sample to heat.

70.12 MULTIDIMENSIONAL NMR

The NMR methods described earlier demonstrate the enormous power of the technique. However, with increasing applications to larger and more complex molecules, and especially to biopolymers, the spectra tend to become extremely crowded and exhibit extensive peak overlap even when the highest available magnetic fields are employed. One solution to the crowding problem is to spread the peaks into more than one dimension, and that is exactly what multidimensional NMR techniques do. However, merely dispersing the peaks better, as important as that may be, is only one of the many advantages of multidimensional NMR. Indeed, this approach can provide a tremendous range of information quickly that was previously impossible, or at least difficult and slow, to obtain by 1D methods. As a result, multidimensional methods have revolutionized the practice of NMR since their introduction in the 1970s.

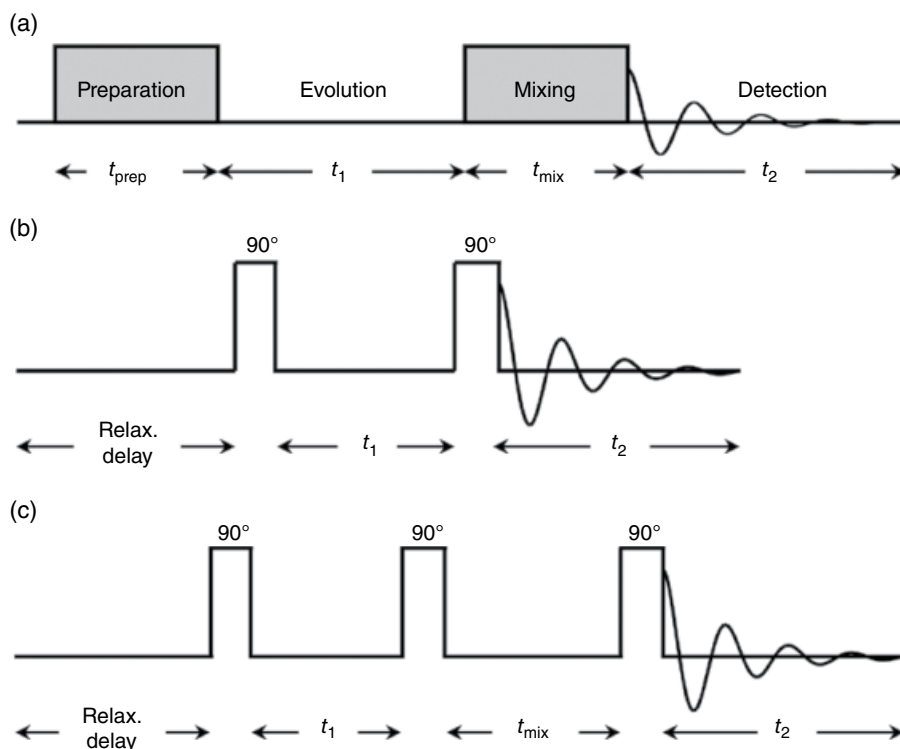


FIGURE 70.34 Two-dimensional NMR pulse sequences. (a) General characteristics. (b) The basic pulse sequence for correlation spectroscopy (COSY). (c) The basic pulse sequence for nuclear Overhauser effect spectroscopy (NOESY).

This brief discussion will be limited to 2D techniques (2D NMR), although extension to additional dimensions is straightforward. In addition to the relaxation delay, short receiver recovery delay, and RF pulse width, a 1D NMR spectrum is normally acquired by using only one other time period: the acquisition time during which the FID is recorded. We define that time period as t_2 (not to be confused with the spin–spin relaxation time!). The principle of 2D NMR, first expressed by Jean Jeener at an AMPERE Conference in 1971, involves using a pulse sequence that contains *two* independent time periods t_1 and t_2 . These are called the evolution and detection periods and are illustrated in Figure 70.34. The actual FID is acquired during period t_2 , just as in 1D NMR, but any evolution the spin system undergoes during period t_1 will also be reflected in the FID. In addition to these two periods, the 2D pulse sequence also includes an initial preparation period, which is usually a relaxation delay of fixed length followed by a 90° RF pulse to excite the spin system. A few sequences include a mixing time, also of fixed length, to allow the exchange of information between different spins.

Figure 70.34 shows the basic pulse sequence for the “correlation spectroscopy” (COSY) experiment. It begins with a relaxation delay and a 90° RF pulse. That is followed by the time period t_1 . On the first pass, a t_1 value of essentially zero is used. Then

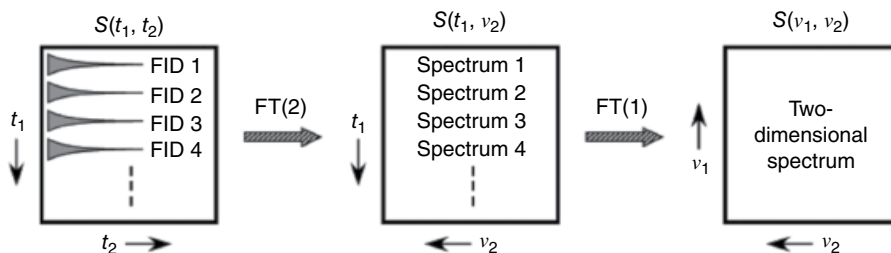


FIGURE 70.35 Steps in processing a 2D NMR data set. Each FID is first Fourier transformed in the t_2 direction, as usual. The first points in each of resulting spectra can be assembled into a set and considered to be another FID but this time along the t_1 direction. Likewise, the set of second points in all the spectra form another FID in the t_1 direction and so on. These FIDs are Fourier transformed in the t_1 direction. Ultimately, the data set that began as a function of (t_1, t_2) has been transformed into the (ν_1, ν_2) frequency domain: a 2D NMR spectrum.

the second RF pulse is applied, and the FID is acquired and stored in the computer. We will define it as FID 1. The sequence is then repeated by starting again with the relaxation delay and 90° pulse. This time, we use a t_1 delay that is increased by a “dwell” time (DW) of, often, several hundred microseconds to give the spin system time to evolve. Then the second pulse is applied and FID 2 is acquired and stored separately from FID 1. The whole process is repeated again, this time using $t_1 = 2 \times \text{DW}$ and saving FID 3. It is common to repeat the sequence $n = 256\text{--}1024$ times to acquire an array of n FIDs, with each one representing an evolution time of $t_1 = (n - 1) \times \text{DW}$.

Now, what do we do with the set of n FIDs? As illustrated in Figure 70.35, we first Fourier transform each FID in the usual manner. Since they were all acquired during the period t_2 , the frequency-domain axis is designated as ν_2 . Inspecting the resulting n frequency domain spectra reveals that their peak amplitudes change (oscillate) from one spectrum to the next. So, we Fourier transform in the t_1 direction to create frequency-domain spectra along a ν_1 axis in that direction. The data set has now been transformed from a 2D time-domain function $s(t_1, t_2)$ to a frequency-domain function $S(\nu_1, \nu_2)$. This is a 2D NMR spectrum. If nothing “interesting” occurred during the evolution period t_1 , then the 2D NMR spectrum will simply contain the conventional 1D spectrum spread along its diagonal axis. However, “interesting” things normally do happen during t_1 , resulting in the appearance *extra peaks* in the 2D spectral plane to accompany the 1D spectrum that lies along the diagonal.

When extra peaks appear in a 2D NMR spectrum, their meaning depends on exactly which pulse sequence was used to acquire the data. Figure 70.36 contains a schematic illustration of a COSY spectrum, where cross peaks arise only when the spins are J coupled. In the diagram, the cross peaks labeled “A” appear at the intersection of the chemical shifts of peaks 1 and 4 along the diagonal, meaning that the spins that produce peaks 1 and 4 are spin coupled. The cross peaks labeled “B” show that the nuclei of peak 4 are also coupled to the nuclei that produce peak 2. The cross peaks “C” also indicate coupling between the nuclei that give rise to peaks 4 and 5.

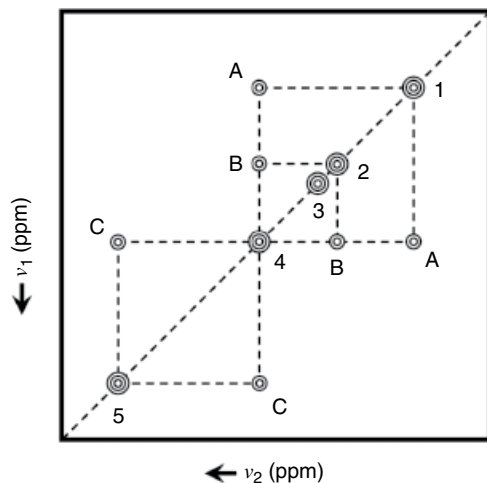


FIGURE 70.36 Schematic diagram of a COSY spectrum shown as a contour plot. The 1D NMR spectrum lies along the diagonal. Cross peaks correlate various peaks in the 1D NMR spectrum, showing that they are spin coupled. The cross peak positions shown lead to the following conclusions. The nuclei that produce peak 1 are coupled to those that produce peak 4 (cross peaks A). The peak 1 nuclei are not coupled to those of peak 2 (no cross peaks). Peak 2 is coupled to peak 4 (cross peaks B). Peak 3 is not coupled to any other peaks. Peak 4 is coupled to peaks 1, 2, and 5. Peak 5 is coupled only to peak 4 (cross peaks C).

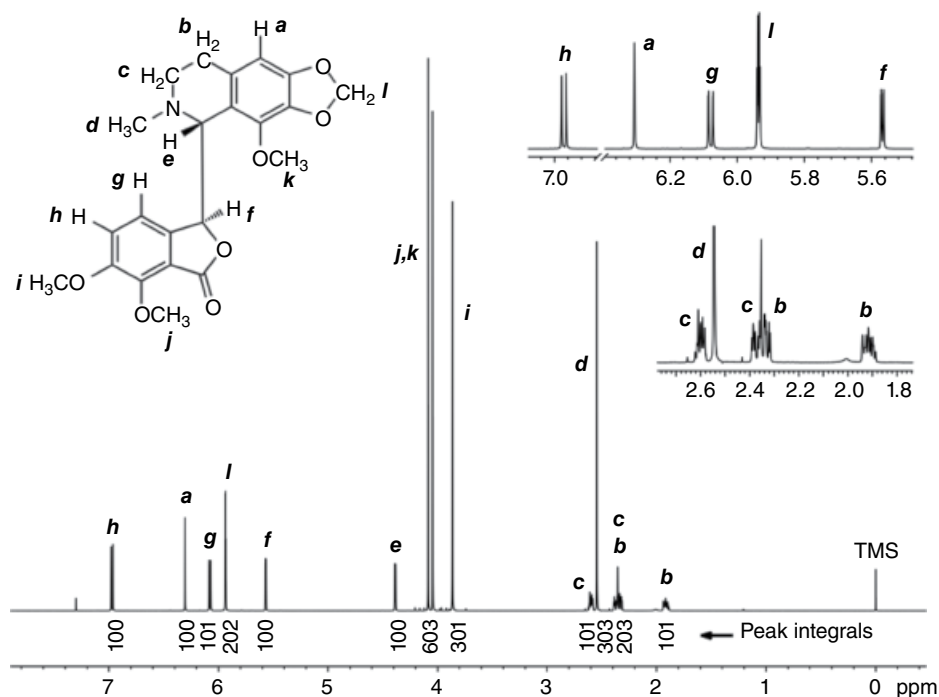


FIGURE 70.37 ^1H 1D NMR spectrum of 150 mg noscapine in 1 mL CDCl_3 . Acquisition parameters: 599.7 MHz; 25°C; 30° tip; 11.9 s total cycle time; 8 averages; 20 Hz spin rate; TMS chemical shift reference. Some of the peaks can be assigned by using the 1D spectrum alone, such as the lone aromatic proton *a* singlet, the singlet for the two *l* protons that are highly deshielded by adjacent oxygen atoms, and the singlet for the three methyl protons *d*. Some of other assignments are more challenging and are aided by 2D NMR methods.

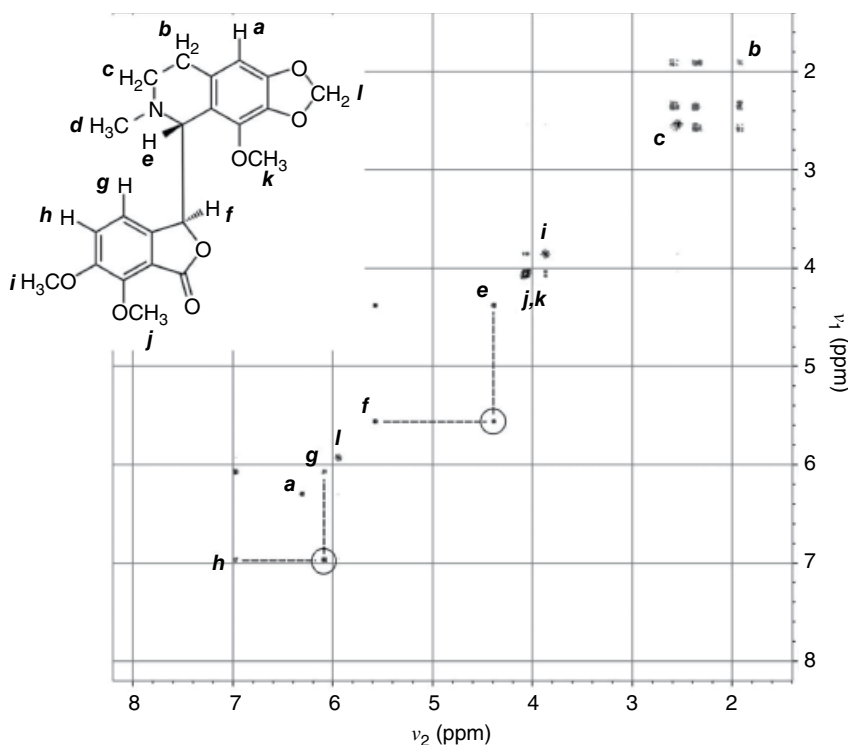


FIGURE 70.38 Experimental ^1H COSY spectrum of 150 mg noscapine in 1 mL CDCl_3 . Acquisition parameters: 599.7 MHz; 25°C; 90° tip; 1.0 s relaxation delay; 256 t_1 values; 8 averages per t_1 value; 20 Hz spin rate; TMS chemical shift reference. Total data acquisition time was 41.5 min. Cross peaks confirm J coupling between the three upfield multiplets due to protons **b** and **c**. Cross peaks also show coupling between protons **e** and **f**, as well as between **g** and **h** (dashed lines). The apparent cross peaks between proton peaks **i** and **j**; **k** are artifacts due to overlap at the bases of these intense peaks.

An experimental ^1H 1D NMR spectrum of noscapine is shown in Figure 70.37. Some of the peaks can be assigned unambiguously, such as the singlets. However, the assignment of other peaks is more difficult to perform and can benefit from 2D NMR data. Figure 70.38 shows the COSY spectrum, which not only confirms J coupling between peaks **e** and **f**, as well as **g** and **h**, but also shows that all three multiplets in the 1.8–2.6 ppm region are mutually coupled.

Even more informative is the NOESY spectrum (Fig. 70.39). In NOESY, the pulse sequence includes a mixing time that allows nuclear Overhauser interactions to occur between excited spins. The NOE is a *through-space* interaction, so the appearance of cross peaks in a NOESY spectrum is evidence that the corresponding nuclei are near each other in space. If the mixing time is short, cross peaks appear only if the spins are very near each other. A longer mixing time allows spins that are more widely separated to produce cross peaks. In the case of noscapine, the NOESY spectrum confirms that

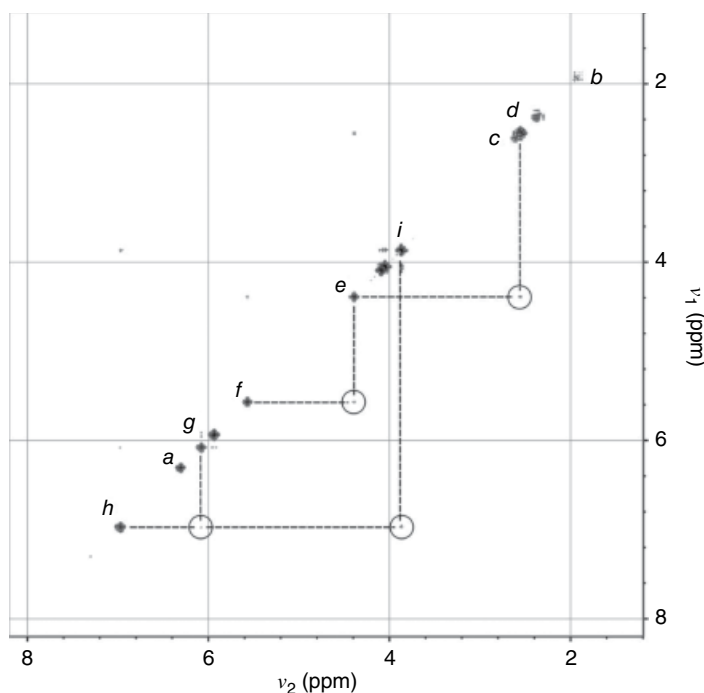


FIGURE 70.39 Experimental ^1H NOESY spectrum of 150 mg noscapine in 1 mL CDCl_3 . Acquisition parameters: 599.7 MHz; 25°C ; 90° tip; 2.0 s relaxation delay; 256 t_1 values; 4 averages per t_1 value; 500 ms mixing time; 20 Hz spin rate; TMS chemical shift reference. Total data acquisition time was 1.5 h. All cross peaks appear symmetrically relative to the diagonal, but for clarity, only those in the lower half of the spectrum are circled. The cross peaks show that proton *e* is spatially near protons *d* and *f* but that *d* and *f* are not near each other (no cross peaks link the two). Likewise, proton *h* is near protons *g* and *i*, but *g* and *i* are not near each other (again, no cross peaks link them directly).

proton *e* is spatially near protons *d* and *f*, although protons *d* and *f* are not near each other. It is therefore logical that proton *e* should occupy a position between *d* and *f*, as shown in the chemical structure. In addition, the peak for proton *h* exhibits cross peaks to the peaks of both protons *g* and *i*, again consistent with the proposed structure. That also allows the singlet for the methoxy protons *i* to be assigned to the peak at 3.87 ppm, narrowing the possible assignments of the other two methoxy groups *j* and *k* to the intense singlets at 4.05 and 4.08 ppm (though the order is not clear). 2D NMR methods provide more information than can be obtained from 1D NMR spectra alone, and the figures demonstrate how that additional information can sometimes prove crucial in making NMR peak assignments.

A great many different 2D NMR experiments have been developed, but the great majority of studies only employ a few of them. Table 70.4 lists the most important ones and includes brief descriptions and other information.

TABLE 70.4 Some Common Two-Dimensional NMR Pulse Experiments^a

Experiment Name	Typical Speed	Features
COSY: <u>C</u> orrelation <u>S</u> pectroscopy	Relatively fast (e.g., 10–40 min)	This homonuclear experiment is widely used for proton spectra. Cross peaks reveal which spins are scalar coupled. It works best if $J = 3\text{--}15\text{ Hz}$ (values typical of geminal or vicinal protons). Excellent for molecular structure analysis and assigning spectra.
TOCSY: Total <u>C</u> orrelation <u>S</u> pectroscopy	Relatively fast (e.g., 10–40 min)	Like COSY, cross peaks show which pairs of spins have scalar coupling. Compared to COSY, this experiment is more sensitive to weak homonuclear coupling and yields cross peaks even from widely separated spins with small J values.
HSQC: <u>H</u> eteronuclear <u>S</u> ingle- <u>Q</u> uantum <u>C</u> orrelation	Moderately fast (e.g., 20–60 min)	A heteronuclear 2D method with the spectrum of one nuclide I (e.g., ^1H) along axis v_2 and that of the other nuclide S (e.g., ^{13}C) along the other axis v_1 . A cross peak appears if an I spin is directly J coupled to an S spin. Very useful for assigning spectra. In the v_2 direction, the cross peaks are multiplets split by $^1J_{IS}$. The HSQC signal is detected from the abundant I spins, so the experiment is relatively fast.
HMQC: <u>H</u> eteronuclear <u>M</u> ultiple- <u>Q</u> uantum <u>C</u> orrelation	Moderately fast (e.g., 20–60 min)	Like HSQC, this heteronuclear experiment yields a 2D spectrum with the I nuclide (e.g., ^1H) along one axis and the S nuclide (e.g., ^{13}C) along the other. A cross peak indicates direct J coupling of an I spin with an S spin. The S peaks show splitting due to coupling between I spins and, so, are a little broader than those in HSQC.
HETCOR: <u>H</u> eteronuclear <u>C</u> orrelation	Slow (usually hours)	This is also a heteronuclear experiment. It is analogous to HSQC, but the signal is detected from the low-abundance S spins (e.g., ^{13}C), resulting in relatively poor signal/noise ratio per unit time.
NOESY: <u>N</u> uclear <u>O</u> verhauser <u>E</u> nhancement <u>S</u> pectroscopy	Slow (usually hours)	The HETCOR experiment has been displaced by HSQC or HMQC for most purposes. A homonuclear experiment, usually performed with ^1H . Cross peaks occur between spins that are near each other in space. (The cross peak intensities and growth rates depend on r^6 .) Especially useful for small molecules. Molecules with formula weights of 1000–3000 often yield weak or no cross peaks due to their unfavorable molecular correlation times.
ROESY: <u>R</u> otating <u>F</u> rame <u>O</u> verhauser <u>E</u> nhancement <u>S</u> pectroscopy	Slow (usually hours)	Similar to NOESY in that cross peaks indicate spatial proximity between different spins. Unlike NOESY, molecules with formula weights of 1000–3000 produce strong cross peaks. ROESY is often a bit faster than NOESY but can still require hours.
HOESY: <u>H</u> eteronuclear <u>O</u> verhauser <u>E</u> ffect <u>S</u> pectroscopy	Slow (usually hours)	This is a heteronuclear analogue of NOESY. A cross peak appears in the spectrum when an I spin is spatially near an S spin. It can be used, for example, to investigate ^1H - ^{13}C and ^1H - ^{31}P distances within molecules.

^aListed are some 2D NMR experiments and their general features. The ones given here constitute a small fraction of all such experiments that have been developed, but they are among the most widely used. (The most popular is probably COSY.) Some of the experiments are *homonuclear* (i.e., the spectra along the v_1 and v_2 axes are from the same nuclide, such as ^1H). Other experiments are *heteronuclear*, having the spectrum of nuclide I (such as ^1H) along one axis and the spectrum of a different nuclide S (like ^{13}C) along the other axis. The experiments are divided into (i) homonuclear correlation methods, and (ii) heteronuclear correlation methods, and (iii) methods that detect the spatial proximities of different spins.

70.13 CONCLUSION

From the information presented in this chapter, the power and enormous range of applications of NMR should be apparent. While it is not possible to cover 70 years of development in a single chapter, the goal has been to provide a clear introduction to both the principles of NMR and to many of its modern techniques. It is hoped that this information will serve as a useful overview and a springboard for those who seek additional information in the field.

REFERENCES

1. W. Gerlach and O. Stern, *Ann. Phys.* 1924, 74, 673–699.
2. I. I. Rabi, J. R. Zacharias, S. Millman, and P. Kusch, *Phys. Rev.* 1938, 53, 318.
3. G. J. Gorter and L. F. J. Broer, *Physica* 1942, 9, 591–596.
4. F. Bloch, W. W. Hansen, and M. Packard, *Phys. Rev.* 1946, 69, 127.
5. E. M. Purcell, H. C. Torrey, and R. V. Pound, *Phys. Rev.* 1946, 69, 37–38.
6. W. D. Knight, *Phys. Rev.* 1949, 76, 1259–1260.
7. W. G. Proctor and F. C. Yu, *Phys. Rev.* 1950, 77, 717.
8. R. R. Ernst and W. A. Anderson, *Rev. Sci. Instrum.* 1966, 37, 93–102.
9. R. N. Bracewell, “*The Fourier Transform and its Applications*,” 2nd edition, McGraw-Hill Book Company, New York, 1978, p. 104.
10. G. A. Morris, *J. Magn. Reson.* 1988, 80, 547–552.
11. K. R. Metz, M. M. Lam, and A. G. Webb, *Concepts Magn. Reson.* 2000, 12, 21–42.
12. L. M. Tolbert, *Acc. Chem. Res.* 1992, 25, 561–568.
13. Y. Gaoni, A. Melera, F. Sondheimer, and R. Wolovsky, *Proc. Chem. Soc.* 1964, 397–398.
14. A. Kuhn, P. Sreeraj, R. Pöttgen, H.-D. Weimhöfer, M. Wilkening, and P. Heitjans, *Angew. Chem. Int. Ed.* 2011, 50, 12099–12102.
15. L. L. Borer, J. G. Russell, R. E. Settlage, and R. G. Bryant, *J. Chem. Educ.* 2002, 79, 494–497.
16. R. K. Harris and B. E. Mann (eds.), “*NMR and the Periodic Table*,” Academic Press, New York, 1978.
17. P. Laszlo (ed.), “*NMR of Newly Accessible Nuclei*,” Vol. 1 and 2, Academic Press, New York, 1983.
18. L. M. Jackman and S. Sternhell, “*Applications of Nuclear Magnetic Spectroscopy in Organic Chemistry*,” 2nd edition, Vol. 5 (International Series of Monographs in Organic Chemistry), Pergamon Press, Oxford, 1969.
19. J. B. Stothers, “*Carbon-13 NMR Spectroscopy*,” Academic Press, New York, 1972.
20. R. M. Silverstein, F. X. Webster, and D. J. Kiemle, “*Spectrometric Identification of Organic Compounds*,” 7th edition, John Wiley & Sons, Inc., Hoboken, NJ, 2005.
21. J. B. Lambert, H. F. Shurvell, D. A. Lightner, and R. G. Cooks, “*Organic Structural Spectroscopy*,” Prentice Hall, Upper Saddle River, NJ, 1998.

22. P. Crews, J. Rodriguez, and M. Jaspars, “*Organic Structure Analysis*,” 2nd edition, Oxford University Press, New York, 2010.
23. C. J. Pouchert and J. Behnke, “*The Aldrich Library of ^{13}C and ^1H FT NMR Spectra*,” 1st edition, Vol. 1–3, Aldrich Chemical Company, Inc., Milwaukee, WI, 1993.
24. M. Foroozandeh, R. W. Adams, N. J. Meharry, D. Jeannerat, M. Nilsson, and G. Morris, *Angew. Chem. Int. Ed.* 2014, 53, 6990–6992.
25. K. R. Metz and L. K. Dunphy, *J. Lipid Res.* 1996, 37, 2251–2265.
26. K. R. Metz and L. K. Dunphy, *J. Lipid Res.* 1996, 38, 1275.
27. R. K. Harris, “*Nuclear Magnetic Resonance Spectroscopy*,” Pitman Books Limited, London, 1983, pp. 121–126.
28. L. T. Scott, M. M. Hashemi, and M. S. Bratcher, *J. Am. Chem. Soc.* 1992, 114, 1920–1921.
29. A. L. Van Geet, *Anal. Chem.* 1968, 40, 2227–2229.
30. W. Sikorski, A. W. Sanders, and H. J. Reich, *Magn. Reson. Chem.* 1998, 36, S118–S124.

NEAR-INFRARED SPECTROSCOPY AND ITS ROLE IN SCIENTIFIC AND ENGINEERING APPLICATIONS

BRAD SWARBRICK

Quality by Design Consultancy, Sydney, New South Wales, Australia

71.1 INTRODUCTION TO NEAR-INFRARED SPECTROSCOPY AND HISTORICAL PERSPECTIVES

71.1.1 A Brief Overview of Near-Infrared Spectroscopy and Its Usage

Until recently, the near-infrared (NIR) region of the electromagnetic spectrum was not discussed in too great detail in undergraduate university courses. This was primarily due to NIR spectra not containing the sharp, well-defined absorbance bands, typical of methods such as the mid-infrared (MIR) or nuclear magnetic resonance (NMR) spectroscopy, therefore making identification of functional groups difficult. However, during the past 20-year period, the availability of high power computers and the development of mathematical methods such as chemometrics (discussed in detail in Chapter 65) has helped scientists and engineers utilize this most informative region of the spectrum.

The NIR region lies between the visible and MIR regions (i.e., between 700 and 2500 nm) and is associated with bond stretching phenomena between molecules with large dipole moments, in particular, the stretching frequencies associated with molecules containing carbon–hydrogen (C–H), oxygen–hydrogen (O–H), and nitrogen–hydrogen (N–H). The specificity for these large dipole moment bonds has allowed the NIR method to be used in a number of industrial applications for the

prediction of quality parameters at the point of manufacture, including sectors such as agricultural, pharmaceutical, and food and beverage, to name just a few.

The main benefits of NIR spectroscopy lie in the high signal-to-noise (S/N) of the instrumentation and the minimization of sample preparation required to collect reliable spectra. Industries, such as the agricultural sector, have used the technique for many years to measure constituents such as protein, moisture, starch, and oil in grains and oilseeds when they are delivered to grain handlers during harvest (refer to Section 71.6.1 for more details). In these applications whole grains are loaded into the spectrometer and scanned “as is,” yielding in one measurement as many as 10 or more constituents, using chemometric models. Of even greater value to the end user has been the implementation of NIR to pharmaceutical applications (Section 71.6.2). Whole pharmaceutical tablets can be analyzed nondestructively for the quantification of active ingredient content, again at the point of manufacture, thus bringing the laboratory to the process operator and providing greater assurance of quality through an entire manufacturing run. A much greater discussion of the many applications of the NIR method is presented in Section 71.6 of this chapter.

NIR spectrometers are also highly versatile. They can be installed in environments such as clean laboratories, pharmaceutical manufacturing plants, right through to harsh conditions found in feed mills, and petrochemical plants. The instrumentation can typically be used to standalone besides the process (at line) or can be integrated directly to the process through the use of specialized interfaces and fiber optic cables (in line). Recently, a number of manufacturers have developed so-called microspectrometers, which can be adapted to the most difficult of sampling situations. The various technologies used for developing NIR spectrometers are discussed in more detail in Section 71.3 of this chapter.

As the challenges of manufacturing high-quality products at competitive pricing continue into the future for all industrial sectors, the NIR technique has found and will continue to find many more applications for meeting such challenges. NIR can now be considered to be a mature technology and has widespread acceptance as an alternative testing method to establish reference methods, for example, the British, European, and US Pharmacopoeias have dedicated chapters to the NIR method for use in the pharmaceutical and biopharmaceutical sectors. Many groups such as the American Society for Testing and Materials (ASTM) and the US Food and Drug Administration (US FDA) have provided guidance on how to implement NIR into a number of industrial situations [1, 2].

NIR is also highly suited to research and development applications. The method can be highly sensitive to both constituents of interest and the entire sample matrix. Many research groups have used the NIR technique for classifying existing samples into known classes and also isolating new classes, particularly in biological and ecological research [3]. This is again attributed to the ability of NIR to measure the sample, as it exists in nature, without the need for sample preparation. Heterogeneity of natural samples is a typical challenge for the NIR scientist to address; however, with the

portability of some instrumentation, the use of fiber optic probes, or using some smart sampling accessories, multiple spectra from a single sample can be collected rapidly, averaged, and used as a composite spectrum that is representative of the entire sample, even in the presence of heterogeneity.

One of the major hurdles for a new practitioner to the NIR technique is the development of robust calibration models. Calibration models are typically developed using multivariate analysis (MVA), also known as chemometric techniques. There are no real shortcuts to model development, and this requires excellent subject matter knowledge of the application at hand and a good working knowledge of the methods used for sampling and model development. This is covered in more detail in Section 71.5 of this chapter and a dedicated chapter on chemometrics (see Chapter 65). The good news is that once a calibration model has been developed, the end user of the method does not need to have a working knowledge of chemometrics. As long as they follow standard procedures of sampling and instrument usage, nonskilled workers can use the technology in their day-to-day tasks without the supervision of an expert.

The NIR method is highly sensitive to moisture content in samples, and in some cases, when the moisture content is too high, the detector can saturate quickly. NIR is therefore not a suitable method for analyzing highly aqueous solutions, although there are some applications where this is possible [4]. Methods such as Raman spectroscopy may be better suited for these applications, but for such a small limitation, the NIR technique is highly versatile for the many other applications that exist.

This chapter provides a concise overview of the NIR methods from basic theory right through to end user applications and their development. It should be used as a first reference point, and the literature cited in this chapter should be used if more detailed explanations are required.

71.1.2 A Short History of NIR

The discovery of the NIR region of the electromagnetic spectrum has been attributed to Frederick William Herschel [5]. Using a simple apparatus consisting of a prism and a thermometer, Herschel was performing experiments to disperse sunlight into its component colors in the visible region of the spectrum. He observed that toward the red part of the visible region, the temperature of the dispersed radiation increased, and just beyond the visible region, where the radiation became invisible, the temperature reached its maximum. Herschel attributed his finding to be a thermal band beyond red, which is termed infrared (IR) from the Latin word “infra,” and considered this region to be different from light [6]. This principle is shown diagrammatically in Figure 71.1.

It was not until Ampere’s experiment using a newly developed thermocouple that NIR was determined to have similar optical characteristics to that of visible radiation [7].

Sometime later, Maxwell, Planck, and some of the great scientists of the nineteenth and twentieth centuries developed theories that better understood the nature of the electromagnetic spectrum, and today, we understand the NIR region to lie between

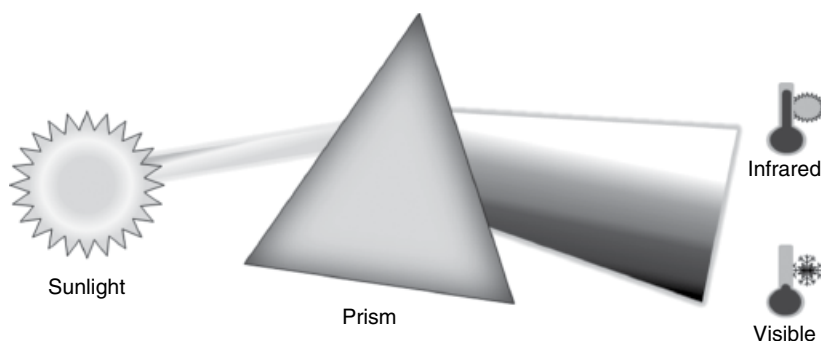


FIGURE 71.1 Dispersion of polychromatic light into its components using a prism. Toward the higher wavelength region of the visible spectrum and into the NIR region, the temperature of the radiation increases.

Visible	Near-infrared		Mid-infrared
	Short wave NIR	Long wave NIR	
400–700 nm	700–1,100 nm	1,100–2,500 nm	2,500–16,000 nm

FIGURE 71.2 The NIR region of the electromagnetic spectrum with respect to the visible and MIR regions.

the visible and MIR regions [8]. Figure 71.2 provides a diagram of where the NIR region of the spectrum lies with respect to the visible and the MIR region.

Coblentz [9] developed a primitive spectrometer for measuring the NIR spectrum of a number of materials and thus provided a means for chemists to elucidate the structure of compounds. For a more detailed discussion of the early history of NIR and the first instrumentation used to collect IR spectra, the interested reader is directed to the work of Burns and Ciurczak [6].

For a long period of time, the use of NIR remained relatively limited with few applications being published. It was not until the 1960s that the pioneering works of Ben-Gera and Norris of the US Department of Agriculture (USDA) paved the way for the modern success of NIR [10]. This initial work focused on the analysis of fat and moisture content in agricultural products and is still cited as one of the most influential references in NIR applications [11].

Agricultural and food applications dominated the NIR literature throughout the 1970s and 1980s with workers such as Williams [12], Osborne [13], Shenk [14], Flinn [15], Blakeney [16], and Batten [17] all making significant contributions. Agricultural applications of NIR are discussed in more detail in Section 71.6.1.

Up until the late 1990s, instrumentation remained the key limitation for the widespread use of NIR in the process industries, particularly for real-time monitoring. Swarbrick [18] reviews the evolution of instrumentation since 2000 and discusses how these advances in

instrument speed and portability have allowed NIR to be utilized in process applications, particularly in the pharmaceutical sector. Section 71.3 discusses the various instrumentation types available for generating NIR spectra.

In the pharmaceutical sector, the work by Ciurczak and Drennen [6], Mark [19], and Ritchie [20, 21] has paved the way for the modern usage of NIR for applications such as raw material identification, intact tablet analysis, monitoring of drying operations, and recently the monitoring of bioprocesses. A more in-depth discussion of NIR applied to pharmaceutical and biotechnology applications is provided in Section 71.6.2.

Other major applications of the NIR method can be found in industries such as petrochemical, wool, wood, dairy, and textile industries. As instrumentation and education in chemometrics improve over the coming years, there will be many more applications of NIR published due to its versatility and ruggedness. This chapter can only provide a brief outline of the myriad of applications where NIR is currently used, and the interested reader is encouraged to investigate the many forums that exist in the NIR community and to attend the two major conferences dedicated to the use and promotion of the NIR method.

In particular, the International Conference on NIR Spectroscopy (ICNIRS) [22] is a biannual meeting held globally that brings many of the leading practitioners together to present their work and discuss their ideas. During this meeting, the leading researcher in NIR spectroscopy (as nominated by their peers) is presented with the Tomas Hirschfeld award. Tomas Hirschfeld was an internationally recognized analytical scientist and inventor with over 100 patents to his name [23]. Tomas envisioned that research into microsensors and microinstruments would be the future and very rewarding to those who invested into this area [24]. After his unprecedented death in 1986, the ICNIRS inaugurated the Tomas Hirschfeld award for recognition of a significant contribution to the science of NIR spectroscopy, including research and development of new technology. Tomas' legacy is still recognized to this day as a key contributor to this important area of science and technology.

Every second year, in between the ICNIRS meetings, the Australian Near Infrared Spectroscopy Group (ANISG) [25] is held in Australia and New Zealand. A smaller conference with respect to the ICNIRS, the ANISG attracts a quality audience that come together to share ideas and stories regarding the development and application of NIR spectroscopy in a wide range of applications.

Overall, it is again stated that NIR is a mature technology, utilized in many research and industrial applications. The advancements in instrumentation, particularly since 2000, have enabled this technology to be implemented into many business critical and novel applications. The next section provides details on the theory behind the NIR method and how its characteristics allow it to have the versatility it has as an analytical, classification, and process monitoring tool.

71.2 THE THEORY BEHIND NIR SPECTROSCOPY

71.2.1 IR Radiation

The IR region of the electromagnetic spectrum is located between the visible and microwave regions (700–111,000 nm; refer to Fig. 71.2). It is associated with stretching and bending modes that occur prevalently within covalent bonds of organic molecules; however this definition extends to other chemical and physical bond types, including hydrogen bonding and other molecular interactions.

The IR region of the spectrum can be further broken down into three subregions:

1. The far infrared (FIR): This region lies between 16,000 and 111,000 nm primarily used for rotational spectroscopy and is associated with the measurement of inorganic materials and applications in astrophysics.
2. The MIR: This region traditionally has been used due to the large amount of chemical information related to molecular structure present in the sharp spectral features produced. This region lies between 2,500 and 16,000 nm.
3. The NIR: This region lies between the visible and MIR regions (700–2500 nm) and is associated with combination bands and overtones generated in the MIR region. Unlike the MIR, the spectral features of a typical NIR spectrum are broad and show much band overlap and require methods such as chemometrics to extract important information.

This chapter is concerned with the NIR region of the spectrum, and a more detailed description of the theory behind its interaction with matter is provided in the following sections.

71.2.2 The Mechanism of Interaction of NIR Radiation with Matter

When MIR radiation interacts with matter, depending on the frequencies, various molecular stretching and bending motions are induced in a molecule. The intensity of the spectral bands generated is a function of both the intensity of the incident radiation and the molar absorptivity of the chemical bond being excited. These excitations are known as the fundamental frequencies.

71.2.2.1 Overtone Frequencies Using the analogy of ringing a bell, when the bell is initially struck with a hammer (or other devices), the first sounds heard are loud and highly distinct in their audible frequencies. These are the fundamental notes of the bell sound. As the bell is left to vibrate, the sound intensity decreases rapidly, and when listened to closely, oscillating sounds can be heard after the fundamentals have decayed. These are known as the overtones of the fundamental frequencies and typically occur at integer values of the fundamental, that is, if the fundamental frequency occurs at a

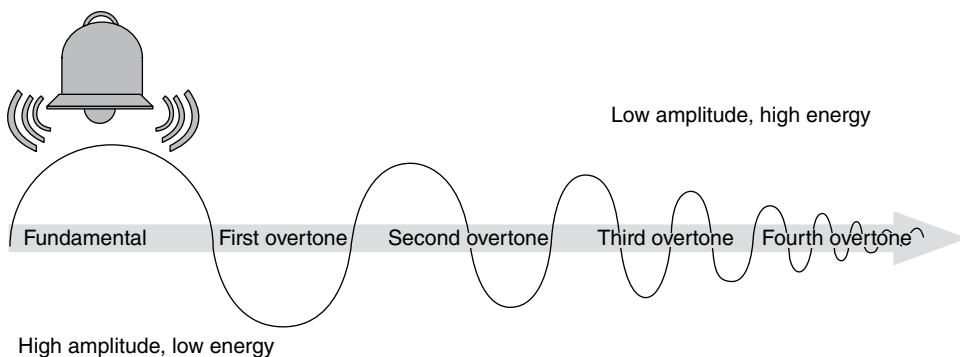


FIGURE 71.3 Analogy between ringing a bell and the overtones generated in the NIR region of the electromagnetic spectrum.

value f , theoretically, the first overtone will occur at $2f$, the second overtone at $3f$, etc. Figure 71.3 provides a diagrammatic representation of the fundamental and overtone frequencies using the bell analogy related to the NIR spectrum.

The earlier analogy of ringing a bell translates across to the MIR–NIR spectrum. When a molecular bond absorbs energy at a particular frequency and a bond stretching mode is induced in the molecule at fundamental frequency f , then at approximately $2f$, the first overtone band of the fundamental will occur in the NIR region of the spectrum. As with the case of the bell ringing, the overtone frequency has an intensity of approximately an order of magnitude less than the fundamental. Successive overtones diminish in intensity by an order of magnitude from the last overtone.

In general, these overtone frequencies have a low molar absorptivity coefficient (ϵ) with respect to the fundamental frequency, and this property lends itself extremely useful in practical implementations of NIR spectroscopy (see Section 71.6).

71.2.2.2 Combination Frequencies In general, stretching frequencies contribute mainly to overtone frequencies in NIR; however, there are some strong bending vibrations that also contribute [8]. The region of the NIR spectrum between 2000 and 2500nm is commonly known as the first overtone and combination band region. Combinations occur when the overtones generated in the MIR combine to form bands of higher intensity than would occur from the overtone alone, that is, they arise from the sharing of NIR energy between two and more fundamental absorptions.

The region of the NIR spectrum between 1100 and 2000nm is typically where the second overtones are located. There are also some combination bands that are strong enough in intensity to generate bands in this region. In the region between 700 and 1100nm, this is where third and fourth overtones occur. Their intensity is typically too small to reveal any combination bands of any practical use. Figure 71.8 in Section 71.2.3.2 provides an overview of the NIR region of the electromagnetic spectrum showing where combinations and overtones occur and also how the molar absorptivity varies across this region of the spectrum.

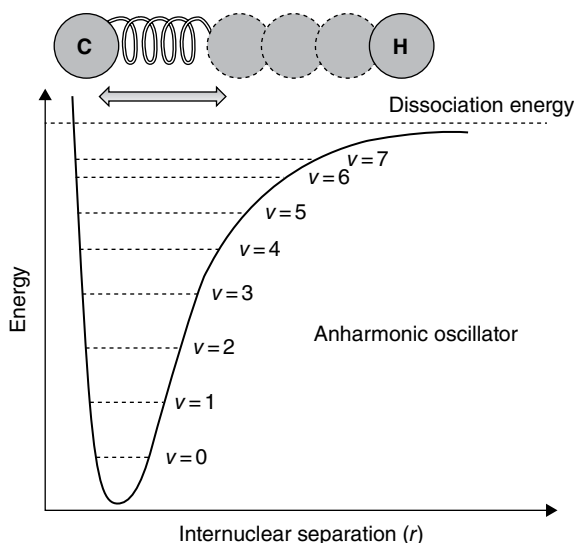


FIGURE 71.4 Description of molecular vibrations in terms of the anharmonic oscillator model.

71.2.2.3 Dipole Moments The atoms in a molecule can be considered as being held together by weak springs [8]. At their ground state, the molecules will naturally vibrate, and as more energy is applied (i.e., in the form of NIR energy in this case), the molecules will vibrate with a greater frequency. The energy levels of the vibrational states allowed are governed by the rules of quantum mechanics and in particular can be described as an anharmonic oscillator. Figure 71.4 shows this concept diagrammatically.

In the case of a two-atom molecule, only stretching of the bond between them can occur; for three or more atoms, bending of the bonds can also occur. Since the NIR region measures the combinations and fundamentals arising from the MIR region of the spectrum, the magnitude of the bands observed are 1–3 orders of magnitude lower than the fundamental bands. For molecular bonds containing C–H, O–H, and N–H bonds, there is a “large” difference in the atomic masses of carbon, oxygen, and nitrogen with respect to hydrogen. When NIR radiation is applied to molecules containing these functional groups, this sets up large dipole moment changes in the molecules, and the energy absorbed by them results in the bands observed in an NIR spectrum. Figure 71.5 provides a description of a dipole moment for the C–H bond.

Since the functional groups have different affinities (attraction) to each other, the dipole moments vary in magnitude. This fact allows for a distinction of bond type in the IR region of the spectrum, and this observation was first made by Coblenz [9] in his pioneering work on structural elucidation of molecules. For a more theoretical discussion on the theory behind NIR, the interested reader is referred to the literature [26].

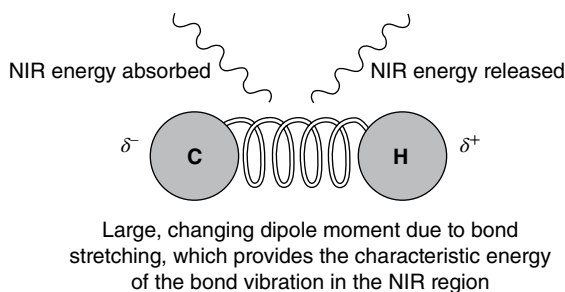


FIGURE 71.5 Changes in dipole moments between two atoms of largely differing atomic mass result in the bands typically observed in the NIR region of the electromagnetic spectrum.

71.2.3 Absorbance Spectra

71.2.3.1 Calculating an Absorbance Spectrum

A brief overview of the theory behind NIR was presented in the previous sections. This section puts the theory into practice by describing how an NIR spectrum is collected. Section 71.3 describes the instrumentation used to collect NIR spectra, and Section 71.4 discusses the sampling methods commonly employed. An absorbance spectrum is calculated as the negative logarithm of the reflectance or transmittance of NIR light from a sample. The concepts of reflectance and transmission are discussed in more detail in Sections 71.4.1 and 71.4.2. The formal calculation of absorbance is provided as follows:

$$A_i = -\log\left(\frac{1}{X_i}\right)$$

where

A_i = absorbance calculated for the i^{th} wavelength of the spectrum.

X_i = the reflectance or transmittance of a sample at the i^{th} wavelength.

Absorbance is measured on a unitless scale and is represented by the symbol AU (absorbance units). Since the scale is logarithmic, each AU represents an order of magnitude less light intensity than the incident light source. Figure 71.6 shows this principle diagrammatically.

For example, an absorbance value of 1 represents 10 times less light incident on the detector compared to the original light source, and an absorbance value of 5 represents 100,000 times less light detected compared to the original light source. A major advantage of NIR spectrometers is the detection systems used have high S/N ratios and in some cases can generate reliable spectrum up to 6 AU. For more details of the detectors used in NIR spectroscopy, refer to Section 71.3.9.

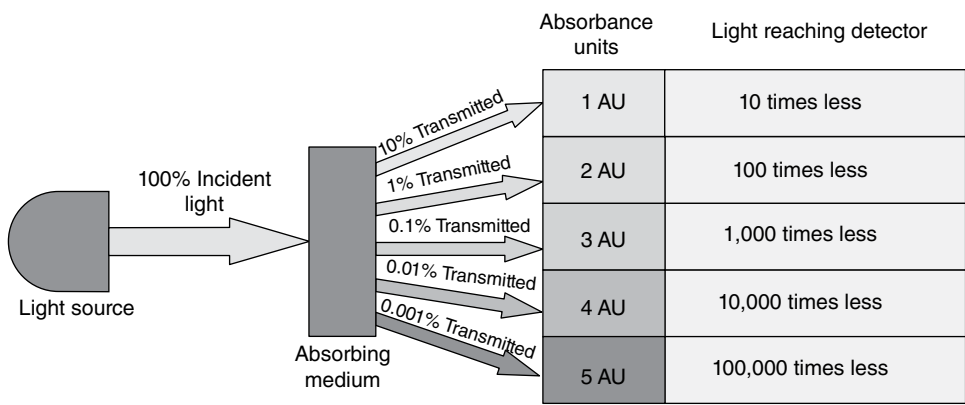


FIGURE 71.6 The relationship between light absorbed by a sample and the light intensity that reaches a detector.

A general process for collecting a spectrum on an NIR instrument typically proceeds as follows:

1. Collect a dark current (DC) spectrum. This is usually collected where the light source of the instrument is turned off and the electronic noise of the detector is measured for all wavelengths.
2. Collect a reference (Ref) spectrum. In transmission mode (Section 71.4.1) this is the signal from the light source unhindered by any sample, and for reflectance mode, the light source is reflected off a standard highly reflective material measured for all wavelengths (typically a material called Spectralon®).
3. Collect a sample (Sam) spectrum by placing the sample in front of the light source and collecting the light either transmitted through or reflected off the sample.

The raw reflectance or transmittance scan is calculated as follows:

$$X_i = \frac{(\text{Sam}_i - \text{DC}_i)}{(\text{Ref}_i - \text{DC}_i)}$$

In the case of transmission, $X_i = T_i$ where T_i is the transmittance at the i th wavelength, and for the case of reflectance, $X_i = R_i$ where R_i is the reflectance at the i th wavelength. Figure 71.7 provides examples of spectra collected in transmission and diffuse reflectance (Section 71.4.2) mode.

71.2.3.2 Characteristics of NIR Spectra The NIR region of the electromagnetic spectrum is split into two general subregions:

1. The short wave (SW) region (also known as the Herschel region) spans the wavelength range of 700–1000 nm [27]. This region is predominantly used for diffuse transmission measurements.

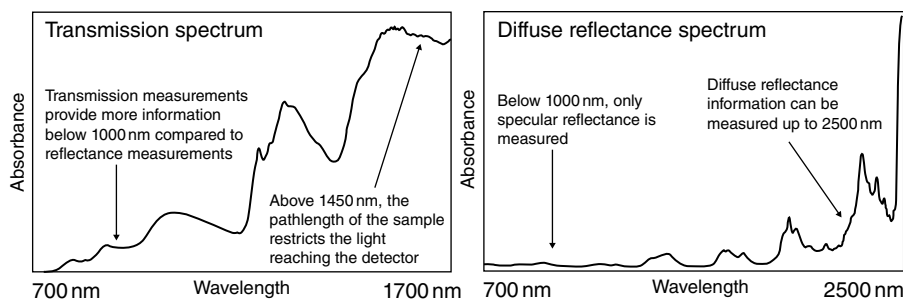


FIGURE 71.7 Examples of NIR spectra collected in transmission and diffuse reflectance modes.

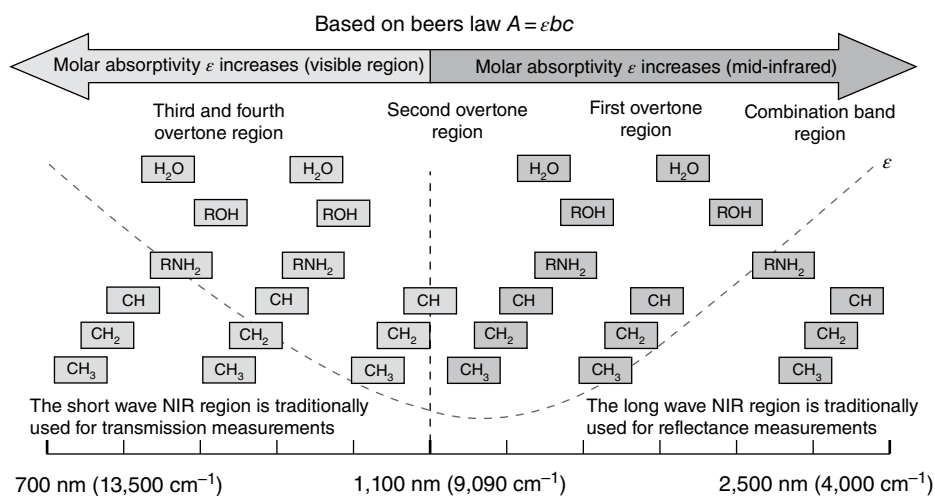


FIGURE 71.8 Characteristics and chemical information available in the NIR region of the electromagnetic spectrum.

2. The long wave (LW) region spans the region from 1000 to 2500 nm. This region is predominantly used for diffuse reflectance measurements.

Figure 71.8 shows the various regions of the NIR spectrum.

In particular, the region between 850 and 1100 nm has a very low molar absorptivity coefficient (ϵ) associated with it. Based on the Beer–Lambert law, this means that more light can be passed through samples in this region, and this is why this region is ideal for transmission measurements. This region contains the third and fourth overtones, and although the signals generated in this region are 2–3 orders of magnitude smaller than the fundamental that they arise from, the combination of high pathlength and low molar absorptivity means that light can be passed through pathlengths up to 30 mm thick to generate high-quality spectra.

Between 1000 and 2500 nm this is where the LW region of the spectrum occurs. Up to 1500 nm transmission measurements are still possible; however, beyond 1500 nm, two mechanisms combine that make transmission measurements difficult:

1. The wavelength of the radiation becomes similar in size to the particles of the sample being measured.
2. The molar absorptivity coefficient increases rapidly as the MIR region is approached.

According to the Beer–Lambert law [28], absorbance is related to the concentration of active sample constituent via the relationship

$$A = \varepsilon bc$$

where

A = absorbance

ε = molar absorptivity coefficient

b = pathlength that the radiation has to pass through

c = concentration of the constituent in the sample

In clear, nonscattering liquid samples, the pathlength b of the sample can be controlled by fixing it with a cuvette. According to the Beer–Lambert law, if the pathlength is fixed and the measurement is performed at a single wavelength, absorbance is proportional to concentration; therefore a linear calibration model can be developed where concentration can be predicted from the sample absorbance (provided the absorbances lie in the range 0–1).

When this principle is extended to the LW NIR region, the fact that the wavelength of the radiation is similar in size to the particles being measured means that its interaction with the particles is more elastic than specular in nature. This phenomenon is known as diffuse reflectance. In both diffuse reflectance and diffuse transmission measurements, the pathlength that the radiation has to pass through is not constant and cannot be made constant due to the different ways solid samples pack each time they are prepared. As a result, typical diffuse reflectance NIR spectra have a quadratic baseline beyond 1500 nm (refer to Fig. 71.7).

In the section on preprocessing (Section 71.5), it is explained how additive and scatter correction algorithms can be used to correct for these spectral features caused by physical effects in the sample.

Overall, NIR spectra contain broad spectral features arising from overtone and combination bands that are highly overlapping. Until the rise of the personal computer and chemometric methods, the NIR region of the spectrum was thought (and taught) to contain little information useful for structural elucidation or any other application. Today, the number of applications is growing because of the development of fast and reliable instrumentation. The next section provides an overview of the instrumentation available for collecting NIR spectra and their modes of operation.

71.3 INSTRUMENTATION FOR NIR SPECTROSCOPY

This section provides a brief overview of the NIR instrumentation currently available at the time of writing of this chapter and is not meant to be an exhaustive description of each instrument type. The interested reader is referred to the referenced literature for more information regarding the theory and construction of each instrument type discussed.

71.3.1 General Configuration of Instrumentation

The development of instrumentation for NIR spectroscopic applications has seen a massive growth and improvement since the mid 1990s. The first instruments were based on specific filters (typically 10–20 individual wavelengths), which were chosen for a particular application; however, today's modern spectrometers can generate thousands of data points per spectrum. The following sections provide a discussion of the instruments currently available and their applicability to various applications.

In general, all spectrometers consist of the following components:

1. A suitable light source (typically a tungsten halogen lamp optimized for the IR region)
2. A suitable sampling device (see Section 71.4)
3. A light dispersion device (monochromator)
4. A detector

Figure 71.9 shows a generic spectrometer setup used in most instruments.

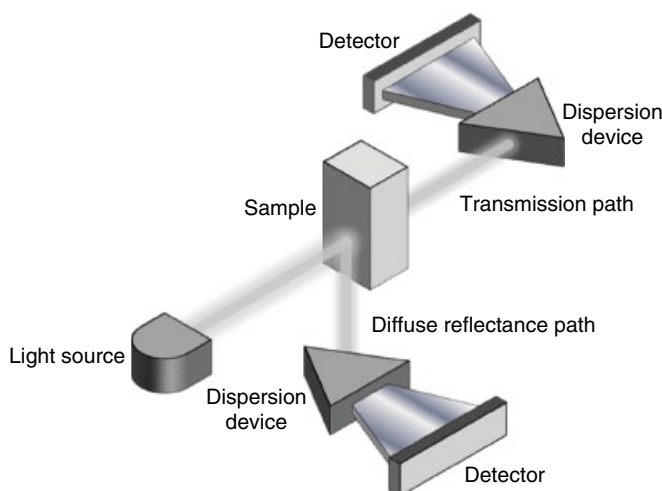


FIGURE 71.9 Generic configuration of a spectrometer used to collect NIR spectra.

71.3.1.1 Diffuse Reflectance Configuration Many of the modern NIR instruments today are configured to collect spectra in diffuse reflectance mode. Figure 71.10 provides a general layout of a typical diffuse reflectance setup.

In any spectrometer configuration, there are two main ways in which the instrument works:

1. Predispersive instruments place the monochromator just after the light source and pass the dispersed light through the sample. These systems have a single, larger surface area detector that synchronizes with the dispersion system such that a spectrum can be generated over many wavelengths.
2. Postdispersive instruments pass the light from the source through the sample and then send the reflected/transmitted light typically to a stationary monochromator, which disperses the light over a diode array detector, such that the entire spectrum is generated in one operation.

71.3.1.2 Transmission Configuration Transmission instruments are typically designed for two main applications in NIR:

1. Diffuse transmission, where an intense light source is passed through a solid/semisolid medium, and what light emerges from the sample is used to generate a spectrum.
2. Normal transmission, where the spectrometer acts as a typical instrument for collecting spectra through clear, nonturbid liquids (typically meeting the requirements of the Beer–Lambert law).

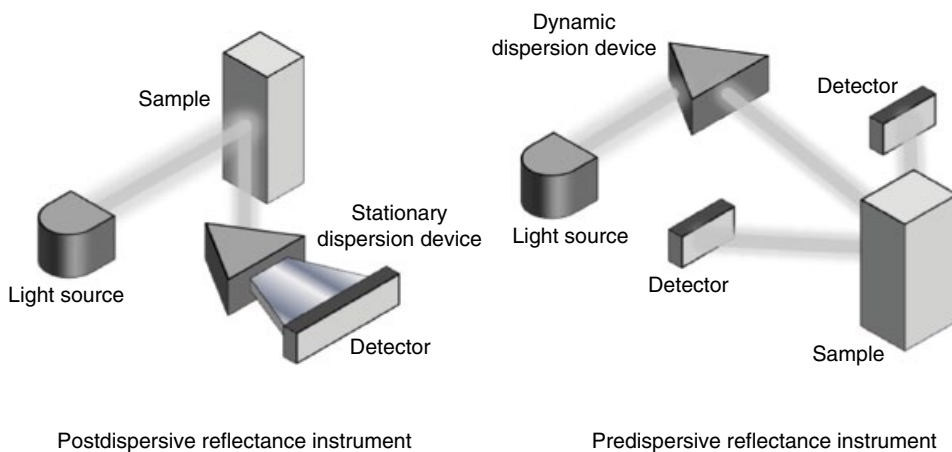


FIGURE 71.10 General NIR instrument configuration for collecting spectra in diffuse reflectance mode.

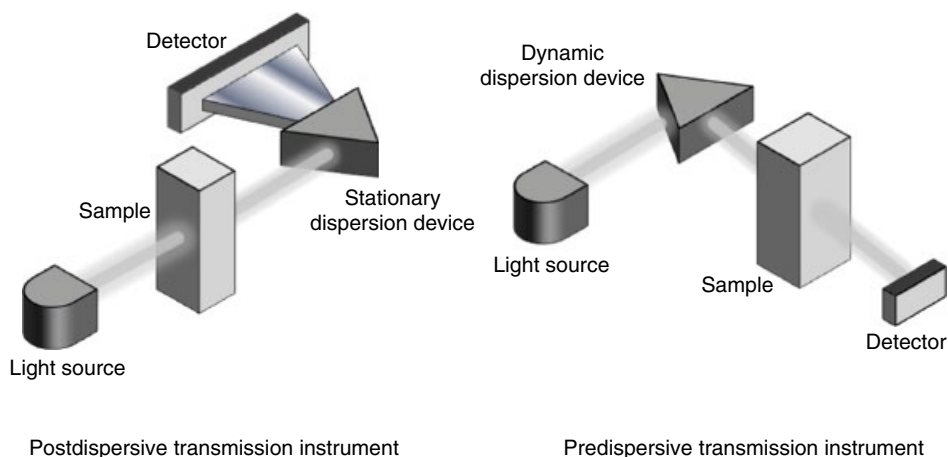


FIGURE 71.11 General NIR instrument configuration for collecting spectra in transmission mode.

Figure 71.11 provides a general layout of a typical transmission setup.

Although there are other sampling techniques available (such as interattance probes; Section 71.4.4.1), in general, these techniques are a variant of the two modes of collection described previously. These configurations are typically used with the instrument types described in the following sections.

71.3.2 Filter-Based Instruments

Figure 71.12 provides a schematic view of how a typical filter-based instrument works.

Light from the source is incident on a spinning filter wheel, which dispersed the light into the component wavelengths defined by the individual filters. A chopper was typically implemented to separate the wavelengths detected after the NIR radiation was passed through (or reflected off) a sample. In a filter instrument, a continuous spectrum cannot be generated; only a set of absorbance values are generated when the sample signal is divided by the reference material and the logarithm of the transmission/reflectance value is taken (see Section 71.2.3.1 for details on the calculation of an absorbance spectrum).

Filter instruments typically consist of a filter wheel with anywhere between 5 and 30 individual filters that represent different wavelengths in the NIR region. The main reason why filters were used in earlier instrumentation was the limitation that mathematical tools and fast computers are able to handle the data generated by modern spectrometers. The main method of analysis of the data generated by a filter instrument was multiple linear regression (MLR), discussed in Chapter 65.

The analyst would take the data collected on a predefined set of samples and use either a step-forward analysis, where wavelengths (filters) were added to a model to see if it improved the fit, or a step-backward approach, where all wavelengths (filters) are added and successively removed from the analysis until an optimal model could be generated.

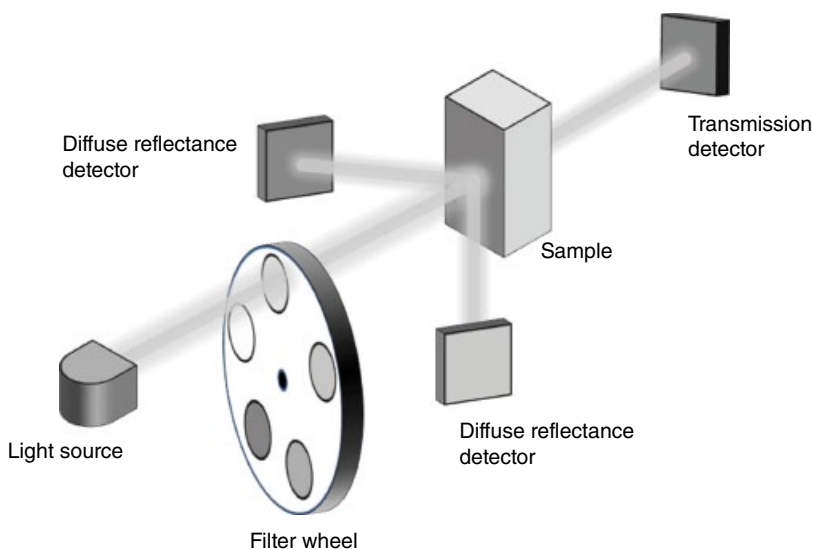


FIGURE 71.12 Filter-based NIR spectrometer.

Filter instruments still find practical use today as moisture analyzers in agricultural and pharmaceutical applications; however, their use is limited as newer and faster (and cheaper) instruments provide more versatility.

71.3.3 Holographic Grating-Based Instruments

A holographic grating for NIR applications is typically a highly polished concave glass surface with a thin coating of gold and blazed with a series of parallel lines that form a diffraction grating. Light incident onto the grating may be used in two ways:

1. Moving grating system, where the holographic grating is mounted onto a precise stepper motor (known as an encoder), and as the grating sweeps through an arc, defined by the encoder, the incident light is dispersed into its component wavelengths.
2. Stationary gratings are fixed in a predefined position such that the light is dispersed over an array detector for a specified wavelength region. This type of instrument is discussed in more detail in Section 71.3.4.

The moving grating system will be discussed in more detail in this section. These instruments can be configured to be either pre- or postdispersive, and as the grating moves through a single arc, the entire wavelength range of the spectrum is generated. Each component wavelength is passed onto the sample, and the detector is used to measure the signal. In order to generate a spectrum, the encoder is used to synchronize the wavelength measured to the detector signal.

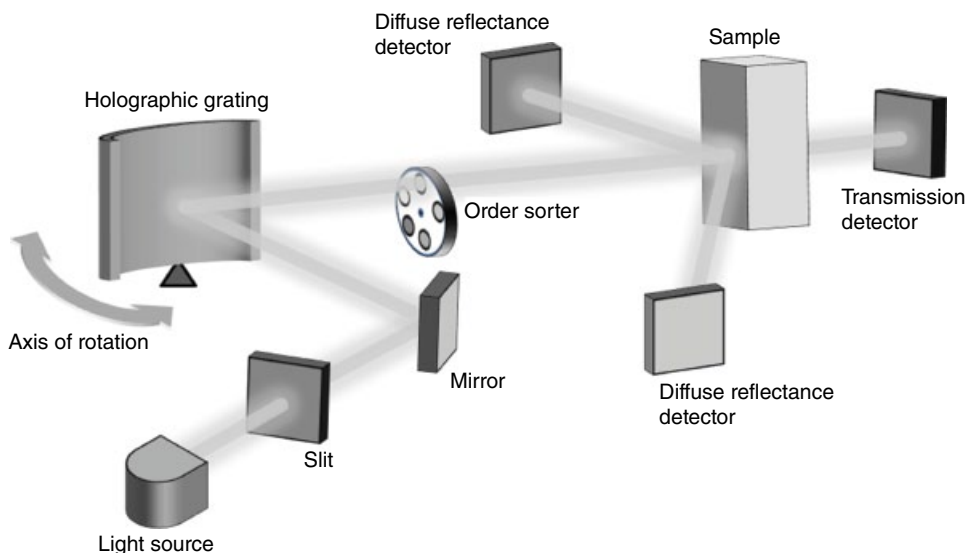


FIGURE 71.13 General instrument configuration of a predispersive holographic grating-based NIR spectrometer.

Holographic gratings also generate higher- and lower-order diffraction patterns [29] that must be eliminated in order to generate as pure a spectrum as possible. These instruments are also equipped with mechanical order sorters, which work in synchronization with the encoder and the detector. Thus, the holographic grating instruments can result in a highly mechanical system, which, when any service or repairs need to be performed, may lead to intensive recalibration of the instrument.

Figure 71.13 provides a schematic view of a typical predispersive grating-based spectrometer.

The spectral resolution of this spectrometer type is typically determined by the slit width of the instrument. There is a trade-off to be made here; if the slit width is too large, the resolution of the instrument decreases; however, more light can be passed onto the detector, thus increasing the S/N ratio. Conversely, if the slit width is made too small, very little light can reach the detector (particularly when diffuse transmission is used); therefore the S/N ratio is decreased; however, the resolution is increased.

Typical grating-based instruments use a 10 nm bandpass for collecting a spectrum. This means that the true resolution of the instrument can discriminate between spectral features separated by 10 nm distances. Spectral resolution must be distinguished from point to point resolution. Most spectrometers will generate spectra with anywhere from 5 to 0.5 nm point to point spacing. This is because the instrument internal software will interpolate in between the spectral resolutions of the instrument to generate a smoother spectrum.

This principle of optical efficiency is also applicable to diode array-based systems using a stationary grating as the slit width also determines the resolution. Also in this

case, the slit width is typically chosen to match the pitch of the detector spacing in the diode array detector.

In general, moving grating instruments are slow with respect to modern spectrometer setups that utilize no moving parts, and reliable spectrum generation usually requires the moving grating to complete 32 passes (coadds) to generate a single spectrum. Their mechanical nature requires them to be housed in rugged and environmentally controlled housings to minimize mechanical shock and temperature fluctuations. For this reason, this type of instrument finds limited use in real-time monitoring applications; however, for general laboratory and research applications, these instruments find many applications.

71.3.4 Stationary Spectrographic Instruments

In order to overcome the mechanical limitations of the moving grating instruments discussed in Section 71.3.3, a postdispersive configuration of the holographic grating monochromator can be used. Figure 71.14 shows a schematic of the general configuration of a fixed spectrograph instrument.

In this instrument type, incident light is passed through or reflected off a sample and passed onto the fixed monochromator. The position of the monochromator is fixed such that it disperses the incident light onto a diode array detector over the desired wavelength region of the spectrum. As with any grating-based instrument, spectral resolution is determined by the slit width. Since the detector is based on a diode array, the

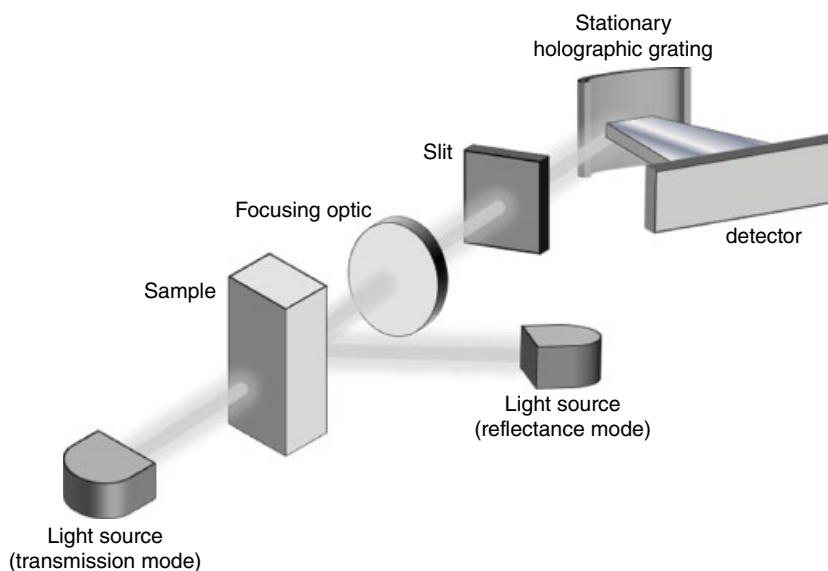


FIGURE 71.14 General instrument configuration of a spectrograph utilizing a diode array detection system.

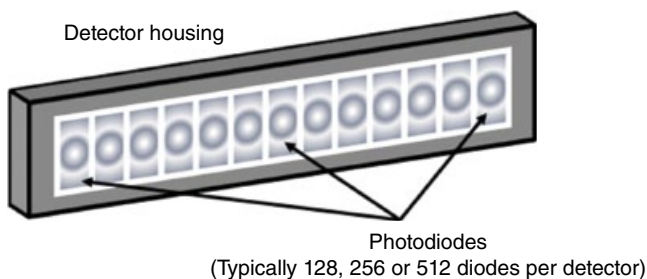


FIGURE 71.15 Diagrammatic representation of a diode array detector.

entire spectrum is collected in one step. This provides a major speed advantage over the moving grating systems. The same S/N issues exist, that is, if the slit width is too small, very little light will reach the monochromator; however, this can be improved by increasing the integration time, that is, the exposure time the light is incident on the detector for collecting a single scan. Therefore, like the moving grating instruments, spectra can be coadded to generate a spectrum with a higher S/N ratio, and since the typical integration time of a diode array detector is in the range of 10–100 ms, a high-quality spectrum can be generated in 3 s or less.

A generic diode array detector is shown diagrammatically in Figure 71.15. There are many different detector types that can be used, and these are discussed in more detail in Section 71.3.9.

Diode array detectors come in sizes as orders of the power of 2, that is, 2^k with the most common pixel numbers being 128, 256, and 512 arrays (although 1024 and 2048 systems do exist). It is a general rule that the pitch between the detectors is matched to the slit width; thus to take advantage of any resolution gains by using larger arrays, the slit width may be too restrictive for collecting high-quality spectra. If the slit and array are not matched, then cross talk between diode array pixels may result, meaning the same information is collected on consecutive pixels. Instrument vendors spend much time and resource into the development of instrumentation in order to minimize such inefficiencies [30].

Overall, stationary gratings coupled with diode array detectors offer a speed advantage over moving grating systems. Since they also contain no moving parts, they are essentially ruggedized for use in process applications. Early instruments of this type suffered from issues arising from overheating and drift; however, today's manufacturing processes and the introduction of more stable detector systems have seen a rise in the use of such systems in real-time applications.

71.3.5 Fourier Transform Instruments

For many years, the moving grating-based instruments dominated the NIR market particularly in agricultural applications and were seen as the gold standard of instrumentation. In the late 1990s–early 2000s the Fourier transform near infrared (FT-NIR)

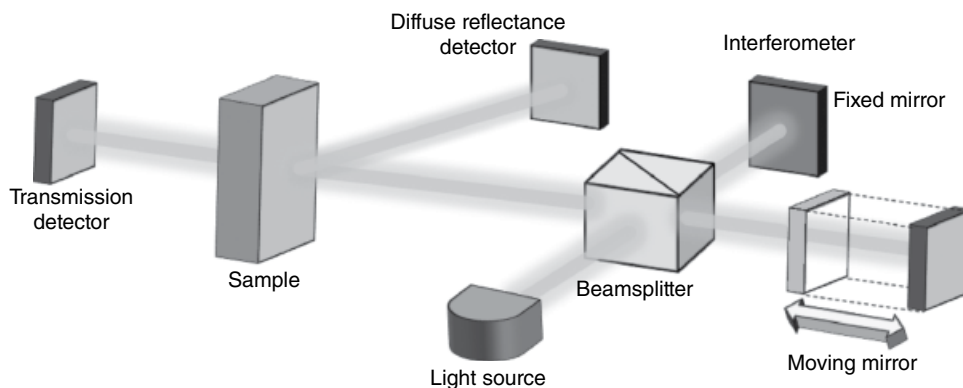


FIGURE 71.16 General instrument configuration of a Fourier transform NIR spectrometer utilizing a Michelson interferometer.

spectrometer was introduced as an alternative to the currently available grating-based instruments.

Fourier transform (FT) instruments are not slit width limited (Section 71.3.3) and therefore provide the advantage of allowing more light to be incident onto a sample or detector (known as the Jacquinot advantage [31]). FT-NIR instruments are typically predispersive, passing the incident light onto an interferometer (in particular, Michelson interferometers [32] are typically used). Figure 71.16 provides a schematic representation of an FT-NIR instrument that utilizes a Michelson interferometer.

The interferometer uses a combination of a fixed and moving mirror system to generate interference patterns. This packet of light patterns is passed onto a sample in either diffuse reflectance or transmission mode. Since the signal is sent to the sample in the time domain, by applying a fast Fourier transform (FFT) [33], the time-domain signal is transformed into a frequency-domain signal, and therefore an NIR spectrum can be generated. This is known as the Fellgett advantage [34]. The theory behind the FT is outside of the scope of this book, and the interested reader is referred to the literature for more details [33].

The accuracy of the wavelength scale is determined by an onboard laser source in the instrument (also known as the Connes advantage [35]). The position and frequency of the laser pulse allow FT-NIR instruments to be set to any spectral resolution desired by the developer. Since the wavelength accuracy is determined by a laser, the resolution of an FT-NIR spectrometer is a true spectral resolution. Typical resolutions that are used in FT-NIR applications range from 8 to 64 cm^{-1} depending on the application. It is noted here that resolution in an FT-NIR spectrometer is linear in the cm^{-1} scale and nonlinear in the nm scale. It is therefore very important to be aware of this fact when comparing results from an FT instrument to a grating-based instrument (which is linear in the nm scale).

In practical terms, the major benefit of the FT systems over the moving grating systems is the minimization of moving parts in the FT system. The so-called resolution

advantage is not really an advantage unless a user is interested in water vapor analysis (or similar applications). For most practical industrial applications, using a resolution of 16 cm^{-1} typically results in good calibration models. When the resolution is improved to below 16 cm^{-1} , the size of the spectral files increases and the inherent noise (shot noise) [30] of the spectrum increases since the instrument is using true spectral resolution and the final signal is not interpolated and smoothed. In terms of time benefits, both FT and grating instruments collect spectra at a similar frequency, and FT instruments also require considerable coadding to generate a high-quality spectrum.

In terms of real-time process applications, FT-NIR instruments offer the following major advantages compared to moving grating-based instruments:

1. They are more rugged by design and require less environmental protection.
2. FT instruments do not suffer from Wood's anomaly [36], a spectral inconsistency related to diffraction gratings, and it is known as passing of orders. These result in spectral features that inconsistently arise (particularly when moving samples are measured) and can result in false predictions if the anomaly becomes too intense.
3. Since the FT instruments can be manufactured in a more consistent manner than grating-based instruments, calibration transfer is easier with FT instruments.

Other than the previously mentioned advantages, FT and moving grating instruments typically have similar performance characteristics. Although preferred over moving grating instruments for process applications, FT instruments are still relatively slow compared to more modern NIR spectrometer designs and are now finding many applications in the analysis of processes that slowly evolve over time or for research applications.

At the time of the writing of this chapter, a new generation of micro-FT-NIR spectrometers has been developed [37]. Based on a continuous manufacturing (CM) process, the entire spectrometer including lamps has been integrated onto a single board with reported interferometer frequencies of 400 Hz. It will be interesting to monitor the progress and performance of such instruments in the future.

71.3.6 Acoustooptical Tunable Filter Instruments

Acoustooptical tunable filter (AOTF) spectrometers were first proposed by Harris and Wallace in 1969 [38]. They utilize a crystal of tellurium dioxide (TeO_2) as the light dispersion device, and through the application of specific radio frequencies (RF), the crystal acts like a tunable bandpass filter, therefore the name acoustooptical. Figure 71.17 provides a schematic of a generic AOTF instrument.

When RF waves are passed through the crystal, the lattice successively compresses and relaxes, and the resulting effect is similar to a transmission or Bragg refractor [39], the main difference being that the AOTF device only emits one wavelength band at a

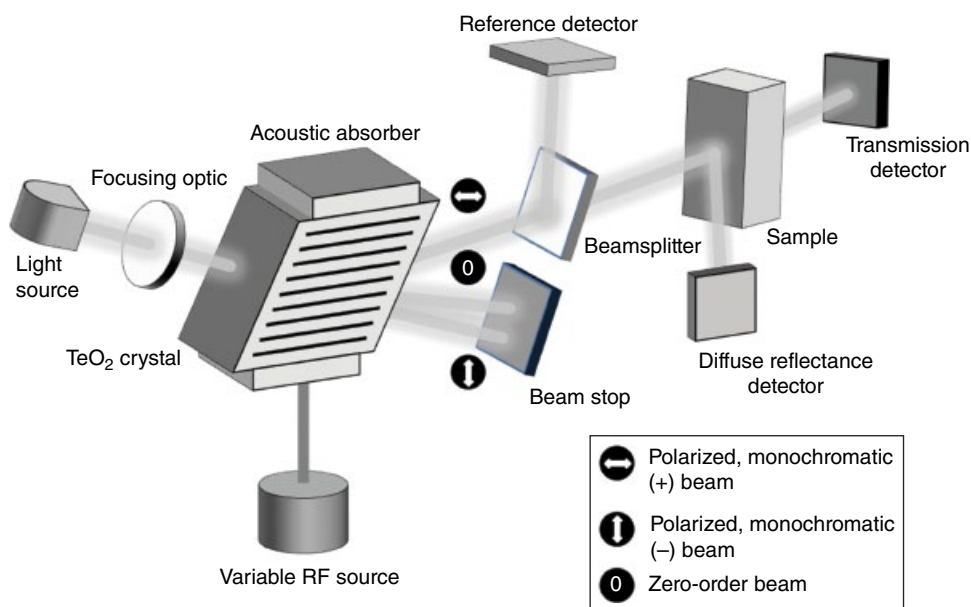


FIGURE 71.17 General instrument configuration of an acoustooptical tunable filter (AOTF) NIR spectrometer.

time. As is shown in Figure 71.17, the light emitted from the AOTF crystal is two orthogonally polarized beams. To use the AOTF as an NIR spectrometer, a beam stop is used to exclude all but the desired monochromatic light beam that is directed toward the sample.

A major advantage of the AOTF spectrometer over other spectrometers is the ability to precisely and rapidly tune the system to a specific wavelength. This is useful when a single absorbance band has been found to monitor a process in real time. Since the RF can be tuned in a matter of microseconds, the collection of an entire NIR spectrum is possible in a short time period.

AOTF instruments find widespread usage in high-speed monitoring applications in the pharmaceutical and agricultural sectors. They are most applicable for industrial applications due to their speed, ruggedness, and no moving parts.

71.3.7 Microelectromechanical Spectrometers

Microelectromechanical spectrometers (MEMS) are mass-produced silicon chip-based devices, which consist of a tunable Fabry–Perot filter [40]. The manufacturing process builds various substrate layers onto the surface of the chip, and the result is a microelectromechanical device consisting of two dielectric mirrors, facing parallel to each other and separated by a small distance from each other. The application of precise voltage levels to the device changes the distance between the two mirrors, thus allowing

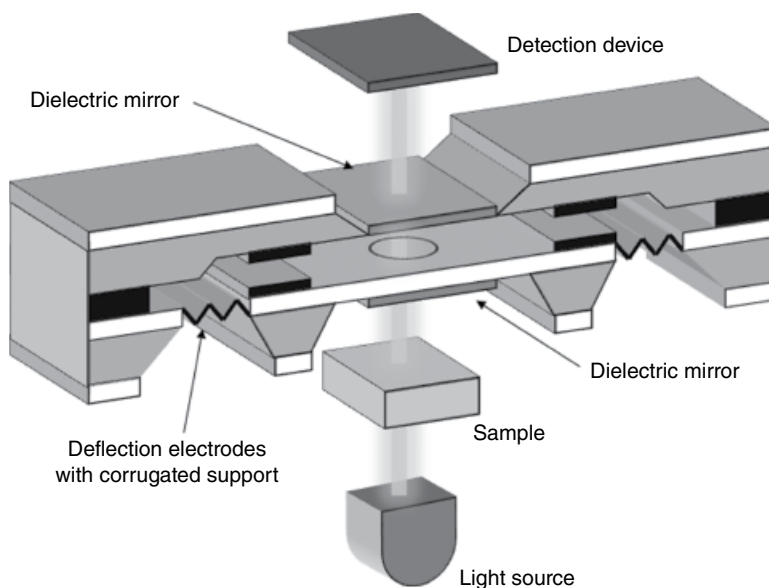


FIGURE 71.18 General configuration of a microelectromechanical spectrometer (MEMS).

the device to be tuned to specific wavelength bands. Figure 71.18 provides a schematic representation of the light dispersion devices for MEMS.

The result is a fast tunable bandpass filter capable of recording a complete spectrum in 50 μ s. MEMS are typically used in small, handheld devices due to their size, and for more information on this interesting class of spectrometer, the interested reader is referred to the literature [30].

71.3.8 Linear Variable Filter Instruments

In recent times, miniaturization of instrumentation has taken a massive leap forward with the introduction of stable micro-NIR spectrometers, which utilize Linear Variable Filters (LVFs). It is interesting to note that the early work on NIR was performed by Professor Karl Norris using instruments based on LVFs [41], and now this technology is making its renaissance in NIR technology.

An LVF is a bandpass filter coating intentionally wedged in one direction. As a result of the varying film thickness, the wavelength transmitted through the filter varies linearly in the direction of the wedge. The LVF is coupled to a linear detector array using a semiconductor manufacturing process such that the entire LVF and detector are seamlessly joined together [42]. Figure 71.19 provides a schematic overview of a currently available LVF device.

As with the new FT instruments discussed in Section 71.3.5, the LVF instruments can be manufactured as a complete unit from highly repeatable manufacturing processes. Much time and investment have been made into the development of these

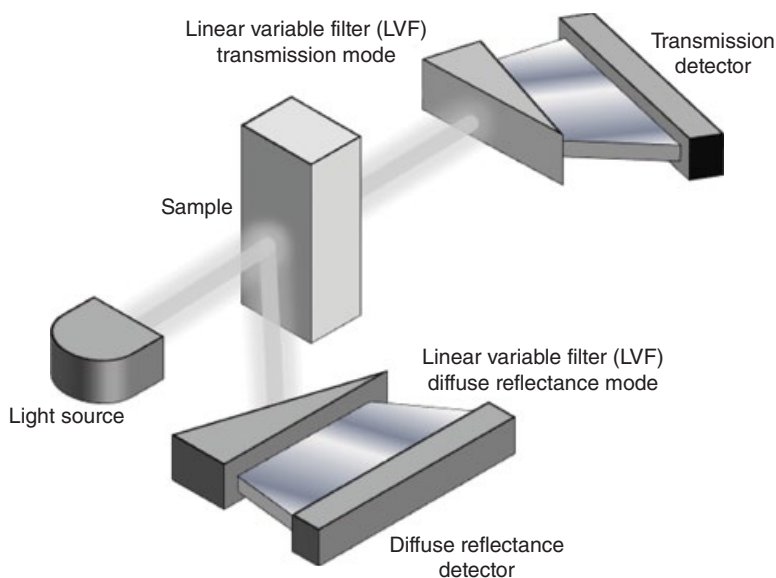


FIGURE 71.19 General configuration of a micro-NIR device utilizing a Linear Variable Filter (LVF).

systems in recent times, and they will form the cornerstone of many process applications, particularly in process analytical technology (PAT) applications moving into the future (refer to Section 71.6.2.3 for more details).

The resolution of these instruments is generally between 0.5 and 1% of the center wavelength of measurement, and they can be configured to cover a wide wavelength range (depending on the detector system fused to the LVF). Scanning times are in the order of 1–2 s, depending on the integration time and scan count used to collect a spectrum. They contain no moving parts at all and are highly ruggedized to be used in harsh conditions.

71.3.9 A Brief Overview of Detectors Used for NIR Spectroscopy

To this point, the discussion has been primarily focused on the dispersion devices that characterize an instrument type. It is, however, the combination of dispersion device and detector type that determines the wavelength region covered by the final instrument.

In Section 71.2.3.2, Figure 71.8 shows the regions of the NIR spectrum and the chemical phenomena measured. Correspondingly suitable detector types must be matched to the region being measured. In particular, only three detector types will be discussed in this section:

1. Silicon (Si)
2. Lead sulfide (PbS)
3. Indium gallium arsenide (InGaAs)

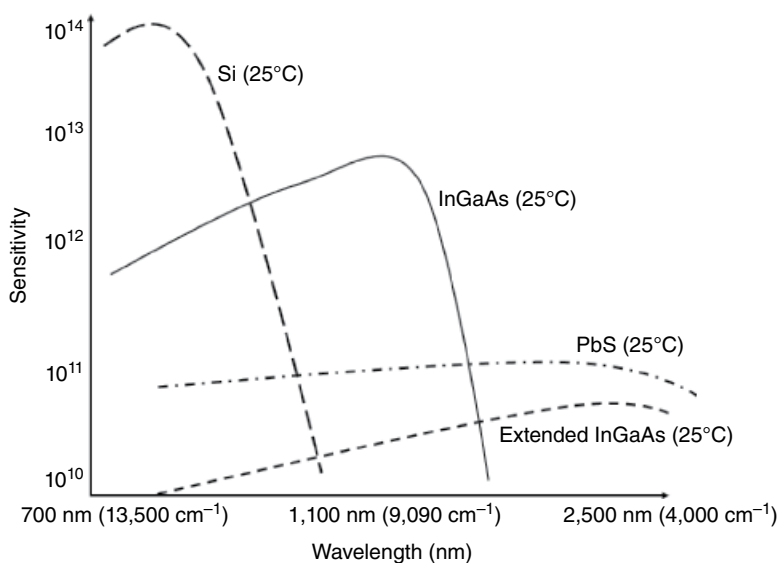


FIGURE 71.20 Comparison of detector sensitivity and wavelength coverage in the NIR region of the electromagnetic spectrum.

Figure 71.20 provides an overview of the detector sensitivities over the NIR region of the spectrum, and a brief discussion of the detector applicability is provided in the following subsections.

71.3.9.1 Silicon Detectors Silicon (Si) detectors are used primarily in the SW NIR region (refer to Section 71.2.3.1). In this region, the third and fourth overtones are measured, and this requires a detector with the high sensitivity characteristics of the Si detector (refer to Figure 71.20). In this region of the spectrum, specular reflectance phenomena (i.e., purely light reflection phenomena) dominate, and therefore, diffuse reflectance methods provide little chemical information of the sample being measured. Therefore, this region of the spectrum is mainly used for transmission measurements, and the Si detector is capable of measuring the small light levels (typically 3–5 AU), which typically emanate from passing light through densely packed solid materials such as grain or pharmaceutical tablets. The Si detector has a maximum sensitivity at 950 nm and has an effective working range between 350 and 1100 nm.

71.3.9.2 Lead Sulfide Detectors Lead sulfide (PbS) detectors are primarily used in the LW region of the NIR spectrum. Typically used for diffuse reflectance applications, the PbS detector has a maximum sensitivity at 2500 nm and an effective working range between 1100 and 2500 nm (actually, the PbS detector can also be used in parts of the MIR).

Some grating-based instruments have utilized a combination of Si and PbS detectors to create an instrument capable of measuring between 700 and 2500 nm. The issue

with this type of instrument configuration is that at the junction between the two detectors at 1100 nm, an inconsistent join is visible in the generated spectrum. This junction has to be removed when developing predictive models.

71.3.9.3 Indium Gallium Arsenide Detectors The indium gallium arsenide (InGaAs) detector is a midpoint between the Si and PbS detectors. The normal InGaAs detector has a maximum sensitivity comparable to the Si detectors with a maximum sensitivity around 1400 nm and an effective working range between 900 and 1700 nm. An extended version of the detector is available by doping the InGaAs with phosphorous. This provides a detector with a wavelength range between 1200 and 2600 nm with a maximum sensitivity around 2000 nm; however, the doping process reduces the sensitivity of the detector with respect to the nondoped version.

InGaAs detectors are the most widely used detectors in NIR instruments since they can cover the 900–1700 nm range where a majority of NIR information can be found, and also it allows some limited applicability to both transmission and reflectance applications.

71.3.10 Summary

This section summarized the many types of NIR instrument available on the market. Each spectrometer type has its advantages and disadvantages with respect to performance. Instruments can be loosely characterized into two categories:

1. Research-grade instruments
2. Portable instruments

Research-grade instruments are typically slower in performance (requiring between 10 and 30 s to collect a spectrum); however, they are highly stable, provide multiple sampling options to be utilized by a single spectrometer, and are highly useful for method development purposes before deploying a smaller, portable system into a real-time process application.

Portable instruments utilize fast scanning dispersion devices, which are typically lower in spectral resolution compared to research-grade instruments; however their speed characteristics more than make up for the resolution deficiencies. The smaller systems are typically used in handheld devices or in process applications where space is limited and where the implementation of a larger system would be excessive in terms of system build and budgetary requirements.

Section 71.4 provides details on the various sampling modes used for the instrument types defined in this section, and Section 71.6 provides example applications where these instruments have found to be highly valuable for business critical operations.

71.4 MODES OF SPECTRAL COLLECTION AND SAMPLE PREPARATION IN NIR SPECTROSCOPY

The low molar absorptivity constant and the high S/N ratio associated with instrument detectors provide NIR with a number of key advantages over many other spectroscopic methods when used for on-line, in-line, and at-line applications. Before continuing with a discussion of spectral collection and sampling, a brief explanation of the general types of application are provided:

- At-line data collection: In this mode, samples are physically taken from the process/bulk material and presented to the instrument for analysis in a remote laboratory.
- On-line data collection: A side port or other sampling systems are used to divert a sample from the main process into a cell (or other devices) for analysis by the spectrometer, with no user intervention required. Typically the sample is not returned to the process after analysis.
- In-line data collection: The sampling system, for example, fiber optic probe, is inserted directly into the process, and samples are measured as they exist in the process.

Based on the earlier definitions, there are a number of ways a sample can be presented to an NIR spectrometer for analysis. As discussed in Sections 71.3.1.1 and 71.3.1.2, to generate an NIR spectrum, the major modes of spectral collection are either transmission or reflectance.

71.4.1 Transmission Mode

In transmission mode, NIR radiation is passed through a sample, usually in the state it exists in, and the light that is not absorbed is collected on the detector for each wavelength scanned and presented as either a transmission or absorbance spectrum (Section 71.2.3). Figure 71.9 provides a diagrammatical overview of the transmission process.

To collect a spectrum in transmission mode, incident light from an NIR source is passed onto the dispersion element of the spectrometer (refer to Section 71.3 for dispersion devices used in NIR spectroscopy), and this serves as the background (or reference) spectrum (I_0). The sample is then presented to the sampling device where light is passed through the sample. This generates a signal based on the light that is transmitted (i.e., not absorbed) by the sample for each wavelength (I). The ratio of I/I_0 for each wavelength generates the transmission spectrum.

Taking the negative logarithm of the transmission values at each wavelength generates the absorbance spectrum. In many cases, the absorbance spectrum is considered the first preprocessing of the data as the logarithm acts to linearize the data before

analysis. Section 71.5 provides more details of the preprocessing methods commonly used for NIR spectroscopy. Depending on the physical nature of the sample, this dictates the wavelength region used to collect a transmission spectrum.

71.4.1.1 Normal Transmission By normal, it is meant that the sample being analyzed is a clear, homogeneous liquid sample contained in a cuvette or sampling cell where the effects of scatter have been minimized. The low molar absorptivity coefficient works into the favor of the method developer and allows the optimization of pathlengths to obtain the best spectral quality for the application.

For example, NIR is highly specific (see Fig. 71.8, band assignment table) to moisture, and if the pathlength is too large, detector saturation at higher wavelengths, that is, greater than 1500 nm, may result. In the region between 900 and 1000 nm, moisture absorbs relatively unhindered by other molecular groups. Therefore the combination of low absorptivity and high pathlength can be used to measure larger sample volumes, using a lower-cost spectrometer setup and providing an effective means to monitor a business critical operation using a simple, yet robust setup.

Using another example, the region of the NIR spectrum between 1200 and 2000 nm lends itself very well to the analysis of octane number and other properties in gasoline using transmission. In particular, NIR is sensitive to aromatic C—H absorbances and can be used to distinguish between aromatic components and straight chain hydrocarbons. In this application, the pathlength of the cuvette or cell is controlled to avoid detector saturation due to increasing molar absorptivity coefficients.

71.4.1.2 Diffuse Transmission To reiterate the importance of low molar absorptivity coefficient in the NIR region, the short wavelength region is most useful for diffuse transmission. The difference between normal transmission (discussed in Section 71.4.1.1) and diffuse transmission (discussed in Section 71.4.1.2) is that the sample is typically solid or a highly particulate suspension/emulsion that absorbs and scatters most of the NIR radiation before it reaches the detector. In sectors such as the agricultural industry (Section 71.6.1), grain traders use NIR to measure commodity prices of the grains either during harvest or after blending. This requires that the NIR radiation passes through up to 30 mm of solid material in order to generate a reliable spectrum.

The absorbance values associated with diffuse transmission are pathlength and material dependent but often are in the 3–5 AU range. Reliable spectra can be recorded in some cases above 5 AU owing to the high S/N ratios of the instrument detector system (see Section 71.3.9).

Another application of diffuse transmission is the on-line measurement of products that are in suspension or emulsion form. The nature of the sample measured is typically heterogeneous and is prone to scatter effects that may distort the spectral consistency between successive scans. There are a number of excellent preprocessing methods used to modify NIR spectra to minimize the effects of scattering due to

particulate matter, oil bubbles, and air pockets that may exist in these samples, and Section 71.5 outlines some of the common methods applied before building NIR quantitative or qualitative models.

71.4.2 Diffuse Reflectance

In many applications, the use of transmission is not always practical, particularly when measuring solid samples in a high-speed environment. While transmission measurements may provide more robust quantitative models due to the greater sampling cross sections achievable to generate a reliable spectrum, up to 120 coadded scans may be required. This is not a feasible solution for high-speed applications.

Diffuse reflectance measurements typically use a direct illumination, fiber optic device, or integrating sphere (Section 71.4.2.1) to collect sample spectra. The process of diffuse reflectance spectral collection is provided in Figure 71.21, and the general instrument setup was previously shown in Figure 71.10.

Since reflectance measurements are made on the surface of a sample, the general assumption that has to be made is that the sample is homogeneous throughout the depth of penetration of the radiation. In NIR spectroscopy, the depth of penetration of the radiation is in the order of 2–10 mm, depending on the nature of the samples and the particle size distribution. Figure 71.22 provides a diagrammatical representation of the depth of penetration and particle size and how it influences the absorbance values of the resulting spectrum.

A practical means of determining the depth of penetration of the radiation for a particular application is to fill a borosilicate glass vial with varying levels of the material

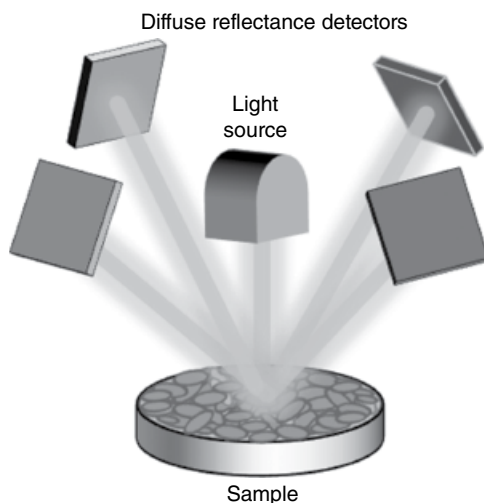


FIGURE 71.21 General instrument configuration for collecting NIR spectra in diffuse reflectance mode.

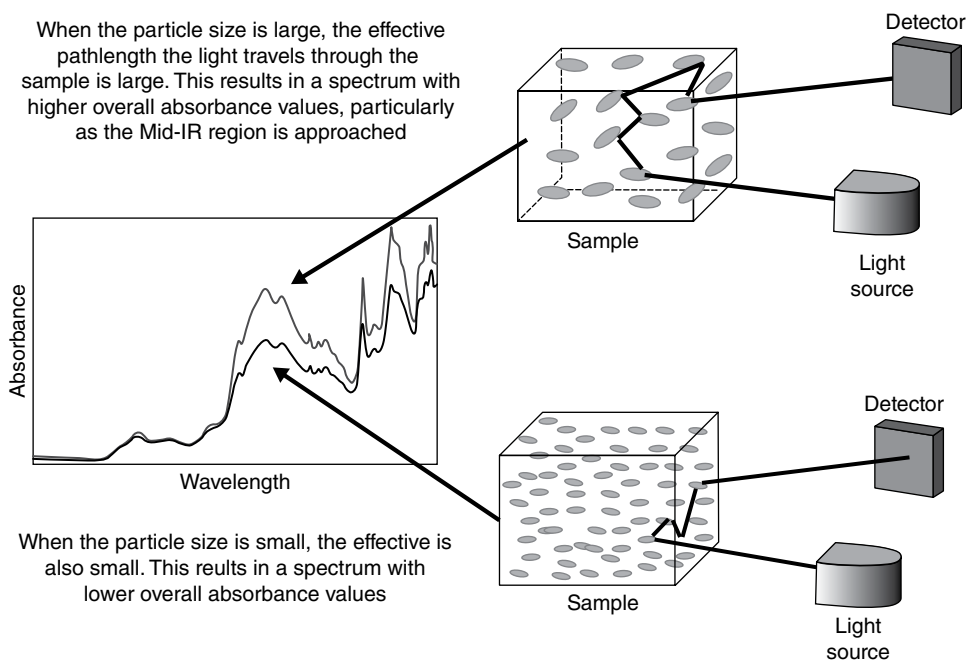


FIGURE 71.22 The relationship between a sample particle size and the absorbance measured by an NIR spectrometer.

to be analyzed and place a material of known spectral characteristics on top of the material. The depth at which only the spectral features of the material to be analyzed are observable in the spectrum can be used as an assessment of the depth of penetration. Another mode of reflectance in NIR spectroscopy is known as specular reflectance. Modern NIR sampling devices are designed to minimize this effect by placing the detectors at 45° to the incident light source. Specular reflectance results in the incident light being reflected back to the source at approximately 180° to the source. This radiation therefore has minimal interaction with the absorbing matter and therefore contains little to no chemical information. The ultimate goal of the NIR method is to collect the most relevant chemical and physical information from the sample, and this information can only be reliably found in a diffuse reflectance spectrum.

71.4.2.1 The Integrating Sphere An extremely useful sampling device in many types of spectroscopy is the integrating sphere. It is particularly well suited to diffuse reflectance measurements and is utilized by a number of instrument vendors. Figure 71.23 provides a diagrammatic representation of an integrating sphere used for diffuse reflectance measurements.

The device consists of a precisely constructed glass sphere coated with a highly efficient material that reflects light (typically gold for NIR applications). The light source is located at 180° to the sample. When light is incident on the sample surface, the

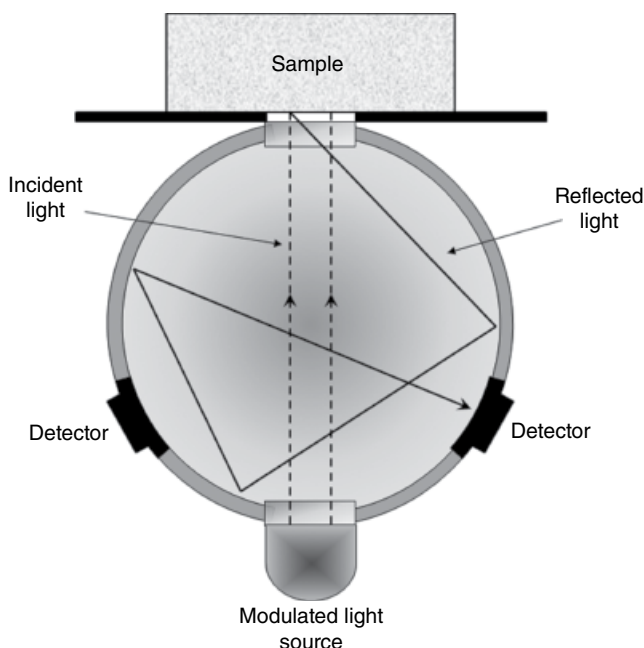


FIGURE 71.23 The integrating sphere.

specular reflectance is sent back to the light source and is therefore not detected. The light that is not specularly reflected can be considered to be diffusely reflected, and this light bounces inside the sphere until it finally reaches the detector.

71.4.2.2 Interactance Measurements Due to the large effects of specular reflectance in the SW region of the NIR spectrum, diffuse reflectance sampling devices provide very little information to the analyst. Figure 71.24 shows this region of the spectrum for a sample measured in diffuse reflectance mode.

In order to use reflectance devices in the SW NIR region, an interactance device is used. This is just a large area of light and detection surface designed to push as much light as possible onto a sample and collect as much of the reflected signal as possible. It is designed to have close contact with the sample and has found much use for the measurement of emulsions and gels, where scatter effects can dominate when measuring at longer wavelengths in the NIR region. Figure 71.28 shows an example of an interactance probe and its fiber configuration.

71.4.3 Sample Preparation

Samples present themselves in many forms, and these can be such items as pharmaceutical tablets, bottled wine, crushed sugarcane, and gasoline, just to name a few. This section provides a practical guideline on what type of sample preparation is required to obtain high-quality NIR spectra for the vast number of applications.

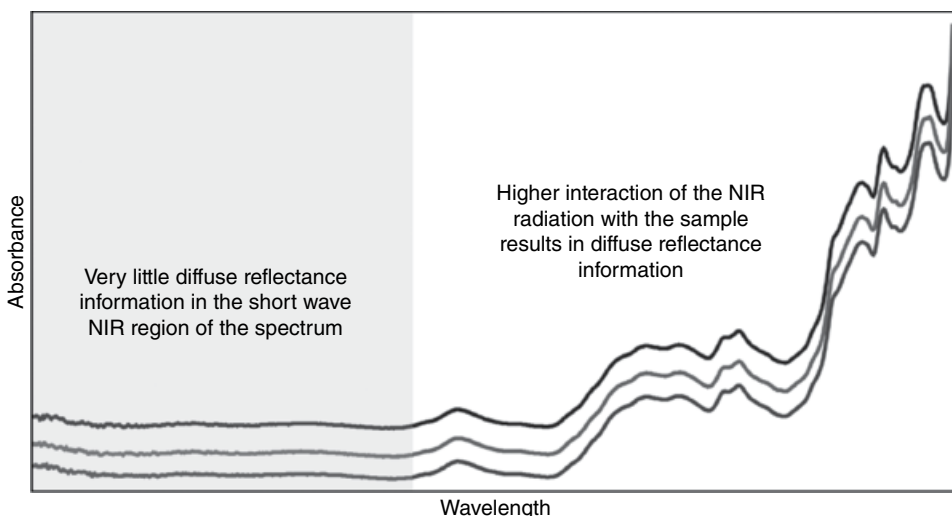


FIGURE 71.24 Sample spectrum measured in diffuse reflectance mode showing little chemical information in the short wave (SW) region.

71.4.3.1 Solid Powders Solid powders can be comprised of a single “homogeneous” material, such as a pure raw material, or may be a blend of two or more components. Powder samples should typically be loose and free of visible clumps. Clumping may be the result of moisture contamination, and if this is not the usual state of the material, it is important to assess the sample quality visually before analysis by NIR spectroscopy.

The method developer must ensure some level of homogeneity of powder blend samples before analysis to avoid nonrepresentative sampling. If this is a concern, multiple scan averages should be collected to avoid heterogeneity problems. Some NIR instruments come with a large sample area device that rotates the sample through a full revolution and takes multiple scans. The average of this scan is used as the final spectrum for analysis or method development purposes.

In the case of flaky materials (usually in the form of small plates), some sample preparation may be required to reduce the particle size and therefore reduce the high sampling variations associated with poorly packed materials. This is easily achieved through a light preparation in a mortar and pestle.

If the powder material is to be measured at line, sampling methods such as borosilicate glass vials can be used to measure the material in a manner that allows visual inspection of the material before scanning. This method also allows complete control over the sample preparation and due to the nondestructive manner of the analysis allows for retention of the same sample used for the analysis. This is highly important in cases where sample traceability is required (particularly in the pharmaceutical industry).

71.4.3.2 Grains and Seeds The agricultural sector utilizes NIR for many of its operations (refer to Section 71.6.1) and, in particular, for the analysis of grains and

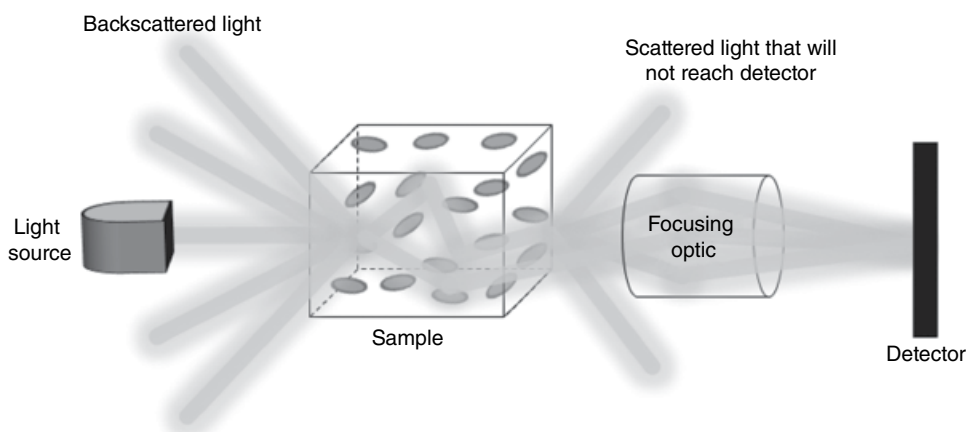


FIGURE 71.25 The process of diffuse transmission as applied to grains and oilseed analysis.

oilseeds. The main mode of analysis has been transmission mode using specially designed grain hoppers. The grain is fed into the top of the hopper of an at-line instrument where it is fed into a fixed pathlength cell. NIR radiation is passed through the sample, and after scanning, the old sample is ejected and a new sample is received. This process is performed multiple times in order to minimize packing differences in the sample and heterogeneity. Figure 71.25 shows the process of diffuse transmission as applied to grains and oilseeds.

71.4.3.3 Silage and Sugarcane During the harvest of grains, the remaining stalks (also known as silage) are also harvested. The material presented to the instrument is either high in moisture or as a dry straw. Sugarcane, after crushing in a mill, also has a similar consistency. To analyze this sample type, the use of overhead NIR instrumentation is required.

Silage is typically monitored in a harvester in real time using a device located on the exit chute to a hopper (Fig. 71.26a) or utilizing an innovative sampling method on a conveyor belt in the case of the sugarcane industry (Fig. 71.26b). Again, due to the heterogeneity of the samples, scans are taken over a specific sample area and averaged to result in a final scan, which minimizes density differences and heterogeneity. More details of these applications are provided in Section 71.6.1.2.

71.4.3.4 Pharmaceutical Tablets Many instrument configurations exist for the analysis of intact pharmaceutical tablets. The most common sampling approach is to develop custom-made holders for the tablet type and analyze the tablets in transmission mode. Figure 71.27 shows this setup diagrammatically.

By measuring tablets in transmission mode, a greater area of sample is measured through the interior of the tablet. In this case, all of the sampling considerations of diffuse transmission must be taken into account, that is, tablet thickness (pathlength

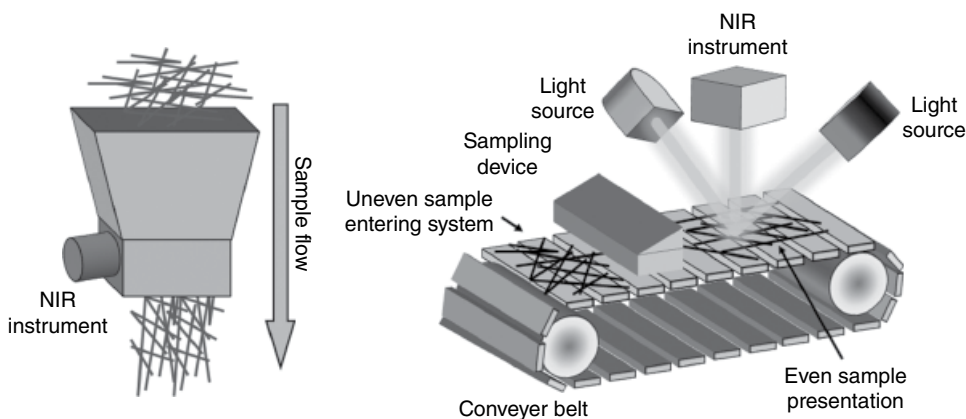


FIGURE 71.26 Analysis of silage (left image) or sugarcane (right image) requires sampling devices that best account of sample heterogeneity.

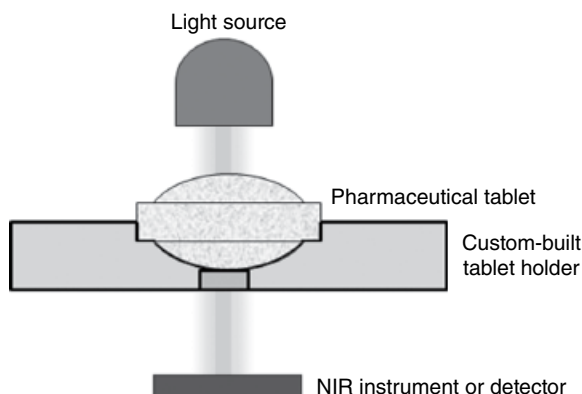


FIGURE 71.27 General sampling device used for measuring intact pharmaceutical tablets in diffuse transmission mode.

considerations), region of the spectrum used, etc. Due to detector response saturation above 1500 nm in transmission mode, most of the spectral information is useless for analysis (refer to Fig. 71.7). It is up to the method developer to exclude these noise parts of the spectrum before a calibration model can be developed. Diffuse reflectance measurements can also be used for measuring intact tablets, provided the following conditions hold:

1. The concentration of the active material in the tablet is high (typically >5% w/w).
2. It can be assumed that the distribution of the active material is homogenous over the surface of the tablet.

This will be discussed more in Section 71.6.2.2.

71.4.4 Fiber Optic Probes

This section provides a brief overview of some of the common types of fiber optic probe currently in use for NIR applications. Fiber optic probes allow for larger instruments or instruments that are sensitive to mechanical shock or environmental conditions to become “portable.”

Fiber optic probes bring the instrument to the process and can be configured in any of the sampling arrangements previously discussed in Section 71.4. Fiber optic probes are most commonly used in process applications, where measurements are performed on the sample as it exists in the process. The use of fiber optic probes is now becoming limited due to the emergence of smaller and more robust instruments, which can make direct contact with the process. While practically advantageous in some applications, fiber optic probes add another level of complexity, particularly when it comes to instrument matching for calibration transfer purposes. The following subsections discuss the various types of fiber optic probe available.

71.4.4.1 Reflectance Probes There are a number of configurations for a direct insertion reflectance probe, and these can be used to measure both solid and liquid interfaces. Reflectance probes are typically designed for the collection of diffuse reflectance spectra, but they also come in designs that accommodate attenuated total reflectance (ATR) and interactance. Figure 71.28 provides a schematic diagram of the probe types available.

These probes are designed such that they can be permanently mounted into a vessel or pipe (with a self-referencing and self-cleaning mechanism, as is a design feature of

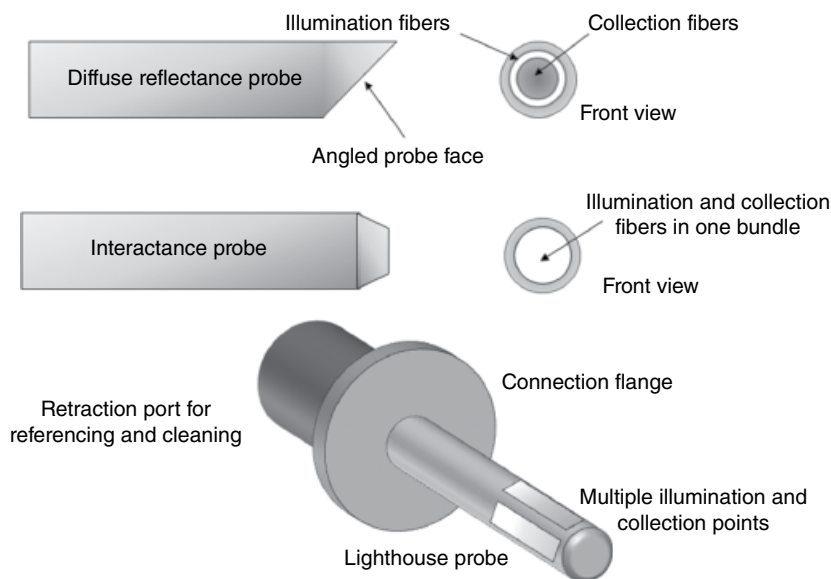


FIGURE 71.28 Examples of reflectance probes commonly in use for NIR process monitoring applications.

the lighthouse probe) or designed to be inserted and withdrawn from a suitably designed housing that controls the depth of the probe into the sample.

Reflectance probes have found widespread use in applications from the pharmaceutical, biopharmaceutical/biotechnology, and petrochemical sectors where monitoring of critical processes can be performed in real time and changes can be made to the process before they become an issue.

71.4.4.2 Transmission Probes Transmission probes typically find use in on-line process applications where the sample can be sidestreamed (and sometimes conditioned) before analysis. Typically suited for clear to relatively turbid (high scattering) liquid applications, transmission probes can offer some real advantages over reflectance probes due to the higher sample volumes measured. Figure 71.29 shows two common types of transmission probe in use for industrial applications.

When the transmission probe is configured as a direct insert probe, the light path is directed from the source to the sample. The pathlength can be adjusted by the use of a calibrated scale on the probe, which can be precisely set; however, in most cases probes come in a fixed pathlength configuration. Once the light passes through the sample (defined by the pathlength), the nonabsorbed light is bounced off an internal mirror and can be sent back to the detector in one of two ways:

1. It can be reflected back through the sample. In this case, the pathlength is doubled since the light has to pass through the sample twice before it reaches the detector.
2. It can be reflected back. This time the light path is through a separate fiber optic that does not pass through the sample twice.

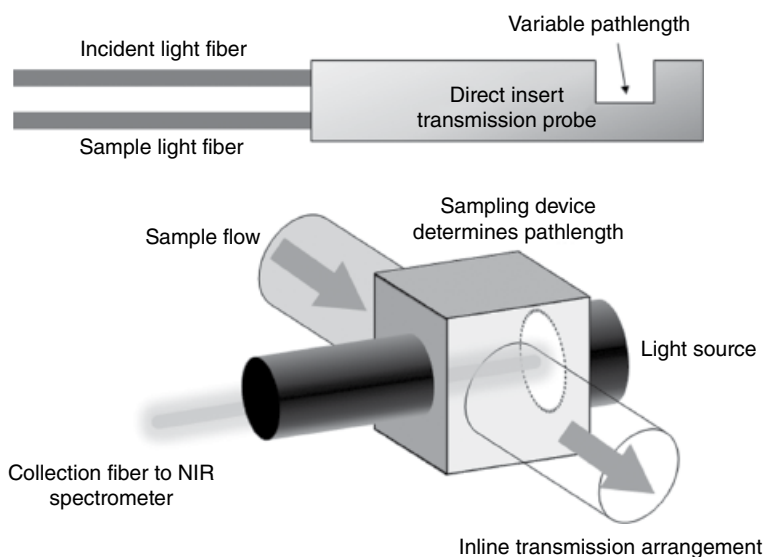


FIGURE 71.29 Examples of transmission probe systems typically used for monitoring processes using NIR spectroscopy.

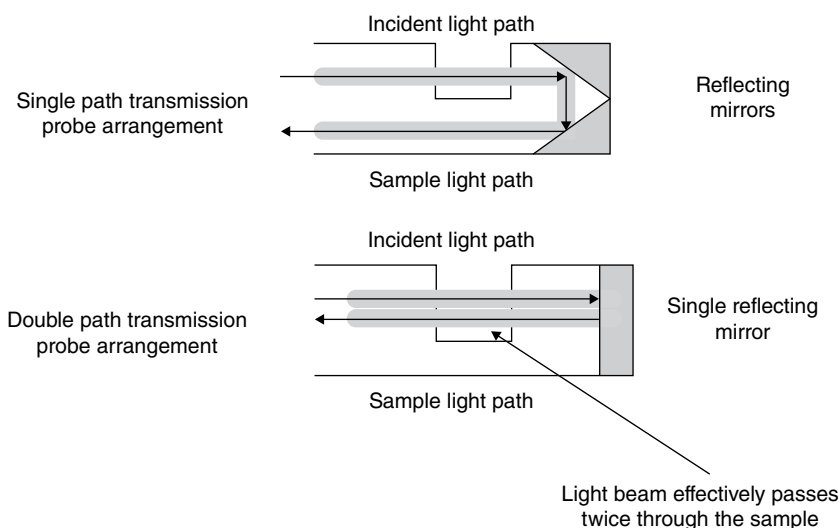


FIGURE 71.30 Configurations of direct insert probes used for measuring transmission NIR spectra.

These two situations are shown in Figure 71.30.

As was the case of reflectance probes, transmission probes can be permanently assembled into a process line or inserted and withdrawn from a fixed housing. The requirement of a transmission probe is that it needs a clear optical path to collect a reference spectrum.

71.4.5 Summary of Sampling Methods

This section provided a brief overview of the sampling methods that can be employed to capture reliable NIR spectra in a laboratory and a process monitoring situation. When approaching the design of a system capable of measuring meaningful NIR spectra for a particular application, the design engineer or scientist must use all of the information available to them in order to optimize the system for the application. This includes ensuring that the following main criteria have been taken into consideration:

1. Ensure that the NIR method is capable of performing its task by conducting feasibility studies.
2. Determine the required speed of analysis as this will determine the type of dispersion device used for the application (refer to Section 71.3).
3. Determine whether spectral resolution is a limiting factor in the analysis and choose an instrument that meets both the speed and resolution requirements.
4. If a moving grating or FT instrument is to be used, the sampling requirements must be optimized for the application. Determine whether the sample can be brought to the instrument (at line) or if the instrument needs to be interfaced with the process via fiber optic probes (refer to Section 71.4.4).

5. If a portable instrument is to be used, can it be interfaced with a minimum or no fiber optic cabling? If so, this is the ideal situation, and this requires that the instrument needs a permanent mounting to keep it in position for analysis. Also assess the environment for the possibility of explosion and ensure that the instrument configuration is rated to comply with the requirements of such environments.
6. Once the instrument configuration is built and the housings are all secured, a suitable graphical user interface (GUI) must be decided upon. This can be as simple as the vendor of the instrument native platform or can be customized for process operators using traffic light systems or simple pass/fail messages.
7. Determine whether the results generated by the NIR analysis are to be used for process control. In this case, the output of the instrument must be sent via a standard communication protocol to the control system such that action can be taken.

Once all of these factors have been taken into account, this should give the end application the best chance of success. It cannot be stressed more; representative sampling is the key to reliable NIR method development. The Theory of Sampling (TOS) is outside of the scope of this chapter, but it must be taken into account before designing a reliable system. The interested reader is referred to the excellent literature on TOS [43].

Representative sampling ensures that the data collected for model building is of the highest quality. There will always be random effects present in spectral data due to particle size of materials and varying sample density. This is where preprocessing of data can be used to minimize such effects, and this is discussed in the next section of this chapter.

71.5 PREPROCESSING OF NIR SPECTRA FOR CHEMOMETRIC ANALYSIS

NIR spectroscopy in general does not produce the sharp absorbance bands, which are typical of other spectroscopic methods such as MIR, Raman, and NMR spectroscopy. For this reason, early researchers did not investigate this region as it provided little structural information that could be interpreted in the traditional ways.

It is true to say that NIR and chemometrics (see Chapter 65) go hand in hand. This section does not go into the details of the algorithms used to generate predictive or classification models but focuses primarily on the preprocessing methods used to minimize unwanted effects before the modeling process is performed.

It is stressed here that preprocessing is not a substitute for good sampling! Ninety percent of the initial effort for implementing an NIR system should be given to optimal sampling. If the data that are generated by the instrument are nonrepresentative or substandard, the final model is either bad or requires more components to account for the

bad sampling. This results in a complex model that, in the case of failure, may not be interpretable. This is a case of garbage in–garbage out.

71.5.1 Preprocessing of NIR Spectra

NIR spectroscopy is typically used to measure solid samples either in powdered form or as grains (agriculture). This introduces variability into the packing as the solid sample cannot be packed the same twice in a row. This results in packing density variations that affect both reflectance and transmission measurements, and these are commonly known as additive effects, since they primarily affect the baseline of the spectra.

Varying particle size distributions also contribute to spectral variations between samples. In the NIR region of the spectrum, the wavelength of the radiation greater than 1600nm becomes of comparable size to the particles in the sample, and thus elastic rather than specular interactions occur between the incident radiation and the sample (refer to Section 71.4.2). The overall effect is to introduce nonlinear (or scatter effects) into the spectra, also known as multiplicative effects. The following sections briefly describe the most common preprocessing methods used in NIR spectroscopy for minimizing both additive and multiplicative effects.

71.5.2 Minimizing Additive Effects

An additive effect can be considered as a constant offset in the spectra. To remove such effects, a differencing factor must be introduced such that a common baseline is achieved. In NIR spectroscopy, only in certain cases (particularly the analysis of nonscattering liquids) is the baseline ever linear. The most common approach to minimizing purely additive effects in NIR spectra is by the use of derivatives. There are a number of commonly used approaches to derivatives, and most software packages either employ:

1. The Savitzky–Golay derivative [44] or the
2. Segment–Gap derivative [44]

These two derivative types are discussed in more detail as follows.

71.5.2.1 How Derivatives Work By definition, derivatives (also known as differentiation) are (generally speaking) calculated as the difference between the second point in the spectrum and the first (divided by a constant centering factor). This acts to center the entire spectrum around the zero line. It also measures the slope of the spectral features (since by definition, derivatives measure the rate of change of data).

Figure 71.31 provides some Gaussian curves of various intensities offset from each other in a linear manner along with their first derivatives.

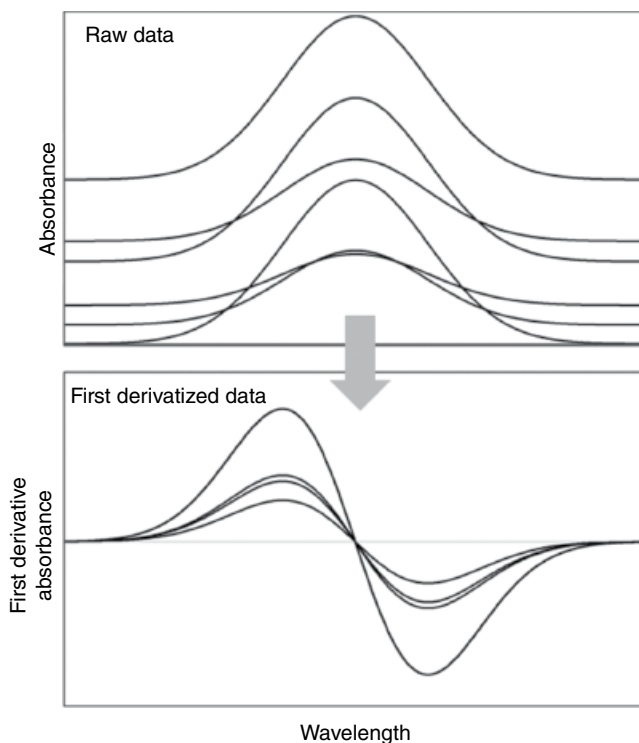


FIGURE 71.31 Gaussian curves of various intensities and offsets along with their first derivatives.

By taking the first derivative of these spectra, Figure 71.31 shows that the data are all centered around zero and the relative intensities of the curves can now be compared on a quantitative basis.

To further explain how the first derivative works, consider a single Gaussian curve and its corresponding derivative in Figure 71.31. Moving from left to right in the Gaussian curve, the slope of the curve is close to zero. From there it rapidly increases until the inflection point is reached. The slope then rapidly decreases to zero when the peak maximum is reached in the Gaussian peak (this is the zero point in the derivative curve). The slope then rapidly decreases as the curve moves to the right of the maxima until it reaches the inflection point. From there, the slope then rapidly approaches close to zero when the end of the curve is reached.

The second derivative is defined as the curvature of the original curve or the slope of the first derivative. Consider the points shown in Figure 71.32.

The first and second derivatives can be described mathematically as follows:

General first derivative:

$$\frac{\partial Y}{\partial \lambda} = \frac{(B - A)}{\Delta \lambda}$$

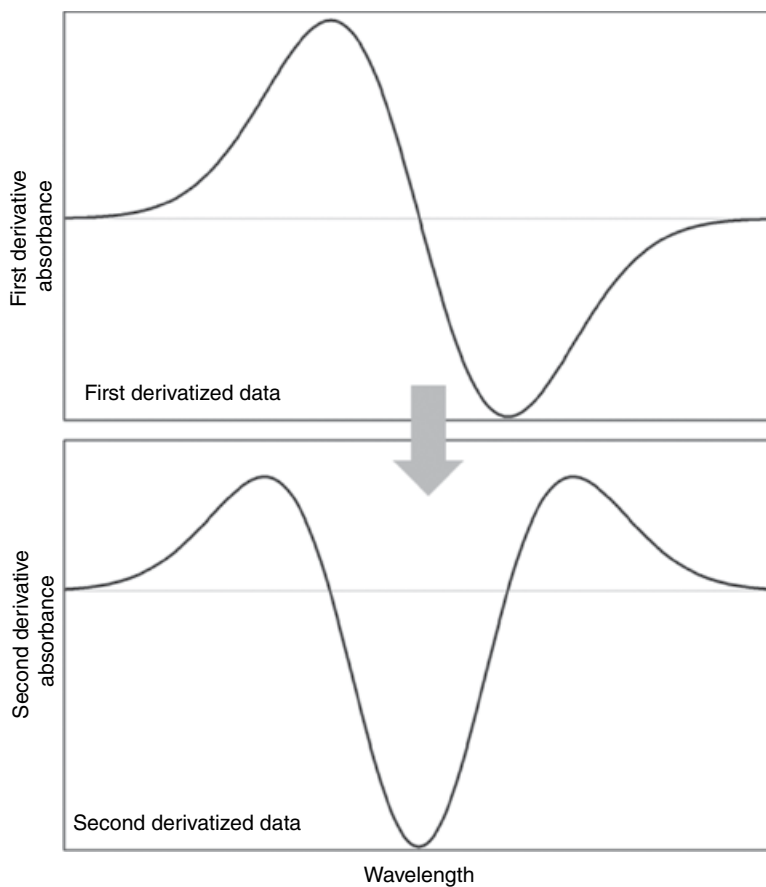


FIGURE 71.32 First and second derivatives of a Gaussian curve.

General second derivative:

$$\frac{\partial^2 Y}{\partial \lambda^2} = \frac{(C - B) - (B - A)}{\Delta \lambda}$$

$$\frac{\partial^2 Y}{\partial \lambda^2} = \frac{C - 2B + A}{\Delta \lambda}$$

where

Y is the y -axis absorbance scale

λ is the x -axis wavelength scale

A , B , and C are the first, second, and third points, respectively, in the spectrum to be derivatized

In the definition of the second derivative, the first equation represents the first derivative of the first derivative. Figure 71.33 shows the application of both first and

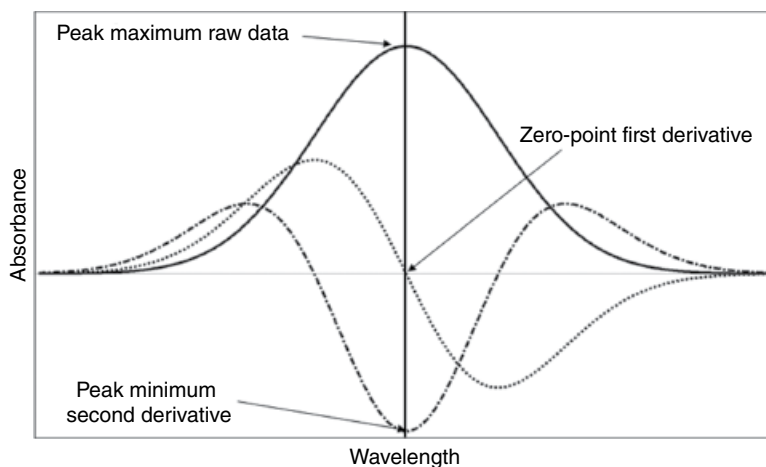


FIGURE 71.33 The relationship of curve maxima, zero point, and minima for Gaussian, first, and second derivative curves.

second derivatives to a Gaussian curve. In the case of the second derivative, the peak maximum in the original curve becomes the peak minimum in the second derivative. This is why the second derivative is preferred sometimes as peak minima are more interpretable than zero points in the first derivative.

Another important feature of the second derivative particularly for application to spectral data that are collected above 1600 nm is that they take curvature into account. Since the baseline of the spectra between 1600 and 2500 nm exhibits quasi-quadratic behavior, the second derivative is particularly useful in minimizing additive effects in this region.

71.5.2.2 The Savitzky–Golay Derivative The Savitzky–Golay algorithm was proposed in 1964 as a digital smoothing filter [44]. It was extended to be a derivative and has found widespread usage in analytical chemistry. The algorithm is based on performing a least squares linear regression fit of a polynomial around each point in the spectrum to smooth the data. The derivative is then the difference of the fitted polynomial at each point. The algorithm includes a smoothing factor that determines how many adjacent variables will be used to estimate the polynomial approximation of the curve segment. Figure 71.34 shows how the algorithm works.

In general, the derivative works by fitting a polynomial to the data defined by the smoothing window. The window size must be an odd number as the smoothed point in the window lies at the center, and this point becomes *A* in the equations defined in Section 71.5.2.1. The window is then moved along by a one-point increment, and a polynomial is fit to the data spanning the smoothing window and becomes the point *B* in the equations listed in Section 71.5.2.1. The point *C* can be calculated in a similar manner. The first and second derivatives are calculated as per the equations. The effect of the polynomial is to provide a smoothed point that is less noisy than the original

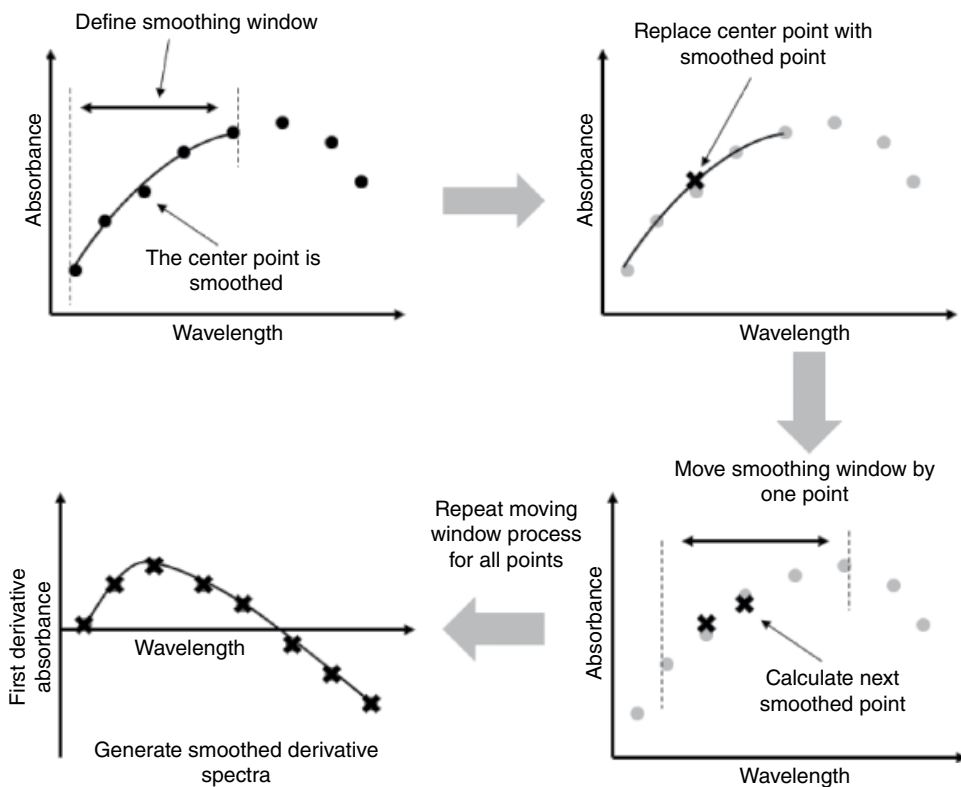


FIGURE 71.34 Diagrammatic representation of the Savitzky–Golay derivative.

data; therefore the final derivative is smooth with minimal noise characteristics. More details on how to set window sizes are discussed in Section 71.5.2.4.

The Savitzky–Golay derivative is used in most software packages that are supplied with NIR instruments. The algorithm was corrected for errors in the original tables by Steiner et al. [45] and is the preferred method by most chemometricians because it fits exact mathematical functions to the data to generate a continuous derivative.

71.5.2.3 Segment–Gap Derivatives The Segment–Gap derivative enables the computation of derivatives using an algorithm that allows selection of a gap factor and a smoothing factor. The principles of the Segment–Gap derivative are based on a modification of the moving average algorithm where a suitable smoothing window is used to calculate the average point in the center of the window. The gap size is set such that different sizes between the windows can be set; however, the most common value for the gap size is 1. For such functions, Norris suggested that derivative curves with less noise could be obtained by taking the difference of two averages, formed by points surrounding the selected x locations [46]. As a further simplification, the division of the difference in y -values or the y -averages, by the x -separation x , is omitted.

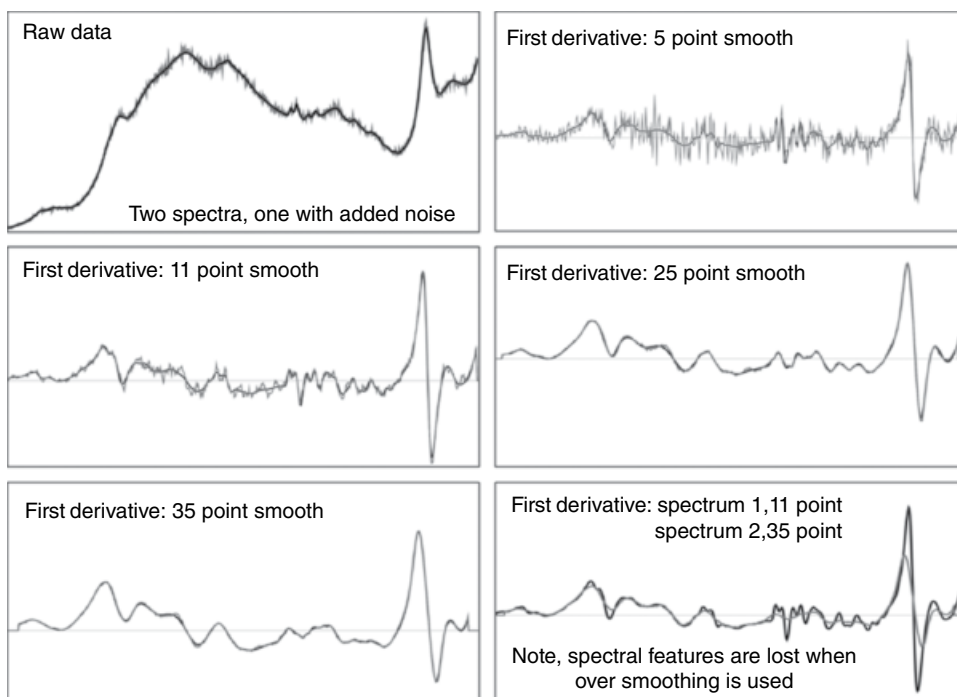


FIGURE 71.35 Comparison of derivative data. One curve is smooth, while the second is the original curve with added noise. It can be seen from this figure that in attempt to remove all of the noise from the spectrum, specific spectral features are smoothed out.

Norris [47] introduced the term segment to indicate the length of the x interval over which y -values are averaged, to obtain the two values that are subtracted to form the estimated derivative. If too large a segment is defined, the resolution of the peaks will decrease. Too narrow a segment (smaller than the half bandwidth of the peak) may introduce noise in the derivative data. This is illustrated in Figure 71.35.

71.5.2.4 Some Pitfalls of Derivatives The definition of suitable window segment sizes is the key challenge to optimizing a derivative. There are a number of factors that must be taken into account when applying a derivative, and if these are overlooked, valuable information may be lost in the spectral data. The following is a checklist that must be taken into account before applying a derivative:

1. Is the baseline shift linear or nonlinear? This aids in the determination of the derivative order to use.
2. Is the wavelength scale linear over the entire range? If a nonlinear scale is used, regions of higher resolution will be smoothed less than regions of lower resolution.

3. The larger the smoothing window, the more noise is removed; however, if too large a window is used, important spectral information may be reduced or even lost through “averaging out.”
4. Derivatives enhance noise! Since the process of derivatization is successive differences between points, depending on the data point resolution of the spectrum, the first derivative scale can be 1–2 orders of magnitude lower in y-axis scale than the original scale of the data. The second derivative has an even smaller scale (sometimes close to the noise level of the instrument).

This is why points 3 and 4 are so important and go hand in hand with each other. It is important to focus any preprocessing on the region of interest to the detriment of other regions. A simple rule of thumb with derivatives is to understand the Full Width at Half Maximum (FWHM) of the most important spectral features and set the smoothing window such that it has minimum effect on the peak shape and intensity.

5. When applying a smoothing window, by definition, the value of the window size is odd such that the center point in the window is the one that is smoothed. This means that the points to the left of the center point on the edge of the spectrum will be eliminated from the preprocessed data; the same also holds at the end of the spectrum for the points to the right of the center point. A method developer must be aware of this fact if they are preprocessing a region of data within the full spectrum range. It is important to leave a number of slack points before and after the region of interest to avoid removing important information.

Overall, derivatives are probably the most used preprocessing technique for NIR spectra. If the previous rules are followed, the preprocessing can be optimized efficiently and can be applied to new data before any predictions are made using chemometric models.

71.5.3 Minimizing Multiplicative Effects

71.5.3.1 Standard Normal Variate and Detrending The Standard Normal Variate (SNV) algorithm is a nonmodel-based scatter correction algorithm first proposed by Barnes et al. in 1989 [48]. The algorithm requires no external spectral information for correction; it uses only the spectrum itself for the correction information. The algorithm works by first calculating the mean value (m_i) and the standard deviation (s_i) over all wavelengths in the spectrum. Then for each wavelength in the spectral region chosen, the mean absorbance is subtracted from each wavelength, and it is then divided by the overall spectrum standard deviation. This can be expressed mathematically as

$$z_{i,j(\text{SNV})} = \frac{z_j - m_i}{s_i}$$

where

$z_{i,j(\text{SNV})}$ is the SNV corrected absorbance value at wavelength (j) for spectrum (i)

z_j is the absorbance value of the original spectrum at the j^{th} wavelength

m_i is the overall mean absorbance value for the i^{th} spectrum

s_i is the standard deviation of all absorbance values for the i^{th} spectrum

Figure 71.36 shows the effect of applying the SNV algorithm to a set of NIR spectra highly affected by scatter.

SNV centers the data around the center of gravity of the spectrum and then scales the spectrum to ± 3 standard deviations around the zero line. The overall effect of this preprocessing method is to minimize the physical scatter and packing variations and reveal chemical information for chemometric modeling. It is a simple scatter correction technique and is implemented by many vendor software packages.

In their original paper [48], the authors also suggest that the detrend algorithm be applied to SNV corrected spectra. Detrending is a transformation that minimizes non-linear trends, thus SNV and detrend in combination reduce multicollinearity, baseline shift, and curvature. The detrending calculates a baseline function as a least squares fit of a polynomial to the sample spectrum. These transformations are applied to individual spectra and are distinct from other transformations that operate at each wavelength in a given set of spectra. As the polynomial order of the detrend increases, additional baseline effects are removed (0 order: offset; first order: offset and slope; second order: offset, slope, and curvature).

Typically, detrending is performed by using a second-order (or higher-degree) polynomial in regression analysis where absorbance values (or y -variables) and the independent variable or x -variable (W) are given by the corresponding wavelengths:

$$z_{\text{Detrend},i} = A_{\text{SNV}} + B_{\text{SNV}}W + C_{\text{SNV}}W^2 + \left[D_{\text{SNV}}W^3 + E_{\text{SNV}}W^4 \right]$$

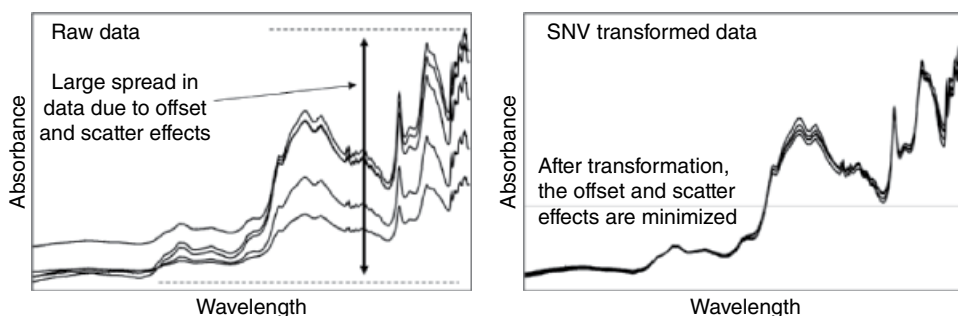


FIGURE 71.36 The Standard Normal Variate (SNV) preprocessing minimizes the effect of offset and scatter in NIR spectra while retaining the original spectral profile.

where A , B , C (and D , E) are the regression coefficients. The base curve in the earlier relationship is given by the fitted values $z_{\text{Detrend},i}$, and thus derived spectral values subjected to SNV followed by detrend become

$$z_{\text{SNV/Detrend}} = z_{\text{Detrend},i} - z_{\text{SNV},i}$$

This calculation removes baseline shift and curvature, which may be found in diffuse reflectance NIR data of powders, particularly if they are densely packed. The use of this transform does not change the shape of the data, as is the case with the application of derivatives (Section 71.5.2).

71.5.3.2 Multiplicative Scatter Correction Algorithms Multiplicative Scatter Correction (MSC) is a model based transformation method used to compensate for pure additive and/or multiplicative effects in spectral data. Extended Multiplicative Signal Correction (EMSC) is an extension of the original method that in addition allows modeling of different types of wavelength-dependent effects that can be found in the spectra.

Multiplicative Scatter Correction MSC (also known as multiplicative signal correction) was originally designed to deal with multiplicative scattering in reflectance spectroscopy [49]. However, a number of similar effects can be successfully treated with MSC such as:

- Pathlength variations
- Offset shifts
- Interference

The simplest form of light scattering can be described by a simple additive baseline shift (a_i). The measured spectra are then given as

$$z_i = a_i + z_{i,\text{chem}} + e_i$$

where $z_{i,\text{chem}}$ is the theoretical spectra without any noise or scattering effect. The error vector e_i describes the random measurement noise and other scattering effects not described by the equation. It should be emphasized that the baseline shift will be different for each individual spectrum and is modeled separately.

A multiplicative effect (b_i) is also included in MSC so that the measured spectra are given as

$$z_i = a_i + b_i z_{i,\text{chem}} + e_i$$

Since the theoretical spectrum $z_{i,\text{chem}}$ is not known, the individual spectra are instead approximated as

$$z_i = a_i + b_i m + e_i$$

where m is some reference spectrum, for example, a typical sample or the mean of a set of spectra. The unknown additive and multiplicative parameters a_i and b_i are then estimated by simple linear regression of the individual spectrum on the reference spectrum. Unlike normal least squares fitting, which aims to minimize the term e_i , MSC aims to model the scatter effects in the data (defined by $b_i m$) therefore removing these effects and maximizing the chemical information in the data. Thus e_i is maximized to retain chemical information.

The coefficients can then be used to calculate MSC corrected spectra:

$$z_{i,\text{corrected}} = \frac{(z_i - a_i)}{b_i}$$

The effect of the MSC algorithm is shown in Figure 71.37.

To investigate whether a set of measured spectra have additive or multiplicative effects, it is often instructive to plot the individual wavelengths against the reference (or average) spectrum. If the spectra have multiplicative scattering, they will have different slopes in such a plot. Additive baseline shifts are seen when the spectra intersect differently with the vertical axis.

As the MSC preprocessing uses the mean spectrum as reference for the data set, the assumption is that this calculated mean is representative for all current and future spectra collected using similar experimental conditions. It is also assumed that the observed scatter effects are not significantly associated with the chemical signal that is to be retained. To safeguard against removing the chemical signal along with the scatter effects, it is recommended to keep known important spectral regions or peaks out of the calculations. If such information about chemical signal is sparse, including the full spectra might still work under the assumption that the chemical signal cancels out when seen across the entire wavelength range.

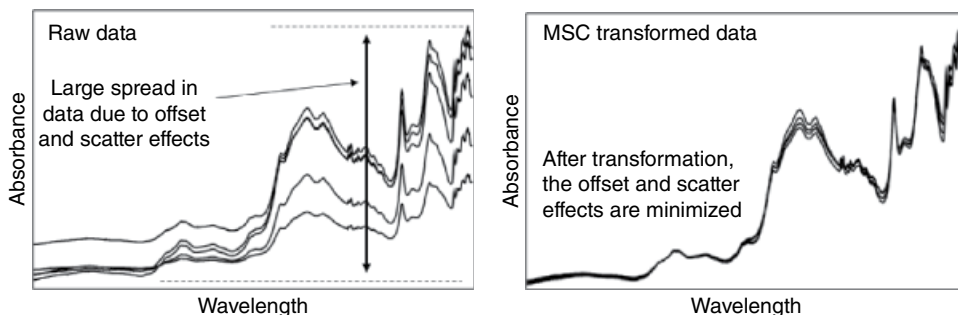


FIGURE 71.37 The application of the MSC algorithm to homogenous data allows the correction of additive and multiplicative effects caused by sample packing and light scatter.

Another way to avoid removing important variation from the data is to include known good spectra in an EMSC model (see section “Extended Multiplicative Signal Correction”).

Extended Multiplicative Signal Correction EMSC is an extension to conventional MSC, which is not limited to only removing multiplicative and additive effects from spectra [50]. This extended version allows a better separation of physical light scattering effects from chemical light absorbance effects, by including wavelength-dependent effects or *a priori* information in the modeling.

The MSC model is then extended with additional terms according to the following equation:

$$z_i = a_i + b_i m + d_i \lambda + g_i \lambda^2 + h_i \text{BS} + I_i \text{GS} + m_i m^2 + e_i$$

where d_i , g_i , h_i , I_i , and m_i are additional effects as explained as follows.

Channel Effects Wavelength-dependent light scattering variations are modeled by the third and fourth terms in the equation ($d_i \lambda + g_i \lambda^2$). The assumption is that the wavelength dependency follows a smooth polynomial of first or second order.

Squared Spectrum The reference spectrum is given as a linear term in the original MSC model. A squared term can also be included to allow for nonlinearity ($m_i m^2$).

Good Spectra The MSC and EMSC models fit a regression equation with the included model parameters explaining the different scattering effects. The residual term from the models will therefore contain the chemical signal of interest as well as any unmodeled scattering effects and noise. This is somewhat different from traditional applications of least squares regression where the residuals should be as small as possible. One assumption behind MSC and EMSC is that the chemical signal of interest is poorly correlated with the model parameters that describe the scattering. If such a correlation exists, parts of the chemical signal will also be removed from the model.

One approach to improve the EMSC model is to include a term modeling the chemical signal of interest. This is done with the term $I_i \text{GS}$ where the GS is a matrix of good spectra. The coefficient I_i is a vector since there will be one coefficient for each of the good spectra. Good spectra in this context might, for instance, be spectra of pure components present in the samples. For applications where the user has no such information, this term may simply be left out from the model equation.

When good spectra are included, the associated chemical signal will be explicitly modeled and hence not included in the residual term. After estimating the model parameters, the term $I_i \text{GS}$ will therefore not be subtracted like the other model terms but retained in the corrected spectra.

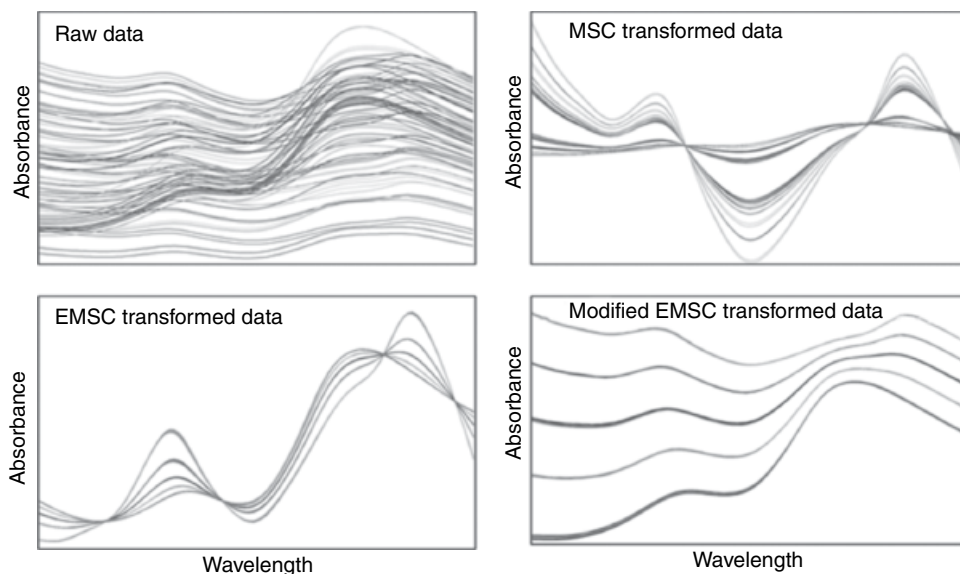


FIGURE 71.38 Effects on nonhomogeneous NIR spectra by the application of MSC, EMSC, and a modified version of EMSC, which takes into account external spectral information.

In a classical experiment conducted by Martens et al. [51] on gluten/starch mixtures, the use of good spectra is presented. The raw NIR data are shown in Figure 71.38 along with the EMSC corrected spectra using a good spectrum.

The good spectrum is a typical example of the underlying chemical constituents and in this case was the difference spectrum of gluten and starch. The overall effect of using a good spectrum is immediately seen in the EMSC corrected data. It is noted here that use of a simple MSC correction was detrimental to the data due to the chemical diversity present in the spectra. In this case, chemical information was confused with scatter information, and this highlights one of the major pitfalls associated with indiscriminate usage of the MSC method.

Bad Spectra Most of the terms in the EMSC equation described so far are used to model specific effects such as additive, multiplicative, or linear/quadratic channel effects. However, the scattering may manifest itself in many different ways. To correct for any general scattering effects, bad spectra can be modeled by the term $h_i BS$. Any spectra that explain the effect that should be corrected for can be included in the algorithm and that has not yet been described by the other terms in the EMSC equation.

In contrast to any good spectra included in the EMSC model, the effects of the bad spectra are subtracted from the input data along with the other estimated parameters.

Reference Spectrum The default and most commonly used reference spectrum is simply the mean of a set of representative spectra. However, it is sometimes more

TABLE 71.1 The Usage and Effect of Preprocessing Techniques Used in NIR Spectroscopy

Transform	Additive	Multiplicative
Derivative	Yes	Yes (second derivative only)
Baseline	Yes	No
SNV	Yes	Yes
Detrending	Yes	No
MSC/EMSC	Yes	Yes

useful to choose a different reference spectrum. If it is known *a priori* that another spectrum is more representative than the mean, this can be used instead.

71.5.4 Preprocessing Summary

Due to the nature of reflectance and diffuse transmission sampling typically used for the collection of NIR spectra, additive and multiplicative effects are a common occurrence and must be minimized in order to develop robust predictive models. In the SW NIR region where transmission measurements are typically performed, additive effects are the major contributor to spectral differences of similar samples. In the longer wavelength region, where diffuse reflectance measurements are typically performed, a combination of additive and multiplicative effects can be observed and in most cases must be jointly corrected for using preprocessing techniques such as MSC or SNV.

Table 71.1 provides an overview of the preprocessing methods commonly used for the minimization of additive and multiplicative effects and briefly describes their intended use.

There also exist many other preprocessing techniques used in NIR spectroscopy; however, this would require an entire chapter in itself to assess, and the interested reader is referred to the excellent literature for more details [52].

71.6 A BRIEF OVERVIEW OF APPLICATIONS OF NIR SPECTROSCOPY

NIR spectroscopy has found application in many industrial and research settings. In its early years, between 1935 and 1980, it was reported that there were 205 publications available [11]. Today, there are thousands of papers and applications with periodicals dedicated to work performed by researchers in NIR.

This section provides a concise overview of some of the industries that have gained the full benefits of the NIR technique. This overview is by no means exhaustive, and the interested reader is referred to the literature references cited in each section for more details of a particular application.

71.6.1 Agricultural Applications

Until the early to mid 1990s, agricultural applications of NIR spectroscopy dominated the literature and practical use of this method. Much emphasis was devoted to this area because the primary methods of analysis available for measuring agricultural products are detailed and time consuming.

NIR is well suited to the analysis of agricultural products since the main constituents of these products are proteins (N—H absorbances), carbohydrates (O—H and C—H absorbances), oils (C—H absorbances), and moisture (O—H absorbances). A major challenge in the development of agricultural methods is sample heterogeneity. Section 71.4 discussed some of the sampling techniques for such materials, and the following sections provide some practical examples of how these have been implemented.

71.6.1.1 At-Line Applications During grain and oilseed harvest time, producers of agricultural products deliver these to receiving points located primarily in rural regions. Before the use of NIR, samples would be taken from each delivery and sent to external laboratories for wet chemical analysis. In peak seasons, this places a tremendous workload on these laboratories and thus delayed the payment of the producer and sometimes leading to great frustration.

The pioneering work of researchers such as Norris [10], Shenk [14], and Williams [12] resulted in the utilization of NIR for the rapid, nondestructive analysis of whole or ground grain at the point of receipt.

A number of instrument vendors started to produce instruments based on diffuse transmission and diffuse reflectance that were capable of measuring multiple constituents in grains and oilseeds simultaneously. In the case of whole grain analyzers, samples collected from the delivery truck are placed into a grain hopper, and the instrument scans the entire lot by subsampling. The spectra collected are then averaged, and a quantitative model(s) is applied. The values of the properties of the grain are then used as the payment method, thus relieving the workload of external laboratories as the analysis time required is no longer than 30 s.

In the case of ground grain analyzers, these systems work in reflectance mode and require some sample preparation before analysis. This is typically performed using an industrial grade mill to produce samples of consistent particle size (thus avoiding the need to minimize such variability using preprocessing; see Section 71.5). Modern sampling techniques have also extended the usage of reflectance techniques to measure whole grains, and the use of ground grain analyzers is becoming limited.

A number of instrument vendors offer turnkey solutions for measuring agricultural products through the development of in-house calibration models. One of the challenges of developing models for the agricultural sector is season to season variability. This requires that calibration models must be developed over many years and is one of the major reasons why NIR is often overlooked as a viable alternative to laboratory

methods by some as it requires a training phase. Some words of wisdom: day 1 starts when an organization makes a conscious effort to build a model. The longer the delay, the longer it takes to get to day 1!

Some instrument vendors have collected spectra for many years and sent the samples used to collect the spectra to certified laboratories for reference analysis. The calibration models developed are based either on local–global calibration approaches [14] or artificial neural networks (ANN) [53].

Other applications of NIR for agricultural products include the analysis of grains and feed pellets for animal consumption and the nutritional and energy outputs of the batches produced [54]. One group is now using NIR to detect high levels of aflatoxin (an extremely toxic material, which results in many human deaths) in peanuts in third world countries [55]. In the area of rice production, fertilization management is a key economic driver, and the work of Batten [56] involved the development of a leaf nitrogen test to understand the fertilization requirements on rice crops during the growing season.

NIR has also been utilized heavily on the wine and grape sector [57]. During harvest, grape quality parameters can be assessed to determine color (anthocyanin content) and taste (tannins, sugars, total acids) at the point of receipt. Attempts to make more automated equipment for the analysis of grapes have met with some technical challenges, and in order to analyze a sample, some detailed preparation work is required.

The next section discusses extensions of the at-line method into real-time processes utilized by the agricultural sector.

71.6.1.2 In-Line Applications The scientific management of fertilization during the growing season is known as precision agriculture [58]. In order to assess the effects of fertilization management on a crop plot, a real-time analysis technology is required; fortunately, this is where NIR has excelled.

Numerous NIR instrument vendors have collaborated with harvest equipment manufacturers to produce systems that can analyze grain and silage during harvest. These NIR instruments are located close to the clean grain elevator or the exit chute of the harvester to make real-time predictions. Since the results of an analysis can be merged with GPS coordinates of the harvester, maps of protein and moisture, for example, can be generated for a particular field, and this information can be used to better understand how to manage fertilizer usage in the coming years and also to understand the water table profile associated with a field. In this way, if fertilizer is placed at the top of a water gradient, it will have more likelihood of dispersing throughout the entire field, rather than just blanket coating the field.

In the area of sugarcane analysis, one vendor [59] has patented a sampling system combined with an overhead NIR analyzer for monitoring parameters such as Brix (sugar), fiber, and other quality attributes. This type of analysis allows for better segregation of materials based on their quality. In the grain industry, traders can use NIR to blend lower-quality grains with premium grains in a way that still meets product specification but allows them to reduce the amount of low-quality material they store.

Figure 71.26 shows some of the in-line setups available for the analysis of agricultural products in real time.

71.6.2 Pharmaceutical/Biopharmaceutical Applications

It is true to say that the pharmaceutical industry (and for that matter any regulated industry) moves at a very slow pace. Quality control (QC) has traditionally been a laboratory-based activity where samples are sent for analysis, and based on the results, raw materials, intermediates, and final products are released.

In the early 2000s, the US FDA released a number of landmark publications to move the pharmaceutical industry into the twenty-first century. The initial work entitled *Pharmaceutical cGMPs for the 21st century: A risk-based approach* [60] defined a road map for pharmaceutical manufacturers to become innovative in the use of modern, state-of-the-art process analyzers. This publication was followed up soon after with the PAT framework guidance document [61], showing the agency willingness to support the concepts of the cGMPs for the twenty-first-century document. Since then, numerous industry groups and standards organizations have delivered specific guidance to industry, and a more detailed discussion can be found in the review by Swarbrick [18]. NIR in the early days was considered to be the driving force behind the PAT initiative and still forms a major part of it. Today, PAT extends to other spectroscopic methods such as Raman, MIR, and technologies such as in-line particle size analyzers.

71.6.2.1 Raw Material Identification and Conformity Based on Coblenz's early work, he found that molecules exhibit unique features or "fingerprints" in the IR region that can be used to distinguish between them [9]. This observation sets the basis for using MIR for raw material identification, and this method was (and still is) employed by pharmaceutical companies. The biggest issue with the use of MIR methods for raw material identification was that the method tended to be subjective and requires sample preparation. An analyst would typically take a sample (either from a single lot or a composite sample) and prepare either a paraffin-based mull or a potassium bromide disc of the material for MIR analysis. Once the spectrum was collected, it would be visually compared with a reference spectrum of the material kept on file, and a decision would be made on its identity.

One of the major advantages already discussed about NIR spectroscopy is based on it requiring minimal to no sample preparation. Ciurczak [62] in his early work described procedures for developing raw material identification approaches using NIR, and a brief description of the method is given here.

To develop a robust NIR raw material identification library, an analyst will collect between 6 and 15 samples of each of a number of raw materials, and using one of the sampling methods discussed in Section 71.4, samples are presented to the spectrometer in a consistent way such that sampling variations are minimized. Then, using a suitable preprocessing method to minimize the residual sampling effects (refer to

Section 71.5), raw material libraries for each material are developed, using chemometric methods, either based on correlation or Principal Component Analysis (PCA) (refer to Chapter 65).

To test the ability of the libraries for correctly classifying the raw materials, the model is used to predict the samples used to make the libraries. This tests for uniqueness of classification. When ambiguities arise between materials, a method developer must consider the use of additional tests to confirm the identity of the material (i.e., using a simple pharmacopoeial test) or through the use of a hierarchical model that identifies known ambiguities and applies second- or third-level models to resolve the ambiguities.

To test the developed model's ability to assess future materials, a set of data, known as a test set, must be used that contains samples not included in the library. The library must also be tested against negative samples, that is, materials not included in the library to test for their rejection. Once the library has been validated in this way, it can be used as an alternative to pharmacopoeial monographs, as stated in the general chapters on alternative testing methods.

The biggest advantage of using the NIR method for raw material analysis is its objectiveness. If the library is based on a set of known, high-quality materials and has been validated using a suitable test set, the spectra are compared via a chemometric model, and the assessment is based on a statistical, not a visual, basis. A secondary but equally important advantage of NIR for raw material analysis is that the sample analyzed is not destroyed and can be kept as a retention sample for further evidence of material quality in the event an audit or if a process investigation is performed.

NIR is used widely as a raw material identification method, and in some companies the instrument is located outside of the QC laboratory, particularly at the point of receipt for instant lot rejection. Other organizations use NIR in the raw material dispensary as a 100% check of lots before they are used for release to production. Other applications of NIR include raw material conformity and process ability prediction. These methods determine whether the material falls within predefined limits (either based on standard deviation limits or multivariate limits) and are used in conjunction with identification results to partition raw material lots to specific products based on knowledge gained over the product history. This is a quality by design (QbD) approach to raw material and process understanding.

71.6.2.2 Intact Tablet Analysis The most common method used to analyze the active pharmaceutical ingredient (API) in tablets, capsules, and gels is by the use of high-performance liquid chromatography (HPLC). This method requires many preparation steps and instrument calibration to known standards before an analysis can be performed. In high-volume pharmaceutical plants, this can put significant pressure on a QC laboratory to release results to keep the manufacturing plant running in an efficient manner.

A number of instrument vendors provide sampling options for measuring pharmaceutical tablets in transmission mode. This can also be performed in diffuse reflectance mode when the API content is high (typically >10% w/w) and where it can be assumed that the API is distributed evenly over the tablet surface. This was previously discussed in Section 71.4.3.4.

In order to develop an intact tablet analysis method, a number of important steps must be followed. These are described as follows:

1. A feasibility study must be performed that shows that the NIR method is both specific and selective for the API over an extended concentration range (typically between 70 and 130% of product label claim).
2. If transmission mode is to be used for the analysis, the tablet thickness must be taken into consideration, and the spectral regions where the API absorbs must show a high-quality signal. Otherwise, if the API concentration is high enough, diffuse reflectance can be considered.
3. Once the feasibility and sampling methods have been performed, the sampling must be optimized, and a calibration set of samples must be chosen over the widest possible production time range of the product. These samples will be highly consistent due to the tight nature of pharmaceutical manufacturing, and they must be extended with samples prepared in the laboratory (using rational designed experiments) to achieve the 70–130% label claim range.
4. Laboratory samples must also be produced at the target label claim range, and all of the laboratory developed tablets should (wherever possible) be produced on the equipment used to make commercial batches.
5. A pool of samples is now developed with a suggested initial target of 180 individual samples. This set will be split 2 : 1 for calibration and test set samples, respectively. The calibration set will contain at least 90 production samples and 30 laboratory-made samples such that their distribution is boxcar in character. The test set will contain 45 production and 15 laboratory samples for the same reasons as for the calibration set.
6. Each sample is now scanned using the spectrometer, and the data are analyzed using PCA (or other techniques) to determine if there is a trend in the active ingredient in the region where the API absorbs and also to confirm that samples produced at the target API level have the same spectral characteristics as the production samples. If this is the case, the laboratory and production samples can be pooled together to make a robust calibration set. Otherwise, further work must be performed on the laboratory samples to make them equivalent to the production samples.
7. Each tablet is now analyzed using the validated reference method that has a known standard error of laboratory (SEL). This is the baseline standard that the NIR model must be compared to. According to PASG and EMA guidelines [1, 63], the standard errors of calibration and prediction (SEC and SEP, respectively) must be less than $1.4 \times \text{SEL}$.

8. Once the individual sample API concentrations have been matched to their NIR spectrum, a chemometric model (based on Partial Least Squares Regression (PLSR), Principal Component Regression (PCR), or MLR) is developed, and the Standard Errors for calibration are compared to the SEL. It is important to note here that the SEC/SEP values can be improved by adding more components to the model; however, the number of components must be commensurate with the complexity of the system. Typically, 1–5 components are the usual number. Since all loadings, loading weights, and regression coefficients of the models must be given a physical interpretation, the use of smaller component numbers is suggested; otherwise the model cannot be used.
9. Once the model has been developed and internally validated using an optimized preprocessing method and test set validation, the model must be applied to new samples to test its ability in a real situation. Since the predicted values of the new samples should span a tight region, the residuals of the predicted values and reference values should be assessed for normality and also to test that no residual exceeds ± 3 standard deviations.
10. Once all of the validation procedures have been followed, the method is ready to deploy for real situation usage.

The previous steps are intended to be a guide for a user to follow when developing a new intact tablet (or any method in general). There will always be exceptions or additional steps required, but if this process is followed, reliable methods can be developed with full traceability. The articles by Ritchie [20, 21] and Moffat [64] are also an excellent source of reference for the development of quantitative methods in the pharmaceutical industry. Figure 71.39 shows the typical output of a chemometric package for the development of an intact tablet method.

Recently, a number of instrument vendors have utilized the diffuse reflectance sampling method by installing it into the tablet press and measuring 100% of tablets as they are pressed. This is a very high-speed application and can be used for trending purposes only. Other systems are available that automatically take a random tablet off the production line and measure a number of properties of the tablet, including NIR analysis for a more accurate assessment of the product state during tablet compression.

71.6.2.3 Process Analytical Technology The fundamental premise of the PAT initiative is to gain understanding of the process at the point of manufacture using state-of-the-art, modern process analysis technologies. NIR is an excellent example of a PAT. The following provides a brief list of applications where NIR spectroscopy has been used as a PAT:

1. Granulation: The use of NIR for granulation has been met with success and failure over the years. Optimal placement of probes into a high shear granulator is the key to success, where the measurement of moisture/solvent uniformity can be assessed. The overall granule size and distribution are of key importance. NIR

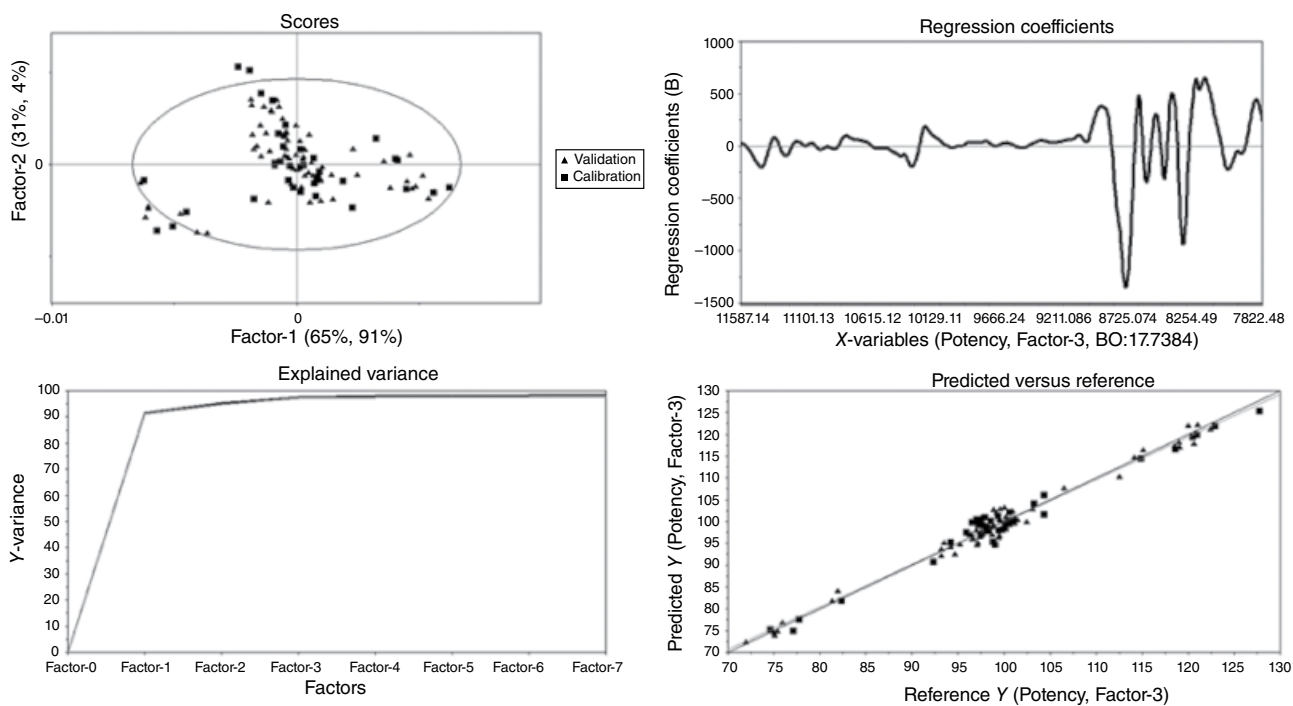


FIGURE 71.39 Results generated by a chemometrics package for the development of a quantitative intact tablet analysis method.

measurements are often supplemented with other measurement systems, including in-line particle size analyzers to provide a measure of uniformity and distribution simultaneously. With the current push by industry to adopt CM [18], granulation has changed from a batch process to a continuous one. This has simplified the measurement of granulation through the use of an analyzer at a fixed point where near 100% measurement of the granulation can be obtained. This all relates to the US FDA's push for QbD and the ability to adjust and adapt a process based on real-time measurement systems.

2. Fluid bed drying (FBD): NIR instrumentation can be integrated directly into an FBD operation to detect drying end points. This avoids overdrying of the granules and thus less potential problems during milling and compression operations. Models can be developed quantitatively (using procedures similar to those outlined in Section 71.6.2.2) or qualitatively by trend analysis, with conformity of PCA-based methods (including batch modeling). Refer to the literature for more information on this application [65].
3. Blend monitoring: NIR analyzers have been developed that integrate directly with tumble and continuous blender systems. They are designed to utilize wireless communication systems, and in the case of a tumble blender, each time the blender rotates a gravity switch or accelerometer triggers the instrument to take a scan when the powder is completely covering the sampling window. This application has been discussed by many authors in the literature [6].
4. Tablet compression: A number of instrument vendors have developed systems for the monitoring of tablets as they are being compressed. In one configuration, the instrument is placed just above the tablet ejection area, and diffuse reflectance measurements are made on tablets in a 100% inspection system. Another configuration utilizes an at-line approach that measures tablets for certain properties at regular intervals. This system was already briefly described in Section 71.6.2.2.
5. Tablet/granule coating: The thickness of a tablet coating is often a critical quality attribute of a pharmaceutical solid dose form for two main reasons. Firstly, for cosmetic reasons, tablet coating provides a quality assurance aspect to the end user when they see tablets evenly coated and thus are acceptable for use. Secondly, coatings may be functional, that is, they contain slow-release properties that may be absorbed by the different parts of the body. Functional coatings may also be applied to granules to achieve differing release profiles. NIR finds use in this application in fluidized bed and pan coaters. The placement of the NIR probe is again the critical factor in coating operations. The probe must be placed where it cannot be coated itself. As the coating builds up on a tablet or granule, the spectral properties of the uncoated tablet change to those of the coating material. Multivariate methods can then be used to classify tablets as being coated or uncoated, and in some cases, the coating thickness can also be monitored.

Figure 71.40 shows the changes in spectral character from coated to noncoated for selected samples during a tablet coating process. The figure also shows how the process can be monitored and controlled multivariately.

6. Packaging: NIR instrumentation can be placed in line with visual inspection tools to determine if the tablet not only has the right appearance but also has the right chemical composition. Samples moving through a packaging system do so at a fast rate; therefore, instrumentation has been developed that can measure 100% of tablets at a high rate and provide the required analysis of surface chemistry. These systems, although they do not provide any intrinsic process understanding in terms of tablet properties and functionality, are often overlooked as “nice to haves” in an operation; however, the detection of one different tablet in a package could be the difference between life and death for a patient in some cases, and the cost of product recall can be much greater than the investment cost of the packaging inspection system in the first place.

It is noted here that the trend toward 100% inspection tools is currently being driven by vendors and proponents of CM systems. A CM manufacturing line is an integrated manufacturing process, joining all of the unit operations together and using PAT to monitor and control the processes, ensuring that the quality of an intermediate leaving one operation is suitable for processing in the next operation. Another initiative in the

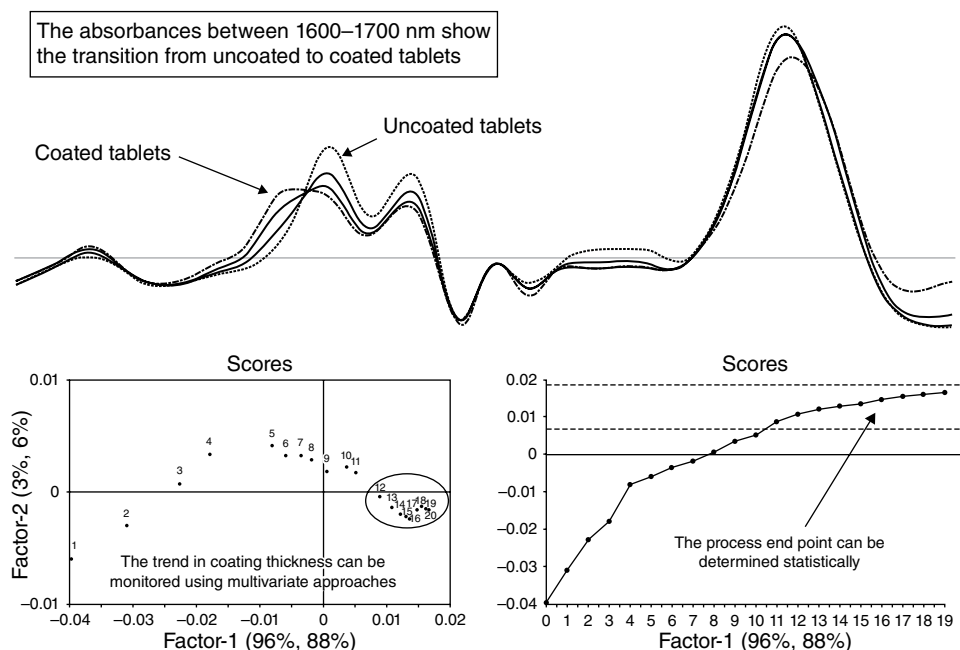


FIGURE 71.40 Spectral changes observed by NIR and monitored multivariately for a tablet coating operation.

pharmaceutical industry is known as QbD. A QbD process is designed to be adaptable to changing material attributes. This topic is briefly discussed in the review by Swarbrick [18], and the interested reader is referred to this article and the references in it.

71.6.2.4 Bioprocesses In recent times, the ability to measure bioprocesses has gained much interest, particularly fermentation reactions where the product value is high and the process can take up to several weeks to complete. There has been much effort directed toward the monitoring of such processes using a combination of process input values and PAT outputs for tracking the trajectory of the so-called golden batch [66].

Many of the bioprocesses currently employed in the biopharmaceutical industry are highly aqueous in nature, and this poses a problem for NIR when the concentration of the proteins being measured becomes too dilute. However, when a specific region has a good S/N ratio, the NIR method can be used to either predict constituent values during the progress of the reaction (by use of hierarchical modeling approaches) or the scores obtained by PCA projection can be combined with process data (using a process called data fusion) to monitor the processes via a physicochemical modeling approach [66]. The major benefit methods such as NIR provides to biopharmaceutical manufacturers are that the process can be controlled analytically and adjustments made to it, as required to bring it back on trajectory (see Figure 71.41 for an example).

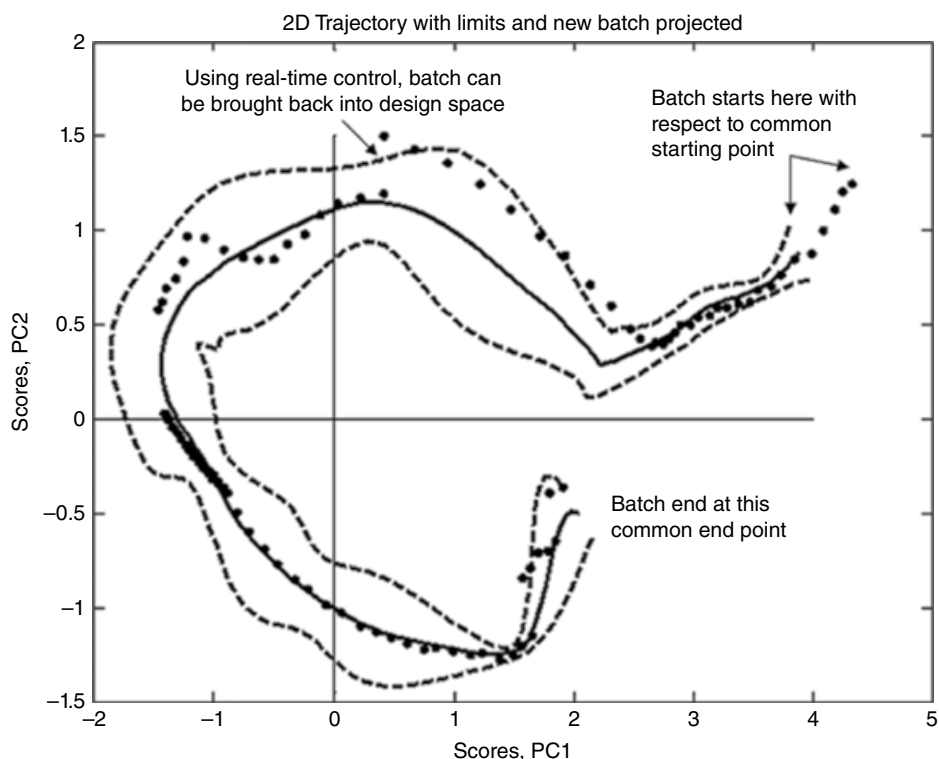


FIGURE 71.41 Process output design space for a fermentation reaction monitored using PAT.

The production of biofuels is also an application where NIR has found suitable application. Since the main product being produced is ethanol, monitoring the evolution of the O—H absorbance makes for an excellent tool for predicting yield and other quality parameters.

71.6.3 Applications in the Petrochemical and Refining Sectors

An oil refinery takes crude oil (a highly complex natural substance) and through the processes of distillation, cracking, isomerization, and alkylation generates base feedstock for many different common products used in everyday lives. Some of the products go to gasoline manufacture; others are sent for polymer production, and NIR has also played a part in the monitoring of such processes, bringing about efficiencies that were not possible without the use of the method.

71.6.3.1 Gasoline Blending With the quality of crude oil declining and the need for more efficiency in the oil refining sector, waste versus profit is a big concern (as it is for any industry). It has been quoted that every 0.2 octane numbers given away can result in a refinery losing anywhere between 50 and 150 K USD per day, based on the refinery output [67].

NIR has long been used to measure multiple properties of gasoline and diesel blends in a manner that reduces such octane (or cetane) giveaway. Depending on the crude oil quality, refineries produce intermediate gasoline products of various octane number values, in general:

1. Light straight run (LSR): This is formed primarily of straight chain alkanes of low octane number. Refineries aim to put as much of this intermediate into final blends as possible as it is a cheap component and can occupy storage tank space quickly if it is not utilized.
2. Aromatics: These are represented mainly by compounds such as benzene, toluene, and xylenes. They are particularly high in octane number, but due to environmental and health organization regulations, there is only a limited amount of aromatics that can be utilized in a gasoline blend. Refineries also like to add aromatics to a gasoline blend as much as legally possible because of their high octane numbers.
3. Isomerates: These are formed by the chemical reorganization of simple alkanes of lower octane number into highly branched compounds that can increase the octane number with respect to the starting materials. These are added to LSR to improve the rating of the final gasoline product.
4. Alkylates: These are highly branched aliphatic compounds that are formed by the reaction of smaller aliphatics (mainly butane) using concentrated hydrofluoric or sulfuric acid. This makes their production a highly dangerous (and expensive)

process, and these compounds have very high octane numbers. They are used in premium-grade gasoline because they provide high octane numbers with no aromatics to worry about.

The task of the refinery now is a balancing act, adds as much LSR and aromatics (within legal bounds), and minimizes the use of isomerates and alkylates such that gasoline of adequate performance characteristics can be produced. NIR has been used both in line and at line for many years to monitor blending of gasoline. Since the monitoring can be performed in real time, some blending systems can adjust the rates of the various components to minimize octane giveaway.

Octane number is not the only property that can be measured; others include benzene content, vapor pressure, alkene content, and many other properties. The analysis is simple; using a liquid transmission device for a laboratory-based system, a sample is typically measured as is, and analysis report is generated. However, when applied in or on line, the need for sample conditioning is sometimes required to remove excess water buildup that may influence the predicted results after application of chemometric models to the NIR spectra. Figure 71.42 shows a general implementation of NIR into a gasoline blending line for monitoring and control of the process.

71.6.3.2 Other Processes There are potentially many uses of NIR spectroscopy for the analysis, monitoring, and control of petrochemical processes. The primary

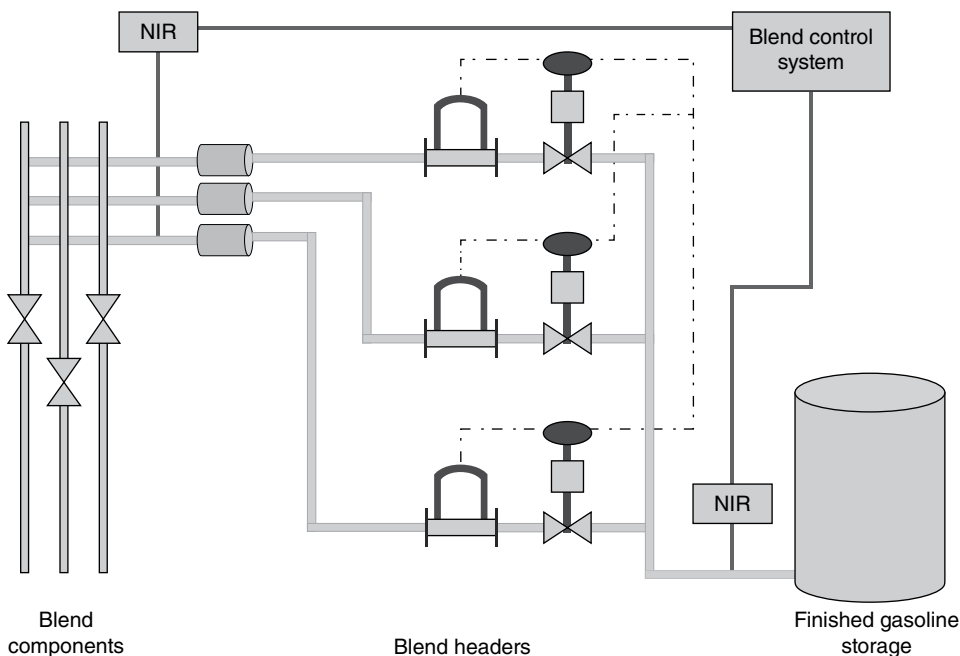


FIGURE 71.42 General implementation of NIR into a gasoline blending operation.

application that dominates in the industry is gasoline blending, but other applications include:

- Measurement of moisture content in hydrofluoric acid for alkylation.
- Measurement of crude oil properties during extraction (particularly for oil sands operations).
- Measurement of input feed, intermediate, and final product output from fluid catalytic cracking (FCC) operations.
- Measurement of sample outputs from the various stages of atmospheric distillation operations allowing assessment of the quality of the various cuts from the crude oil input.

As with the biopharmaceutical applications discussed in Section 71.6.2.4, NIR can be used to monitor chemical reactions in line for the production of specialty chemicals at a petrochemical plant, in particular the production of polymers and related products. The general approach is similar in all industries and applications, find an application suitable for NIR, optimize sampling, and monitor and control in real time.

71.6.4 Applications in the Food and Beverage Industries

Food and beverage applications of NIR are very similar to those discussed in the pharmaceutical industry (Section 71.6.2.). These include but are not limited to:

- Raw material identification and quality, particularly for flours used in baking and spices used in specialty mixtures.
- Product blending of cake and bread mixes and the exact combination of spices for repeatable product quality when used by consumers.
- Monitoring of moisture and food ingredient uniformity, particularly for dough mixing in breads and cake mixtures.

As these situations are very similar in nature to pharmaceutical operations, the reader is referred to the section on pharmaceutical applications for more details. This section will provide a short introduction to the use of NIR for food authentication.

Consumers around the world have come to expect the highest standards in food quality, and in particular, if they purchase what is said to be authentic on the product label, they expect it to be that product. However, in today's society, there is a minor element who would substitute nonauthentic foods as authentic to gain superior profits from inferior products. NIR has been used for the authentication of foods and wines for a number of years now.

Gerard Downey of the Agriculture and Food Development Authority in Ireland (Teagasc) has been active for many years in the research of NIR for food authentication. The following provides a short list of the applications where NIR has been successfully used:

- Adulteration of olive oil with sunflower oil in Mediterranean countries [68]
- Authentication and origin of honey [69]
- Authentication of varietal origins for the classification of commercial white wines [70]
- Authentication of coffee bean varieties [71]

The previous list of selected applications shows that the main areas of food authentication are centered around luxury and fine foods. With the introduction of micro-NIR systems, the future may see the development of consumer-based NIR devices attached to smartphones that will be able to perform these types of analyses at the point of sale and verify that what is being paid for is what is being received.

71.7 SUMMARY AND FUTURE PERSPECTIVES

NIR instrumentation has seen a rapid rise in development and application particularly since the mid 1990s when computing power and chemometric methods allowed for the rapid capture and modeling of data for predictive and classification purposes.

Initially thought of as a tool for agricultural applications due to its nondestructive nature and ability to handle heterogeneous samples, other industries, particularly the pharmaceutical/biopharmaceutical and petrochemical industries, have taken full advantage of the in-line implementation options available.

In the pharmaceutical industry, the PAT initiative in its early days was based primarily around NIR for at-line and in-line applications. Other industries such as the food and beverage, fine chemical, textile, and renewable energy sectors have also utilized the benefits of NIR for business critical applications.

The future of NIR is looking very bright with the introduction of microspectrometers. These devices will find use in consumer applications where they will be integrated with mobile devices such as smartphones for applications such as personalized medicine, food authentication, and even gasoline integrity. By integrating these microanalyzers with smartphones, applications can be written that will access libraries over the Internet for the rapid identification of materials in an easy and reliable manner. This opens up many applications for national security, food integrity, process analysis, and medical applications. This vision of the future is the same vision Hirschfeld shared in his 1985 paper [24].

71.8 TERMINOLOGY

ANISG	Australian Near Infrared Spectroscopy Group.
ANN	Artificial neural network refers to a set of nonlinear mathematical techniques used for the development of classification and prediction models based on machine learning algorithms.
AOTF	Acoustooptical tunable filter refers to a wavelength dispersion system that uses a tellurium oxide crystal and a radio-frequency source to selectively tune the crystal to measure sharp spectral bands in the NIR region.
API	Active pharmaceutical ingredient refers to the active component(s) in a pharmaceutical formulation.
ASTM	American Society for Testing and Materials.
ATR	Attenuated total reflectance refers to a sampling device that utilizes a chemically inert crystal of precise design to make close contact with a sample surface. When electromagnetic radiation is passed through the crystal, a series of internal reflections are set up at the crystal boundary. The light at the surface of the crystal interacts with the sample, and the chemical information is transmitted back to the detector to generate an NIR spectrum.
AU	Absorbance units refers to a unitless measure of spectral intensity measured on a logarithmic scale.
CM	Continuous manufacturing refers to a flow production method used to manufacture, produce, or process materials without interruption.
EMA	European Medicines Agency.
EMSC	Extended Multiplicative Scatter/Signal Correction refers to a scatter correction technique that uses MSC as its foundation and extends it with wavelength-dependent correction terms and can also be extended to include the use of good and bad spectra, thus making the correction a physicochemical model.
FBD	Fluid bed drying refers to a process where air of controlled temperature is passed through the bottom of a bed of wet powder in order to create a fluidized motion that simultaneously dries and regulates granule size.
FCC	Fluid catalytic cracking refers to a process used in the petrochemical industry for taking heavy vacuum distillates and cleaving them through catalytic processes into smaller, more volatile fractions for use in gasoline blending.
FFT	Fast Fourier transform refers to a mathematical procedure that converts a time-domain signal into a frequency-domain signal. In the case of NIR, the frequency-domain signal is a near-infrared spectrum.
FIR	Far infrared refers to the region of the electromagnetic spectrum between 16,000 and 111,000 nm.

FT-NIR	Fourier transform near-infrared spectroscopy refers to a method of spectral collection that utilizes an interferometer as the wavelength dispersion device. The signal sent to the detector is a time-dependent signal. The instrument uses a fast Fourier transform to convert the time-domain signal into a frequency-domain signal, thereby generating a near-infrared spectrum.
FWHM	Full Width at Half Maximum refers to the width of a spectral band at the half-way point on the y-axis. This is used to understand the size of any smoothing window used in derivatives such that the band is not overly distorted or smoothed out.
GMP	Good Manufacturing Practice refers to a set of guidelines used to ensure that a manufacturer consistently meets its quality and safety targets for the products it manufactures.
GUI	Graphical user interface refers to a system that displays the results generated by an instrument in a form that is interpretable to an end user.
HPLC	High-performance liquid chromatography refers to an analytical procedure that separates and quantifies specific components of a mixture using a purpose-designed separation column and an appropriate mobile phase to affect the separation.
ICNIRS	International Conference on Near Infrared Spectroscopy
IR	Infrared refers to any part of the electromagnetic spectrum that produces thermal energy, that is, NIR, MIR, and FIR.
LOD	Limit of detection refers to a statistical analysis that determines at which point a sample signal is distinguishable from background noise. This is typically measured as 3σ above the baseline noise.
LOQ	Limit of quantification refers to the statistical assessment of a sample signal that can be reliably used for building quantitative models. This is typically measured as 10σ above the baseline noise.
LSR	Light straight run refers to a gasoline cut typically of low boiling point and low octane number. This is utilized in gasoline blending to minimize the use of expensive fractions such as isomerates and alkylates.
LVF	Linear Variable Filter refers to a wavelength dispersion device that utilizes a wedge-shaped filter to separate polychromatic light into its monochromatic components, which are then detected on a diode array detector.
LW	Long wave refers to the region of the NIR spectrum between 1100 and 2500 nm. This is primarily where diffuse reflectance measurements are performed.
MEMS	Microelectromechanical spectrometer refers to a wavelength dispersion system that uses two parallel dielectric mirrors, built on a microchip substrate to generate a tunable Fabry–Perot filter. Through the application of electrical voltage, the distance between the dielectric mirrors can be controlled resulting in the generation of an NIR spectrum.

MIR	Mid-infrared refers to the region of the electromagnetic spectrum between 2,500 and 16,000 nm.
MLR	Multiple linear regression refers to a multivariate regression technique that fits a small number of terms (typically <20) in order to fit a linear model that can be used to predict properties of new samples when measured by the variable in the model.
MSC	Multiplicative Scatter Correction, sometimes referred to a Multiplicative Signal Correction, refers to a scatter correction method that corrects for both additive and multiplicative effects by normalizing data to a mean or reference spectrum.
MVA	Multivariate analysis refers to a set of mathematical tools used to analyze more than one variable simultaneously in order to understand variable and sample relationships and also to build regression and classification models that utilize the most important parts of a data set.
NIR	Near infrared refers to the region of the electromagnetic spectrum between 700 and 2500 nm.
NMR	Nuclear magnetic resonance refers to a type of spectroscopy that utilizes a strong magnetic field and radio wave pulses to excite typically protons in a molecule and measure their relaxation times. The result is a highly detailed spectrum that allows an analyst to structurally elucidate the molecule in the sample.
PASG	Pharmaceutical Analytical Sciences Group.
PAT	Process analytical technology refers to an initiative of the US FDA that encourages pharmaceutical and related industries to adopt novel and state-of-the-art process technologies for better understanding of processes and products. NIR has been the major PAT tool used to date.
PCA	Principal Component Analysis refers to an exploratory data analysis method used to understand complex sample and variable relationships in multivariate data. The aim of PCA is to isolate those variables that most contribute to sample patterns observed. PCA models can be further used for developing classification models or predicting the state of new samples with respect to the original model.
PCR	Principal Component Regression refers to a method for relating the variations in a response variable (<i>Y</i> -variable) to the variations of several predictors (<i>X</i> -variables). This method performs particularly well when the various <i>X</i> -variables express common information, that is, when there is a large amount of correlation. Principal Component Regression is a two-step method. First, a Principal Component Analysis is carried out on the <i>X</i> -variables. The principal components are then used as predictors in a multiple linear regression.
PLSR	Partial Least Squares Regression refers to a method for relating the variations in one or several response variables (<i>Y</i> -variables) to the variations of several predictors (<i>X</i> -variables). This method performs particularly

	well when the various <i>X</i> -variables express common information, that is, when there is a large amount of correlation between the variables.
QbD	Quality by design refers to a global pharmaceutical initiative that builds quality into the process such that the overall integrity of a batch can be assured through direct process monitoring and control. Changes to the process are allowed without any deviation reports being generated as long as the process is maintained in its design space. NIR is a PAT tool that enables organizations to monitor and control processes within the design space.
QC	Quality control refers to protocols used to determine the fitness for purpose of a product before it is released to the next step of its life cycle.
RF	Radio frequency refers to electromagnetic waves generated in the radio wave region of the spectrum, typically of wavelengths between centimeters and meters.
SEC	Standard error of calibration refers to the variation in the precision of calibration sample predictions over several samples. SEC is computed as the standard deviation of the prediction residuals and is dependent of the validation method used.
SEL	Standard error of laboratory refers to the baseline precision of a reference analytical method used to calibrate a secondary method such as NIR. The precision of the secondary method must be statistically equivalent to the SEL in order for it to be used as an alternative method.
SEP	Standard error of prediction refers to the variation in the precision of predictions over several samples. SEP is computed as the standard deviation of the prediction residuals when test set validation is the validation method used.
S/N	Signal-to-noise ratio refers to that part of a measurement signal that can be statistically differentiated from baseline noise. This is often used in limit of detection (LOD) and limit of quantification (LOQ) studies.
SNV	Standard Normal Variate refers to a scatter correction technique commonly used in NIR spectroscopy. It used the spectrum itself to correct for both additive and multiplicative effects by subtracting the grand spectrum mean from all points and dividing them by the spectra overall standard deviation. The end result is a spectrum with its center of gravity at zero and a spread of absorbances between ± 3 standard deviations around the mean.
SW	Short wave refers to the region of the NIR spectrum between 700 and 1100 nm. Sometimes called the Herschel region, this is primarily the region where transmission spectra are collected.
TOS	Theory of Sampling refers to a set of protocols that ensure representative sampling techniques are implemented and adhered to for a particular measurement system.
USDA	US Department of Agriculture.
US FDA	US Food and Drug Administration.

REFERENCES

1. European Medicines Agency, "Guideline on the Use of Near Infrared Spectroscopy (NIRS) by the Pharmaceutical Industry and the Data Requirements for New Submissions and Variations", EMEA/CHMP/CVMP/QWP/17760/2009 Rev2, 2014.
2. United States Food and Drug Administration, "Development and Submission of Near Infrared Analytical Procedures, Guidance for Industry", <http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM440247.pdf> accessed June 3, 2015.
3. Richardson, A. D., Reeves III, J. B., and Gregoire, T. G., "Multivariate analyses of visible/near infrared (VIS/NIR) absorbance spectra reveal underlying spectral differences among dried, ground conifer needle samples from different growth environments", *New Phytol.*, 161(1), 2004, pp 291–301.
4. Curcio, J. A. and Petty, C. C., "The near infrared absorption spectrum of liquid water", *J. Opt. Soc. Am.*, 41(5), 1951, pp 302.
5. Herschel, W., "Investigation of the powers of the prismatic colours to heat and illuminate objects; with remarks, that prove the different refrangibility of radiant heat. To which is added, an inquiry into the method of viewing the sun advantageously, with telescopes of large apertures and high magnifying powers", *Phil. Trans. R. Soc.*, 90, 1800, pp 225.
6. Ciurczak, E. W. and Drennen III, J. K., *Pharmaceutical and Medical Applications of Near Infrared Spectroscopy*, Practical Spectroscopy Series Vol 31, Marcel Dekker Inc., New York, 2002, pp 1.
7. Hindle, P. H., "Historical Development", in *Handbook of Near Infrared Analysis*, 3rd Ed, eds. Burns, D. A. and Ciurczak, E. W., CRC Press, Boca Raton, FL, 2008, pp 3.
8. Davies, A. M. C., "An Introduction to Near Infrared Spectroscopy", <https://www.impublications.com/content/introduction-near-infrared-nir-spectroscopy> accessed June 3, 2015.
9. Coblenz, W. W., *Investigation of Infrared Spectra, Part I*, Carnegie Institution, Washington, DC, 1905.
10. Ben-Gera, I. and Norris, K., "Direct spectrophotometric determination of fat and moisture in meat products", *J. Food Sci.*, 33, 1968, pp 64.
11. Burns, D. A. and Margoshes, M., "Historical Development", in *Handbook of Near Infrared Spectroscopy*, Practical Spectroscopy Series Vol 13, eds. Burns, D. A. and Ciurczak, E. W., Marcel Dekker Inc., New York, 1992, pp 3.
12. Williams, P. C., "Implementation of Near-Infrared Spectroscopy", in *Near Infrared Technology in the Agricultural and Food Industries*, 2nd Ed, eds. Norris, K. and Williams, P. C., AACC, Saint Paul, MN, 2001, pp 145–170.
13. Osborne, B. G., "Principles and practice of near-infrared (NIR) reflectance analysis", *Int. J. Food. Sci. Tech.*, 16(1), 1981, pp 13–19.
14. Shenk, J. S. and Westerhaus, M. O., "Population definition, sample selection, and calibration procedures for near infrared reflectance spectroscopy", *Crop Sci.*, 31(2), 1990, pp 469–474.
15. Flinn, P. C. and Downes, G. J., "The importance of Near Infrared Spectroscopy in Deciding Appropriate Feeding Strategies for Australian Livestock", in *Near Infrared Spectroscopy*:

- the Future Waves*, eds. Davies, A. M. C. and Williams, P., NIR Publications, Chichester, 1996, pp 512.
16. Blakeney, A. B. and Flinn, P. C., "Determination of non-starch polysaccharides in cereal grains with near-infrared reflectance spectroscopy", *Mol. Nut. Food. Res.*, 49(6), 2005, pp 546–550.
 17. Batten, G. D., "Forestry and the Environment: Challenges for Near Infrared Spectroscopy", in *Near Infrared Spectroscopy: proceedings of the 11th International Conference*, eds. Davies, A. M. C. and Garrido-Varo, A., NIR Publications, Chichester, 2004, pp 749.
 18. Swarbrick, B., "Review: advances in instrumental technology, industry guidance and data management systems enabling the widespread use of near infrared spectroscopy in the pharmaceutical/biopharmaceutical sector", *J. Near Infrared Spectrosc.*, 22(3), 2014, pp 157–168.
 19. Mark, H., *Principles and Practice of Spectroscopic Calibration*, John Wiley & Sons, New York, 1991.
 20. Mark, H., Ritchie, G. E., Roller, R. W., Ciurczak, E. W., Tso, C., and MacDonald, S. A., "Validation of a near-infrared transmission spectroscopic procedure, part A: validation protocols", *J. Pharm. Biomed. Anal.*, 28(2), 2002, pp 251–260.
 21. Ritchie, G. E., Roller, R. W., Ciurczak, E. W., Mark, H., Tso, C., and MacDonald, S. A., "Validation of a near-infrared transmission spectroscopic procedure, part B: application to alternate content uniformity and release assay methods for pharmaceutical solid dosage forms", *J. Pharm. Biomed. Anal.*, 29(1–2), 2002, pp 159–271.
 22. International Council for Near Infrared Spectroscopy (ICNIRS), <http://icnirs.org/> accessed June 3, 2015.
 23. Shapiro, H. M., "In memoriam, Tomas Hirschfeld (1939–1986)", *Cytometry*, 7(5), 1986, pp 399.
 24. Hirschfeld, T., "Instrumentation in the next decade", *Science*, 230, 1985, pp 286–291.
 25. The Australian Near Infrared Spectroscopy Group, <http://www.anisg.com.au/> accessed June 3, 2015.
 26. Norris, K. and Williams, P. C., *Near Infrared Technology in the Agricultural and Food Industries*, 2nd Ed, AACC, Saint Paul, MN, 2001.
 27. McClure, W. F. and Tsuchikawa, S., "Instruments", in *Near-Infrared Spectroscopy in Food Science and Technology*, eds. Yukihiro Ozaki, W. F. M. and Alfred, A. C., John Wiley & Sons, Inc., Hoboken, NJ, 2006, pp 75–108.
 28. Skoog, D. A., West, D. M., and Holler, F. J., *Fundamentals of Analytical Chemistry*, 5th Ed, W. B. Saunders Company, New York, 1988, pp 466–472.
 29. Tipler, P. A., *Physics*, 2nd Ed, Worth Publishers Inc., New York, 1982, pp 921–923.
 30. Crocombe, R. A., "Miniature optical spectrometers, Part III: conventional and laboratory near-infrared spectrometers", *Spectroscopy*, 23(5), 2008, pp 40.
 31. Jacquinet, P., "Caractères communs aux nouvelles méthodes de spectroscopie interférentielle; Facteur de mérite", *J. Phys. Radium*, 19, 1958, pp 233.
 32. Tipler, P. A., *Physics*, 2nd Ed, Worth Publishers Inc., New York, 1982, pp 905–906.
 33. Brown, S. D., "Signal Processing and Digital Filtering", in *Practical Guide to Chemometrics*, ed. Gemperline, P., Taylor & Francis Group, Boca Raton, FL, 2006, pp 389.

34. Fellgett, P., "A propos de la théorie du spectromètre interférentiel multiplex", *J. Phys. Radium*, 19, 1958, pp 187.
35. Connes, J. and Connes, P., "Near infrared planetary spectra by Fourier spectroscopy I, instruments and results", *J. Opt. Soc. Am.*, 56(7), 1966, pp 896.
36. Berntsson, O., Danielsson, L.-G., and Folestad, S., "Characterization of diffuse reflectance fiber probe sampling on moving solids using a Fourier transform near-infrared spectrometer", *Anal. Chim. Acta*, 431(1), 2001, pp 125.
37. IMP, "NIR Products at Pittcon 2015", <https://www.impublications.com/content/nir-products-pittcon-2015> accessed June 3, 2015.
38. Harris, S. E. and Wallace, R. W., "Acousto-optic tuneable filter", *J. Opt. Soc. Am.*, 59, 1969, pp 744.
39. Bragg, W. H. and Bragg, W. L., "The reflexion of X-rays by crystals", *Proc. R. Soc. Lond. A*, 88(605), 1913, pp 428–438.
40. Crocombe, R. A., "Miniature optical spectrometers: the art of the possible, Part IV: new near-infrared technologies and spectrometers", *Spectroscopy*, 23(6), 2009, pp 26.
41. Norris, K. H. and Hart, J. R., "Direct spectrophotometric determination of moisture content of grain and seeds", *J. Near Infrared Spectrosc.*, 4, 1996, pp 23–30.
42. O'Brien, N. A., Hulse, C. A., Friedrich, D. M., Van Milligen, F. J., von Gunten, M. K., Pfeifer, F., and Siesler, H. W., "*Miniature Near-Infrared (NIR) Spectrometer Engine for Handheld Applications*", eds. Druy, M. A. and Crocombe, R. A., Society of Photo-optical Instrumentation Engineers, Washington, DC, 2012, pp 837404-1-8.
43. Esbensen, K. H. and Paasch-Mortensen, P., "Process Sampling: Theory of Sampling, the Missing Link in Process Analytical Technologies (PAT)", in *Process Analytical Technology*, ed. Bakeev, K., John Wiley & Sons, Ltd, Chichester, 2010, pp 37–80.
44. Savitzky, A. and Golay, M. J. E., "Smoothing and differentiation of data by simplified least squares procedures", *Anal. Chem.*, 36, 1964, pp 1627–1639.
45. Steiner, J., Termonia, Y., and Deltour, J., "Comments on smoothing and differentiation of data by simplified least squares procedure", *Anal. Chem.*, 44, 1972, pp 1906–1909.
46. Hopkins, D. W., "What is a Norris derivative?", *NIR News*, 12(3), 2001, pp 3–5.
47. Norris, K., "Applying Norris derivatives. Understanding and correcting the factors which affect diffuse transmittance spectra", *NIR News*, 12(3), 2001, pp 6.
48. Barnes, R. J., Dhanoa, M. S., and Lister, S. J., "Standard normal variate transformation and de-trending of near-infrared diffuse reflectance spectra", *Appl. Spectrosc.*, 43(5), 1989, pp 772–777.
49. Martens, H., Jensen, S. Å., and Geladi, P., *Multivariate linearity transformation for near-infrared reflectance spectrometry* in Proc. Nordic. Symp. Appl. Stat., Stockland Forlag Publ, Stavanger, Norway, 1983, pp 205–234.
50. Martens, H. and Stark, E., "Extended multiplicative signal correction and spectral interference subtraction: new preprocessing methods for near infrared spectroscopy", *J. Pharm. Biomed. Anal.*, 9, 1991, pp 625–635.
51. Martens, H. Nielsen, J. P., and Engelsen, S. B., "Light scattering and light absorbance separated by extended multiplicative signal correction. Application to near-infrared transmission analysis of powder mixtures", *Anal. Chem.*, 75, 2003, pp 394–404.

52. Gemperline, P., "*Practical Guide to Chemometrics*", 2nd Ed, Taylor & Francis Group, Boca Raton, FL, 2006.
53. Hastie, T., Tibshirani, R., and Friedman, J., "*The Elements of Statistical Learning, Data Mining, Inference, and Prediction*", 2nd Ed, Springer Science+ Business Media LLC, New York, 2009, pp 389–416.
54. Flinn, P., "NIR: A Key Component of the Premium Grains for Livestock Project", Grain Industries Center for NIR, 6th Annual Meeting for Participants, Canberra, 2001, pp 32.
55. Fox, G. and Manley, M., "Applications of single kernel conventional and hyperspectral imaging near infrared spectroscopy in cereals", *J. Sci. Food Agric.*, 94(2), 2014, pp 174–179.
56. Batten, G. D., Blakeney, A. B., Glennie-Holmes, M., Henry, R. J., McCaffery, A. C., Bacon, P. E., and Heenan, D. P., "Rapid determination of shoot nitrogen status in rice using near infrared reflectance spectroscopy", *J. Sci. Food Agric.*, 54(2), 1991, pp 191–197.
57. Dambergs, R. G., Cozzolino, D., Cynkar, W. U., Esler, M. B., Janik, L. J., Francis, I. L., and Gishen, M., "Strategies to Minimise Matrix-Related Error with Near Infrared Analysis of Wine Grape Quality Parameters", in *Near Infrared Spectroscopy: proceedings of the 11th International Conference*, eds. Davies, A. M. C. and Garrido-Varo, A., NIR Publications, Chichester, 2004, pp 183.
58. McBratney, A., Whelan, B., and Ancev, T., "Future directions of precision agriculture", *Precis. Agric.*, 6, 2005, pp 7–23.
59. JEFFRESS, "PRODUCTS—Cane Analyser—InfraCana® II IC02", http://www.jeffress.com.au/products_ic02.htm, accessed June 3, 2015.
60. US FDA, "Pharmaceutical CGMPs for the 21st Century—A Risk Based Approach, Final Report", 2004, <http://www.fda.gov/drugs/developmentapprovalprocess/manufacturing/questionsandanswersoncurrentgoodmanufacturingpracticescgmppfordrugs/ucm137175.htm> accessed June 3, 2015.
61. US FDA, "Guidance for Industry: PAT—a Framework for Innovative Pharmaceutical Manufacturing and Quality Assurance", 2004, <http://www.fda.gov/downloads/drugs/guidancecomplianceregulatoryinformation/guidances/ucm070305.pdf> accessed June 3, 2015.
62. Ciurczak, E. W., "Following the progress of a pharmaceutical mixing study via near-infrared spectroscopy", *Pharm. Tech.*, 15(9), 1991, pp 141.
63. Broad, N., Graham, P., Hailey, P., Hardy, A., Holland, S., Hughes, S., Lee, D., Prebble, K., Salton, N., and Warren, P., "Guidelines for the Development and Validation of Near-Infrared Spectroscopic Methods in the Pharmaceutical Industry", in *Handbook of Vibrational Spectroscopy*, eds. Chalmers, J. M. and Griffiths, P. R., John Wiley & Sons, New York, 2002.
64. Moffat, A. C., Trafford, A. D., Jee, R. D., and Graham, P., "Meeting the international conference on harmonisation's guidelines on validation of analytical procedures: quantification as exemplified by a near-infrared reflectance assay of paracetamol in intact tablets", *Analyst*, 125, 2000, pp 1341.
65. Barla, V. S., Kumar, R., Nalluri, V. R., Gandhi, R. R. and Venkatesh, K., "A practical evaluation of qualitative and quantitative chemometric models for real-time monitoring of moisture content in a fluidized bed dryer using NIR technology", *J. Near Infrared Spectrosc.*, 22(3), 2014, pp 221–228.

66. Westad, F., Swarbrick, B., Gidskehaug, L., and Flaaten, G. R., "Assumption free modeling and monitoring of batch processes", *Chemom. Intell. Lab. Syst.*, 149, 2015, pp 66–72.
67. Emerson Process Management, "Refining", <http://www2.emersonprocess.com/siteadmincenter/PM%20Micro%20Motion%20Documents/Refining-Blending-PSG-MC-00796.pdf> accessed June 3, 2015.
68. Downey, G., McIntyre, P., and Davis, A. N., "Detecting and quantifying sunflower oil adulteration in extra virgin olive oils from the eastern mediterranean by visible and near-infrared spectroscopy", *J. Agric. Food Chem.*, 50(20), 2002, pp 5520–5525.
69. Hennessy, S., Downey, G., and O'Donnell, C., "Multivariate analysis of ATR/FT-IR spectroscopic data to confirm the origin of honeys", *Appl. Spectrosc.*, 62(10), 2008, pp 1115–1123.
70. Cozzolino, D., Smyth, H. E., and Gishen, M., "Feasibility study on the use of visible and near infrared spectroscopy together with chemometrics to discriminate between commercial white wines of different varietal origins", *J. Agric. Food Chem.*, 51, 2003, pp 7703–7708.
71. Downey, G. and Boussion, J., "Authentication of coffee bean variety by near-infrared reflectance spectroscopy of dried extract", *J. Sci. Food Agric.*, 71(1), 1996, pp 41–49.

NANOMATERIALS PROPERTIES

PAUL J. SIMMONDS

*Departments of Physics & Materials Science and Engineering, Boise State University,
Boise, ID, USA*

72.1 INTRODUCTION

In general, a nanomaterial is defined as possessing functional structural features that are less than 100 nanometers (nm) in size ($1\text{ nm} = 10^{-9}\text{ m}$). Nanomaterials exist widely in nature and give certain organisms their unique characteristics. For example, nanostructured hairs on the legs of spiders allow them to walk up walls, while waxy nanostructures give the leaves of plants such as the lotus their striking ability to repel water (Fig. 72.1).

It is, however, the recent development of artificial nanomaterials with tunable properties that has created so much interest and excitement in this field. Whole industries have grown up around nanomaterials, and entire university departments are devoted to their study. Nanomaterials are now essential to fields as disparate as telecommunications, medicine, renewable energy, and quantum computation. Arguably, materials containing structures on other length scales (e.g., milli- or micromaterials) have not made such a huge impact. So what is it about nanomaterials and their properties that make them so important? In order to answer this question, in this chapter we will explore the following three topics:

1. How and why certain properties of materials change as we reduce them in size from bulk (i.e., macroscopic scale) to the nanoscale.
2. How the behavior of certain materials is affected or dictated by the properties of the nanoscale features and structures that comprise them.
3. How nanomaterial properties can be controlled, for example, by engineering size, composition, or structure.

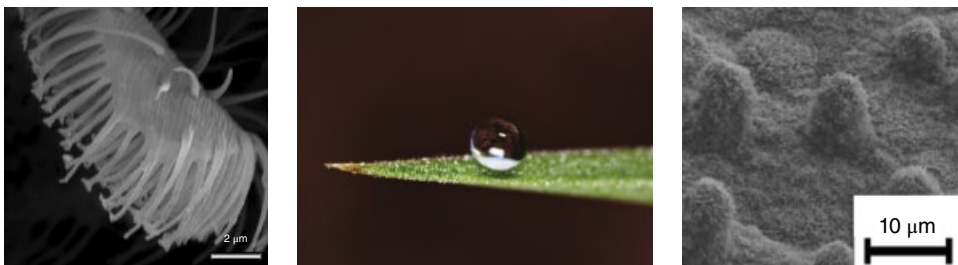


FIGURE 72.1 (Left) Hairs on the leg of a jumping spider, *Evarcha arcuata*. Each of these tiny hairs terminates in a contact pad nanostructure, which, by exploiting van der Waals attraction, allows the spider to crawl on ceilings and walls. Source: Kesel *et al.* [1]. Reproduced with permission of IOP Publishing. (Center) Water landing on the leaves of certain plants balls up into pseudospherical droplets that readily roll off, carrying with them any dirt particles that could otherwise inhibit photosynthesis. Source: http://upload.wikimedia.org/wikipedia/commons/8/8c/Dew_2.jpg, CC-BY-SA-3.0. (Right) This superhydrophobic property is a direct result of the presence of waxy nanostructures that cover the surface of the leaves: in this case, the lotus. Source: Cheng *et al.* [2]. Reproduced with permission of IOP Publishing.

Picture a cube of some material, say a metal or semiconductor, the sides of which are a few centimeters in size. We can easily pick up this macroscopic object and measure its various physical properties, for example, its Young's modulus, magnetic state, or melting point. Now imagine reducing the length of each side of that cube. First we would pass through the milliscale (10^{-3} m), followed by the microscale (10^{-6} m). Continuing to shrink the size of the cube, we would arrive at the *nanoscale*. At each scale we could repeat the measurements to see if any of the material's properties had changed.

What we would find is that many of the cube's material properties do not change in any meaningful way at either the milli- or microscale. This is not to say that there are not plenty of unique and important applications for milli- and micromaterials. In general, however, their properties are simply scaled down from the macroworld, and it is simply the fact that they are smaller that makes them useful. For example, the first *microchips* were created in the 1970s when transistors became small enough that they could be packed together in high-density arrays.

In contrast, upon reaching the nanoscale we see a much more abrupt and important shift in the cube's material properties. Certain characteristics become profoundly different from what they were when the object was larger, and the cube may take on new and sometimes surprising behaviors as a result. It is arguably these properties, that don't scale but change in a fundamental way when one descends to the nanoscale, which are the most interesting.

Due to the enormous breadth of modern nanomaterials, we cannot hope to discuss every family of nanomaterials in a single chapter. Inevitably this means that there are several major areas of current research that I have had to exclude in the interest of

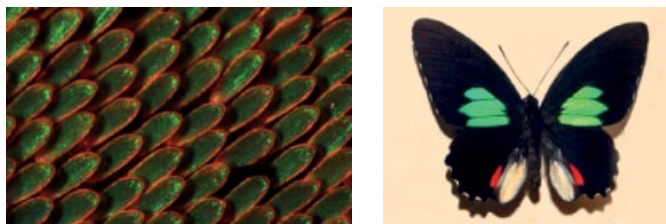


FIGURE 72.2 Naturally occurring photonic crystal nanostructures in the wings of a butterfly, *Parides sesostris zestos* (left). Source: Saranathan *et al.* [3]. Reproduced with permission of PNAS. The structure of this particular photonic crystal gives it the property of absorbing red and blue light, meaning that only green light is reflected. This gives the butterfly its colorful green markings, seen here (right) as three bright stripes on each wing. Source: http://commons.wikimedia.org/wiki/File%3APapilionidae_-_Parides_sesostris_zestos.JPG. CC-BY-SA-3.0.

brevity. Perhaps the most notable omission is the field of photonic bandgaps and optical metamaterials. These use periodic arrays of nanostructures to engineer the optical properties of a material, for example, its refractive index, reflectivity, or color. These take their cue from naturally occurring photonic structures in bird feathers and butterfly wings (Fig. 72.2). Nanostructures such as pillars or holes are created in arrays such that their separation is on the order of half the wavelength of the light of interest. The following references give an introduction to the exciting properties of these nanomaterials, further discussion of which is sadly outside of the scope of this chapter [4–9].

Rather than be an exhaustive list of nanomaterials, this chapter is instead intended to give a general overview of the most important properties that nanomaterials offer us. To do this, we will focus solely on those physical properties that change drastically as an object is reduced in size from the macroscale to the nanoscale. We now know, for example, that the physical and chemical properties of common materials such as carbon, silicon, or gold can be tuned simply by changing their size [10]. Color, magnetic and electronic behaviors, and even melting point can all be changed by shrinking a material down to the nanoscale.

We can broadly divide these significant changes in behavior at the nanoscale into two major groups. *The first group* contains those nanomaterial properties that come from a change in the relative influence of certain classical physical forces. For example, because of their low mass, gravity ceases to play a major role in the behavior of nano-scaled objects. In contrast, an enormous increase in the surface-to-volume ratio of nanomaterials means that electrostatic forces become extremely important. *The second group* contains those properties peculiar to nanomaterials that arise due to a fundamental shift in the way that nano-scaled objects interact with particles such as electrons and photons. As we will see, it is at the nanoscale that a quantum mechanical description of objects becomes most accurate. Compared with classical mechanics, with which we are

all well accustomed at the macroscale, quantum physics famously makes some strange predictions about the way tiny objects behave. It is this quantum framework for looking at the world at the nanoscale that helps explain some of the unexpected properties of nanomaterials.

72.2 THE RISE OF NANOMATERIALS

The recent explosion in nanomaterials research has been driven by some key inventions that allow us to create and look at nanoscaled objects. In the 1970s, the development of molecular beam epitaxy [11], and metal organic chemical vapor deposition [12, 13], made possible the growth of ultrahigh purity crystalline materials with atomic-level control over their thickness and other dimensions. Using increasingly short wavelengths of light, modern photolithographic methods can now create patterns on the surface of a material that have features only a few tens of nanometers in size, a limit that continues to fall [14–18]. Other synthesis approaches that have been critical at some stage in the development of nanomaterials include nanostructured masks; advanced wet, dry, and selective etching techniques; nanoimprinting; nanostructure self-assembly; and self-alignment. As we look at the properties of various nanomaterials in this chapter and discuss how they are made, it is important to distinguish between “top-down” and “bottom-up” approaches [19]. Top-down methods start with a featureless surface and remove material to create nanostructures. Bottom-up approaches start with a featureless surface and add material to create, or “grow,” the nanostructures.

The majority of technologically important nanomaterials are crystalline in nature, for example, most semiconductors and metals. Crystals form a subgroup of solid materials that are characterized by a highly ordered, periodic, symmetric atomic structure, where the interatomic distances are fixed and well known for a given material. Although noncrystalline nanomaterials exist (especially in nature), in the interests of space this chapter will mainly focus on examples of crystalline nanomaterials and their properties.

However, the ability to synthesize extremely high-quality materials with increasingly small sizes is useless unless we are able to look at and measure what we have made. To this end, some critical advances in microscopy were made during the 1980s with the advent of the scanning tunneling microscope and atomic force microscope. For the first time, these techniques allowed researchers to directly image and manipulate nanostructures or even individual atoms. Other key characterization techniques used (or sometimes specifically developed) to measure and interrogate the properties of nanomaterials include X-ray diffraction, photoluminescence, electroluminescence, transmission electron microscopy, and low-temperature magnetoresistance techniques. Many of these essential characterization tools and approaches are covered elsewhere in this book. Due to these tools and a whole host of

other discoveries in the burgeoning field of nanotechnology, nanostructures can now be controllably synthesized from an extensive range of materials for a wide variety of applications.

72.3 NANOMATERIAL PROPERTIES RESULTING FROM HIGH SURFACE-AREA-TO-VOLUME RATIO

72.3.1 The Importance of Surfaces in Nanomaterials

It is important to note that the chemistry deep within the interior of a solid material can be very different from its chemistry at the surface. In a typical crystalline material, atoms in the interior (or bulk) are stable and relatively inert. The distances to their nearest neighbors are well defined in every direction and they exist within a symmetric potential field. The situation at the surface of the material is very different. Think of creating a fresh surface by slicing through the bulk of the crystal to expose a single atomic layer. This newly exposed surface would initially have the same atomic arrangement as did the atoms in the bulk. However, creating this surface required the breaking of many atomic bonds, freeing up a large number of unpaired bonding electrons. These unpaired electrons, often referred to as “dangling bonds,” mean that the surface is extremely energetically unstable. To reduce this surface energy, the atoms almost immediately rearrange themselves into structures that are more stable, but also more complex, often involving double bonds or dimer pairs [20].

Although these new atomic reconstructions do help stabilize the surface, they also create a surface that is fundamentally distinct from the bulk for several reasons. First, the new surface reconstruction changes its chemical and physical properties with respect to that of a region deep in the bulk. Second, even though the surface energy is minimized following reconstruction, it is typically not lowered to its bulk value. Thirdly, and most obviously, atoms at the surface have nothing attached to them in at least one direction. This means that external species (atoms, molecules, etc.) can approach the atoms at the surface and interact with them. This is usually unlikely for atoms deep within the bulk of a material. Overall, the upshot is that surfaces are typically significantly more reactive, both chemically and physically, than the interior bulk of a material.

Going back to our example of a macroscopic cube, the number of atoms at the surface is only a small fraction of the total number of atoms within the object as a whole. As a result, the physical properties within the bulk of the material have a strong influence on the properties of the cube as a whole. However, as we reduce the size of the cube, its surface-area-to-volume (SAV) ratio begins to increase. Reducing the length of each side of the cube by a factor of 10 *increases* its SAV ratio by a factor of 10. Therefore, the SAV ratio of a cube with sides 1 nm will be 10,000,000 higher than a cube with sides 1 cm (Fig. 72.3). In fact, it could be argued that for some nanomaterials, such as single-walled carbon nanotubes or the 2D materials that consist of a single

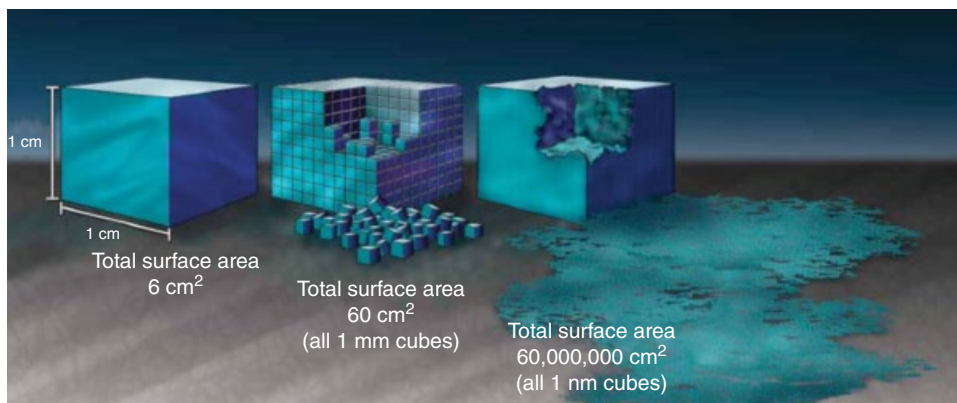


FIGURE 72.3 Illustration of the enormous increase in surface-area-to-volume (SAV) ratio exhibited by nanomaterials compared with macroscale objects. Source: <http://www.nano.gov/nanotech-101/special>. CC-BY-SA-3.0.

layer of atoms, the SAV ratio is infinite since they have no volume in the traditional sense of the word and are essentially all surface.

This enormous increase in SAV ratio is one of the root causes of the strikingly different properties of nanomaterials. Processes and interactions that are highly surface dependent will have a much stronger influence on behavior. Van der Waals forces, surface energy-driven phenomena such as capillarity [21], or surface-dependent interactions including catalysis [22], absorptive capacity, and chemical selectivity [23], can hence be intensified by many orders of magnitude at the nanoscale.

72.3.2 Electrostatic and Van der Waals Forces

The strength of electrostatic and van der Waals forces at the nanoscale means that they are used routinely for moving and placing nanostructures. It was these forces that allowed researchers in 1989 to carefully position 35 xenon atoms to spell out the IBM logo in one of the most iconic early demonstrations of nanotechnology [24]. More recently, in 2013, researchers used a similar approach to create the world's smallest stop-motion film "A Boy and His Atom" (<http://www.research.ibm.com/articles/madewithatoms.shtml>). Of course these accomplishments are entertaining (and good for advertising!), but they also illustrate just how precisely nanoscale objects with large SAV ratios can now be manipulated.

Another scientific breakthrough resulting from the electrostatic manipulation of nanomaterials was the discovery in 2004 of graphene. Graphene consists of a single sheet of carbon, one atom thick, with an sp^2 (honeycomb) bonding structure. Researchers used the van der Waals forces between scotch tape and pieces of graphite to peel off thin flakes of carbon. The flakes were then attached to silicon

wafers, again using van der Waals attraction. Some of these flakes were found to be only one atom thick, and this was the first experimental demonstration of graphene. Graphene has since been shown to possess many remarkable properties from exceptional strength to very high electrical conductivity. Graphene also has a unique electronic structure that arises from the fact that it is only one atom thick. We will discuss the specific properties that result from this unusual electronic structure in Section 72.4.1.1.

72.3.3 Color

The increase in SAV ratio at the nanoscale can have surprising effects on certain fundamental physical properties. For example, we know gold as a shiny yellow metal, but even in medieval times, gold nanoparticles were being used unknowingly to create red and purple stained glass windows. This change in color arises from the way that light interacts with arrays of gold atoms. An oscillation is set up in free electrons at the metal surface, and this oscillation has a resonance called a surface plasmon. Compared with bulk gold, it is much easier for incoming light to excite this plasmon mode at the surface of a gold nanoparticle due to its much higher SAV ratio. The plasmon interaction results in strong absorption of photons with wavelengths close to 450 nm, which gives gold nanoparticles their characteristic red color [25].

72.3.4 Melting Point

The temperatures at which bulk materials melt are well known, typically with high precision. However, as materials are shrunk down to the nanoscale, researchers have shown that their melting points can change. For example, the usual melting points of the elements indium and gold are 157 and 1064°C, respectively. However, as these metals are reduced in size until they are clusters of atoms a few nanometers across, their melting points fall dramatically (Fig. 72.4) [26–28].

The reason for this change in such a fundamental physical property as melting point is once again the increase in SAV ratio at the nanoscale. During the melting of any material, the energy provided by heating enables the bonds between atoms in the solid to be broken to form a liquid. Atoms at the surface of a solid are less stable than atoms deep within the bulk since they are not bound to anything above. Less energy is therefore needed to pull an atom off the surface (i.e. to break the bonds with atoms below and beside it) than to remove an atom from within the bulk, where there are additional bonds above it to break. In a nanocluster, the high SAV ratio means that a far larger proportion of the total atoms are at the surface. As a result, less heat energy is needed to break all the bonds in the nanocluster and so the melting point is lower.

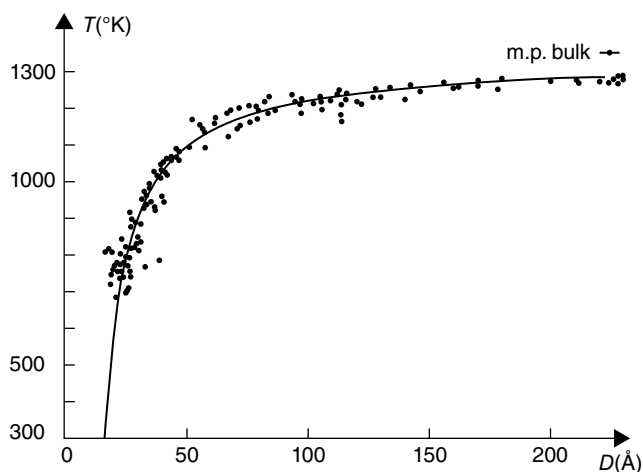


FIGURE 72.4 The change in the melting point of gold nanoclusters as a function of their diameter D . The unit of distance on the x -axis is the Angstrom, \AA , ($1\text{\AA} = 0.1\text{ nm}$). As the size of the gold nanoclusters increases, the melting temperature, T rises toward the bulk melting point of gold of 1337 K (1064°C). Source: Buffat and Borel [26]. Reproduced with permission of the American Physical Society.

72.3.5 Magnetism

One of the most surprising effects of shrinking materials to the nanoscale is perhaps that of turning certain nonmagnetic metals into magnets. Unlike the elements iron, cobalt, and nickel, noble metals such as gold, platinum, and palladium display no magnetic properties in the bulk. However, when nanoclusters are created from these metals, they can suddenly become magnetic [29, 30]. This effect is rather sensitive and appears to be strongly linked to nanocluster geometry [31] and the specific number of atoms involved (i.e., nanocluster radius) [30]. The origin of this effect seems to be highly surface dependent and hence another consequence of the large SAV ratio of nanomaterials [32]. Atoms at the surface of the nanocluster have bonding electrons with nothing outside of the nanoparticle to bond to. These unpaired surface electrons (which don't exist in the bulk metal), combined with the complex electronic structure of the nanoclusters, induce an overall magnetism in the nanoparticle [33]. As the size of nanoclusters is reduced, the SAV ratio rises, amplifying the effect of unpaired surface electrons, thereby increasing the magnetism (Fig. 72.5).

72.3.6 Hydrophobicity and Surface Energetics

High SAV ratios give researchers an opportunity to control a material's surface energy by creating nanostructured coatings. Controlling the surface energy influences the way that these coatings interact with external chemical species (e.g. water). Hydrophobic coatings have very low surface energies and so repel water via capillary

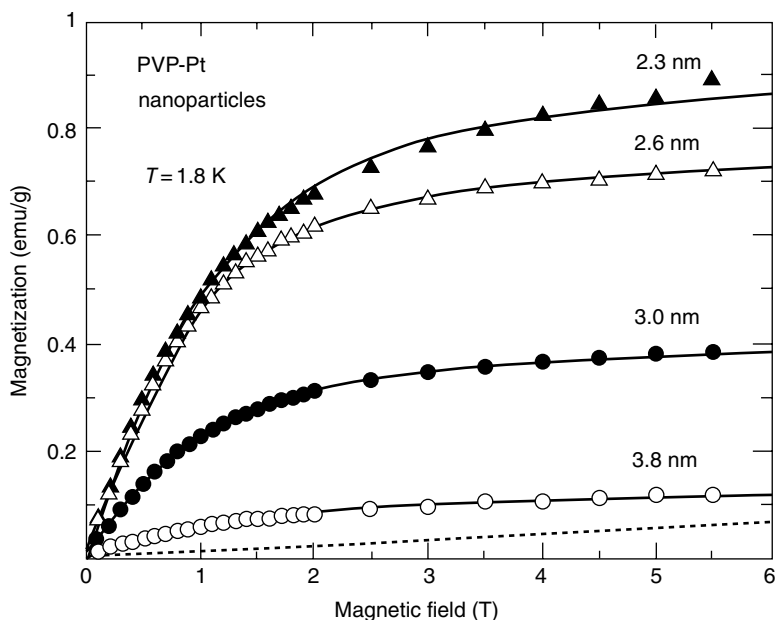


FIGURE 72.5 Dependence of the magnetism of platinum nanoclusters on their radius. As the radius of the nanoclusters is decreased from 3.8 to 2.3 nm, their magnetization increases. The dashed line shows bulk platinum metal and its relative lack of magnetic behavior. Source: Yamamoto *et al.* [30]. Reproduced with permission of Elsevier.

action to prevent it from wetting the surface. Hydrophobic surfaces abound in nature. A common example is the lotus plant (Fig. 72.1), the leaves of which, at the nanoscale, are covered in waxy bumps. Rather than wetting the surface, the waxy nanostructures cause water to ball up into droplets and roll off the leaves. The leaves are hence self-cleaning: the water droplets carry away any particles of dirt that could hinder photosynthesis.

The extent of hydrophobicity is quantified by measurement of the contact angle, θ_c , between a droplet of water and the surface (Fig. 72.6). Hydrophobic surfaces have θ_c less than 90° and the water is said to “wet” the surface; hydrophilic surfaces are characterized by θ_c greater than 90° . Superhydrophobic surfaces have values of $\theta_c > 150^\circ$ and approach the limit of 180° (i.e., perfectly spherical water droplets) where the water is completely repelled by the surface.

Many applications for superhydrophobic nanomaterial coatings exist, from preventing icing of control surfaces on aircraft, to self-cleaning windows. The key to superhydrophobic coatings with low surface energies is that they consist of nanostructures with very sharp protrusions. By using chemical treatments to functionalize the sharp nanostructured surfaces, the water droplets are forced toward the tips of the points, while a layer of air trapped between the points actually prevents the water from reaching the surface (Fig. 72.7a). Incidentally, it is possible to reverse their function by

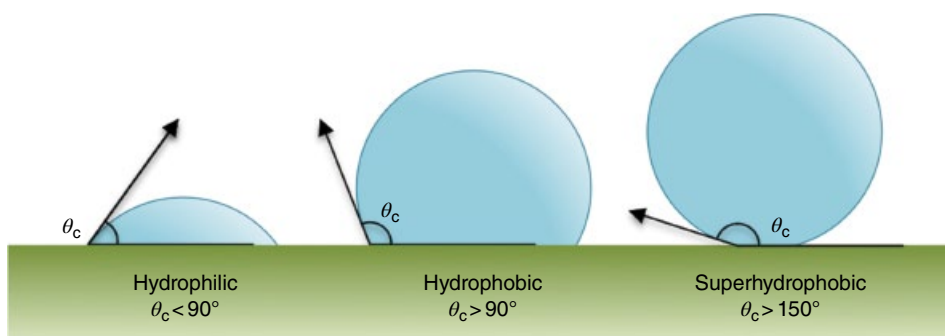


FIGURE 72.6 The degree of hydrophobicity is characterized by the contact angle, θ_c , formed between a water droplet and the surface on which it sits.

instead coating the spikes with a different chemistry that makes the surface superhydrophilic. When water hits a superhydrophilic surface, it forms an almost perfectly uniform, flat layer (see Section 72.3.8).

The nanoscale protrusions can be engineered using numerous approaches, including sharpening the tips of a bundle of silica optical fibers into nanocones (Fig. 72.7b) [35], selectively etching certain parts of a glass matrix (Fig. 72.7c) [36], or even using diatomaceous earth, which contains the skeletal remains of creatures that possess naturally occurring nanoscaled features (Fig. 72.7d) [34].

72.3.7 Nanofluidics

Another area dependent on the unusual capillary properties of nanomaterials is the field of nanofluidics. Nanofluidics is the study of how fluid flow changes at the nanoscale as a result of the increased SAV ratio of nanostructures and nanochannels. Capillary action is a manifestation of the interaction between two materials (often a liquid and a solid) at an interface. At the nanoscale, the number of molecules in the liquid becomes very low, while the effect of the solid surfaces on flow is greatly magnified due to the high SAV ratio. As a result, nanofluidic systems exhibit some unique properties [37–40].

The high SAV ratio in nanofluidic channels means that flow is governed predominantly by surface charge, and this can be used to distinguish between or separate different ions [39]. As a fluid moves across a surface, a double layer of charged particles builds up at the interface. The first layer consists of charged particles that are chemically attached to the surface in some way, and these may be positive or negative, depending on the situation. The second layer is made up of particles with the opposite charge that are coulombically attracted to (and electrically screen) the first layer. When the thickness of this double layer becomes comparable to the width of the nanochannel, only ions of one polarity (i.e., positive or negative) will be able to pass through [39].

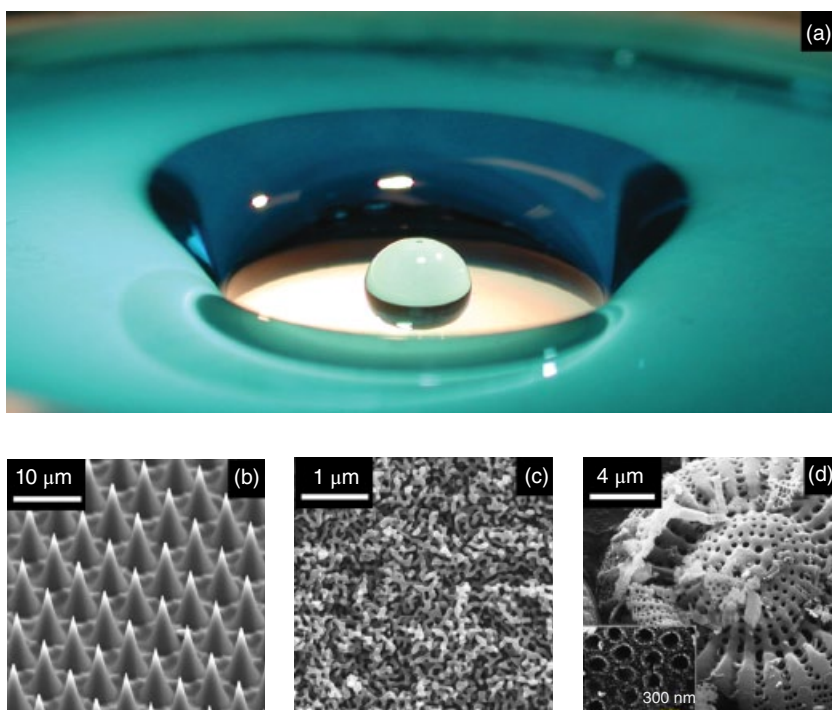


FIGURE 72.7 (a) The “Moses effect” where water is strongly repelled from a nanostructured superhydrophobic surface. A thin layer of air prevents the water from ever touching the structured surface. Source: Reproduced with permission of John T. Simpson. (b) An array of glass nanocones created by etching the ends of bundled and fused optical fibers. Source: Reproduced with permission of John T. Simpson. (c) A borosilicate glass film after spinodal decomposition and selective etching of sodium borate-rich regions, showing the porous nanoscale branched network of the silica-rich film. Source: Reproduced with permission of John T. Simpson. (d) Scanning electron microscopy image of the skeletal remains of a circular diatom (i.e., diatomaceous earth) characterized by micro- and nanoscale surface features. Source: Polizos *et al.* [34]. Reproduced with permission of Elsevier.

It is also the case that nanochannels may be comparable in size to molecules dissolved in the fluid. Larger molecules are thus physically blocked from passing through, while smaller molecules are transported freely [38]. In this way nanofluidic devices are able to act as electrostatic or physical filters, either separating different ions in an electrolytic fluid on the basis of their charge or distinguishing between molecules based on their size. This ability to filter nanoparticles and molecules from fluids is extremely appealing for biotechnology applications, particularly in the areas of DNA and protein analysis [40–42]. Nanofluidic channels can be made in a variety of ways, for example, using electron beam lithography or selective etching to make nanotubes [19, 41–43].

72.3.8 Nanoporosity

Closely aligned with both the nanostructured superhydrophobic surfaces and nanofluidic systems are nanoporous materials. Nanoscaled holes or pores can be used as molecular filters, antifogging coatings, desiccants, ion exchangers, and catalysts [44–47]. Let us take the example of antifogging coating as an example. Nanoporosity is used to engineer surface energies in a similar way to the superhydrophobic materials discussed in Section 72.3.6. The difference is that in this case the nanoporous material has a very high surface energy (either naturally or after chemical treatment) such that it is superhydrophilic (i.e., θ_c approaches 0°). Water is attracted to its surface, creating a continuous sheet of liquid rather than forming into discrete droplets that scatter light [46]. When applied to glass, these superhydrophilic nanoporous coatings hence prevent fogging and keep the glass clear (Fig. 72.8).

A very different application for nanoporous materials is in the technologically important area of terahertz (THz) optoelectronics. The efficient emission and detection of light at THz frequencies is critical to a wide range of fields from defense to biomedicine. Certain bulk semiconductor crystals emit THz radiation when excited with an ultrafast laser, but the strength of the THz signal is usually low [48]. Researchers have shown that increasing the surface area of the semiconductor crystal significantly enhances the THz emission intensity. Of course, with their high SAV ratio, nanostructured surfaces are ideally suited to this purpose. Groups began by demonstrating that in semiconductor nanowires, an enhancement of THz emission of up to 40 times could be achieved [49, 50]. They showed that the nanowires' high aspect ratio (length/diameter) geometry was critical for achieving the desired effect [49, 50]. Researchers then explored the use of nanoporous surfaces (Fig. 72.9a) [52, 53]. Compared with the bulk semiconductor crystal, THz emission enhancements of 100–1000 times have been reported for nanoporous GaP surfaces (Fig. 72.9b) [51].

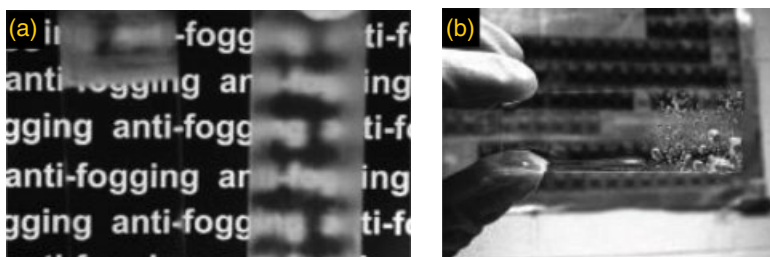


FIGURE 72.8 (a) Comparison of the fogging behavior of a bare glass slide (right-hand slide) and a slide with a superhydrophilic nanoporous coating (left-hand slide). For comparison, the fogged region at the top of the left-hand slide was not coated. (b) Glass slide illustrating the nonuniform water dewetting behavior of normal glass (right half) compared to the uniform wetting behavior of the surface with the superhydrophilic nanoporous coating (left half). Source: Cebeci *et al.* [46]. Reproduced with permission of the American Chemical Society.

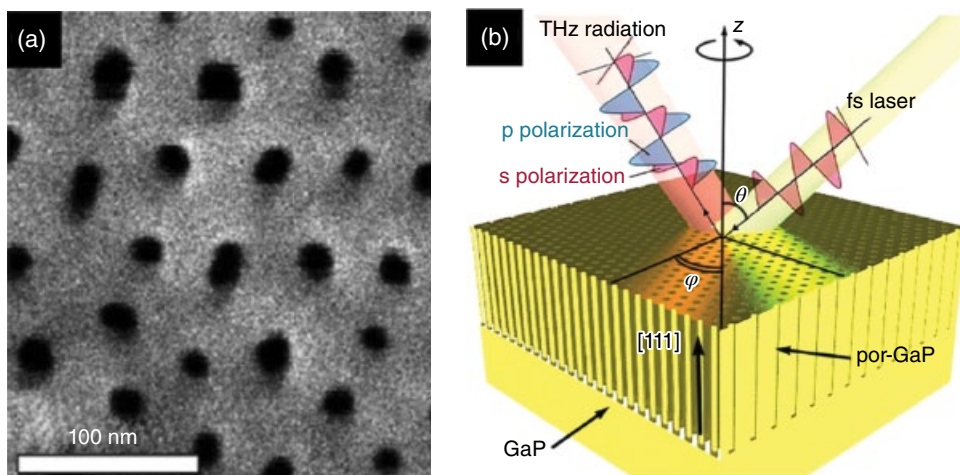


FIGURE 72.9 (a) The nanoporous GaP surface created by ion etching, where the nanopores have an aspect ratio of approximately 1500 [51]. (b) Schematic diagram of the experimental geometry for THz emission from a nanoporous material. The femtosecond (fs) laser coming in from the right excites the nanopatterned porous GaP (por-GaP) surface so that a THz beam is emitted on the left. Source: Atrashchenko *et al.* [51]. Reproduced with permission of AIP Publishing LLC.

A wide range of techniques have been adopted to create nanoporous materials, including the use of large molecules [44, 45], structured arrays [47], nanocasting [54], and dealloying and selective etching [55].

72.3.9 Nanomembranes

Researchers are able to create nanostructured membranes out of various electrically conductive materials. The conductivity of these nanomembranes change when different chemical species interact with their surface. The high SAV ratio of a nanomembrane provides a huge amount of surface area for these species to interact with. As a result, nanomembranes can be used as extremely sensitive chemical sensors [56].

Nanomembranes with thicknesses on the nanoscale can also be designed as micro-mechanical sensors [57]. A 25–70 nm thick polymer/gold composite nanomembrane is freely suspended above a comparatively large opening several hundred microns in diameter (Fig. 72.10) [58]. In response to stress, either from the tip of an atomic force microscope or pressurized gas, the nanomembrane bulges out and its total deflection is measured under different conditions. The nanomembranes have a high elastic modulus, such that deflections up to 40 μm can be readily achieved without plastic deformation or rupture of the membrane [58]. After the load is removed, the nanomembranes return to their natural unstressed shape within a few seconds.

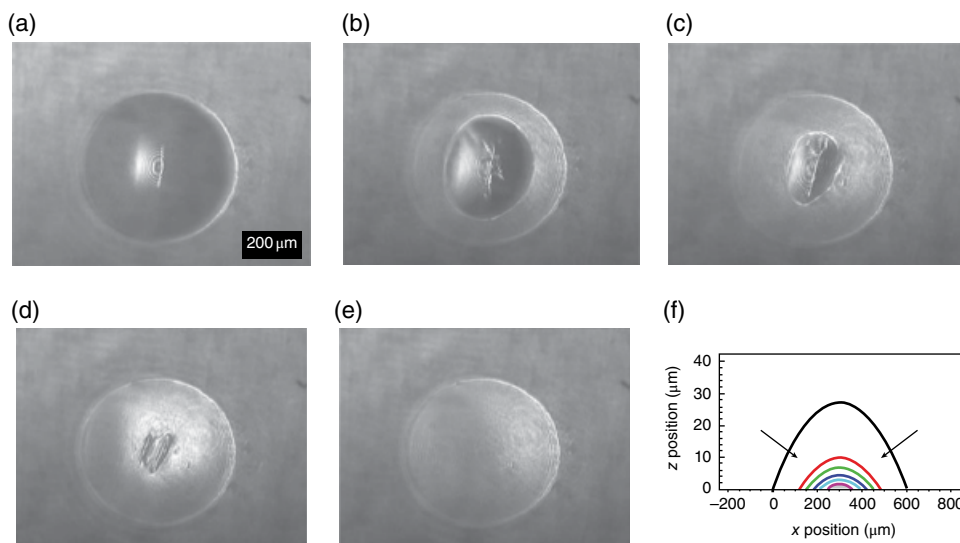


FIGURE 72.10 Auto recovery of a freely suspended nanomembrane subjected to high pressure and long loading time. (a) Fully distorted nanomembrane under high pressure (4 kPa). (b)–(e) Evolution of nanomembrane shape as a function of time after sudden pressure release—images taken at 2 s intervals. The overstretched central portion disappears and the nanomembrane becomes flat. (f) Cross sections through the nanomembrane during recovery. Source: Jiang *et al.* [58]. Reproduced with permission of Nature Publishing Group.

These nanomembranes have an enormous dynamic range: the highest pressure they can detect is 10^8 larger than the smallest measurable pressure [58]. As a result, nanomembranes like this are significantly more sensitive than comparable silicon membranes of the same diameter. Resonant frequencies of around 100 kHz causing vibrational amplitudes of 25 nm are measured for the nanomembranes. These nanomembranes could therefore be used as acoustic or even thermal sensors.

72.3.10 Nanocatalysis

The use of catalysts to start or speed up chemical reactions and to increase their efficiency or yield is ubiquitous. The chemical and oil industries are completely reliant on them, while consumers use them daily, for example, to reduce emissions from their vehicles. Many catalysts work by accelerating the reaction rate at the interface between a solid surface and a fluid. Reactant species adsorb to the surface of the catalyst, a process that lowers the energy barrier for the reaction to occur. The catalyst, often consisting of a transition metal element, does not take part in the reaction itself. The catalyst is therefore not consumed and so only a small amount is typically needed. Since adsorption of reactant species is entirely reliant on the availability of active sites on the surface of the catalyst, increasing a catalyst's surface

area greatly increases its efficacy. Due to their high SAV ratios, catalytic nanomaterials can be extremely effective [22, 59, 60]. What is more, the activity of nanocatalysts can be further enhanced by careful engineering of their structural properties, for example, their shape, alloy composition, or nanostructure [22, 59, 60].

Increasingly, nanomaterials with an electrocatalytic behavior are sought for their ability to increase the reaction rate between hydrogen and oxygen, with the aim being to create lightweight but highly efficient fuel cells for use in vehicles [61–63]. Nanocatalysts are therefore at the forefront of the green energy revolution and an area of intensive research.

72.3.11 Further Increasing the SAV Ratio

Despite the fact that nanomaterials inherently possess very high SAV ratios, for some applications it is desirable to increase the SAV ratio even further. As an example, researchers wished to increase the surface area of a silicon nanowire (Fig. 72.11) so that it could be covered with a greater number of light-sensitive molecules for a detector. Rather than use complex photolithographic approaches to reduce feature size, they exposed the silicon nanowire to a metal-induced chemical etch to create silicon “nanoforests” (Fig. 72.11) [64]. Taking into account each silicon “tree,” it is clear that compared with the original planar nanowire, these nanoforests have an enormous surface area and hence a greatly increased SAV ratio. Similar approaches that try to increase SAV ratio in order to magnify various nanomaterial properties are widely used.

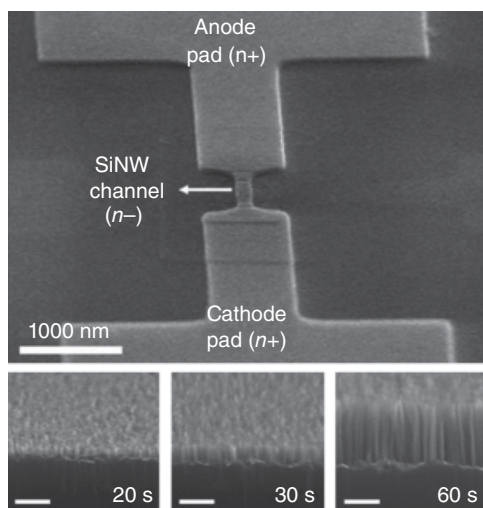


FIGURE 72.11 Top image shows a silicon nanowire (SiNW) connecting a cathode to an anode. The bottom three images show how the surface of the SiNW evolves into a Si nanoforest as the etch time is increased (scale bars=200 nm). Source: Seol *et al.* [64]. Reproduced with permission of the American Chemical Society.

72.3.12 Nanopillars

The high aspect ratio nanoforests in Figure 72.11 are created using a top-down approach. Similar structures, known interchangeably as nanopillars or nanowires, can also be created using a bottom-up approach. Nanopillars can be made to grow vertically out of a flat substrate surface using various epitaxial (i.e., crystal growth) techniques. Depending on the desired application and the growth method used, these nanopillars are either randomly located across the substrate surface [65] or arranged in ordered arrays (Fig. 72.12) [66].

The versatility of the bottom-up approach means that the nanopillars can be created in arrays with predefined geometries [66, 67], even on traditionally incompatible substrate materials [65, 68], or combining multiple materials in the same pillar in so-called core-shell or axial geometries to build electronic or photonic device structures [69–71]. Randomly arranged pillars are typically created using metallic nanodroplets to catalyze the vertical growth [72, 73]. In contrast, nanopillar arrays are formed by first depositing a mask on the substrate surface. E-beam or nanoimprint lithography is used to open nanoholes in the mask to expose the substrate only in certain areas. During growth, the nanopillars form only in the holes and so by designing the arrangement of the nanoholes, nanopillar arrays with almost any geometry can be readily engineered [67, 74]. This can be put to use to create photonic crystals for light trapping or to act as a laser cavity [69, 75, 76].

Nanopillars typically have radii of 25–100 nm but can be several microns in length. This enormous aspect ratio exaggerates the already large SAV ratios possible in nanostructured materials and can be exploited in a wide range of applications including light-emitting diodes [71, 74], lasers [76, 77], photodetectors [78], solar cells [69, 79], transistors [80–82], and electrically gated modulator waveguides [83]. Taking solar cells as an example, the large SAV ratio of the pillars is particularly attractive for several reasons: the large surface area increases photon absorption probability; the light-trapping properties of the nanopillar array further improves photon capture; and

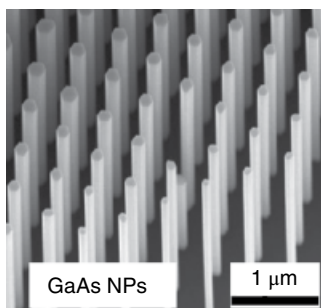


FIGURE 72.12 An ordered array of GaAs nanopillars grown via metal organic chemical vapor deposition on a nanopatterned GaAs substrate. Source: Lin *et al.* [66]. Reproduced with permission of IOP Publishing.

the ease of radial charge collection due to the high aspect ratio of the nanopillars raises the efficiency of the solar cell [69].

72.3.13 Nanomaterial Functionalization

Preceding sections have touched on the fact that, the large SAV ratio of nanostructured materials can be exploited through “functionalization.” Chemical or biological species that serve some active, useful purpose are attached to the surface of nanomaterials. Because of the large surface areas, enormous numbers of these functional molecules can be attached, making their combined action highly effective. Functionalization hence allows the chemical or physical properties of a nanomaterial to be engineered for specific applications [84].

Functionalization has been demonstrated for various nanomaterial families, including graphene [85–87], carbon nanotubes [88–90], nanowires [91–93], and semiconductor nanoparticles known as colloidal quantum dots [94–96]. Applications for functionalized nanomaterials are enormously varied but include electrical devices [85], improving the structural properties of composite materials [86], and chemical sensing [91, 92]. However, it is the field of biotechnology that is finding some of the most exciting uses for functionalized nanomaterials. Biofunctionalized nanomaterials are used as light-emitting fluorophores for *in vivo* imaging, labeling, diagnostics, and sensing [94–96]; measuring DNA hybridization in real time [93]; and the delivery of drugs to specific areas of an organism for targeted disease treatment [87, 95].

Taking colloidal quantum dots as an example, these semiconductor nanoparticles are extremely promising candidates for cancer treatment. They exist in solution and are small enough that they can pass into an organism and interact with it at the cellular level. Quantum dots are first coated to prevent them from degrading once they enter the organism [95]. Molecular species that target a specific kind of cancer are then attached to their exterior surfaces. Finally, drugs that attack that cancer type are added. The size and selectivity of these functionalized quantum dots means that they are able to deliver the drugs directly to where they are needed. Importantly, the light-emitting properties of the quantum dots of their movements can be simultaneously tracked throughout the organism to verify that they end up in the correct location. In this way the quantum dots are also monitored to see whether they are eventually excreted by the organism or ultimately build up in certain locations (Fig. 72.13) [95].

This approach is extremely versatile. Various quantum dots or nanoparticles (different sizes, light emission wavelengths, etc.) can be combined with a wide range of surface functionalizations (different drugs, target organs, etc.). The ability to control multiple behaviors in a single dose of nanoparticles enables multiple treatments or diagnostic experiments to be run simultaneously [95, 97–100].

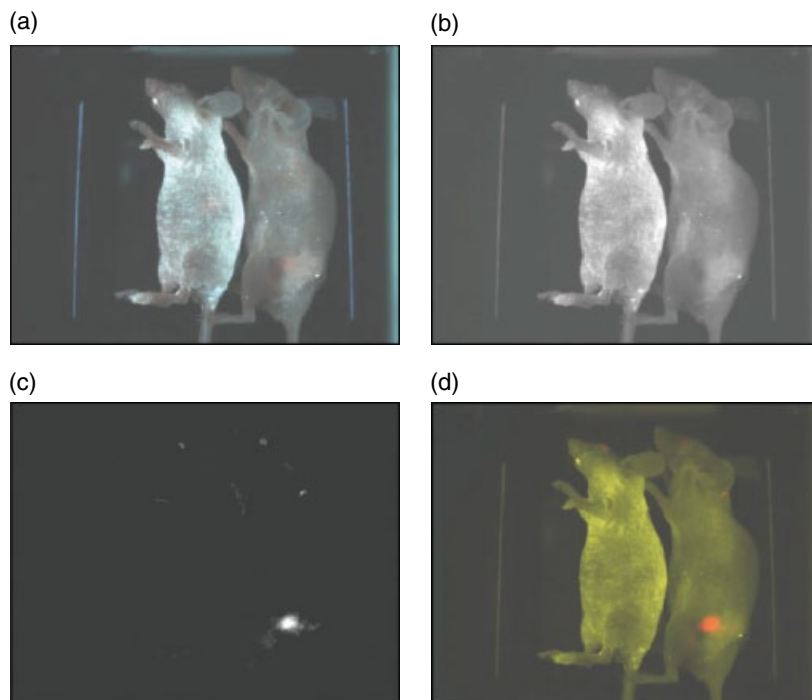


FIGURE 72.13 Spectral imaging of live animals injected with quantum dots that were functionalized to specifically target prostate cancer cells. Each image shows a healthy mouse (left) and a mouse with a prostate tumor (right). Fluorescence signals from the quantum dots, visible as brighter areas in the images of the sick mouse, indicate that they localize in the prostate tumor. The same dose of quantum dots injected into the healthy mouse showed no localized fluorescence signals. (a) Original image, (b) unmixed autofluorescence image, (c) unmixed quantum dot image, and (d) superimposed image. After *in vivo* imaging, the sick mouse was dissected to confirm that the quantum dot signals indeed came from an underlying tumor. Source: Gao *et al.* [95]. Reproduced with permission of Nature Publishing Group.

72.3.14 Other Applications for High SAV Ratio Nanomaterials

In addition to the applications already discussed, the high SAV ratio of nanomaterials (and the properties that result from this) makes them ideally suited for use in a wide range of technologically important areas. These applications include nanoengineered batteries [101–106], solar cells [107, 108], fuel cells [60–63], next-generation 3D computer chip architectures [109, 110], graphene membranes for electron microscopy [111, 112], and interaction with biological systems [113].

72.4 NANOMATERIAL PROPERTIES RESULTING FROM QUANTUM CONFINEMENT

So, an inherently large SAV ratio gives rise to nanomaterials with a dizzying array of properties. However, the second family of nanomaterials that we will look at acquires

their unique characteristics simply because of their size. These nanomaterials are designed so that their interactions with electrons are fundamentally different than at the macroscale, and this gives them their extraordinary properties. Within semiconductor materials, the de Broglie wavelength of an electron (which is inversely proportional to the electron's momentum) is typically on the order of a few tens of nanometers. For example, in GaAs, the de Broglie electron wavelength at 300 K is 24 nm [114]. When nanomaterials are created with feature sizes smaller than the de Broglie wavelength, electrons within the material suddenly begin to behave very differently.

A different physical model is hence needed to describe and predict the electrons' strange behavior: quantum theory. Compared with classical mechanics with which we are all well accustomed at the macroscale, quantum physics famously makes some bizarre predictions about the way tiny objects behave. It is quantum physics that is responsible for giving this second family of nanomaterials their dramatically different properties. This quantum framework allows us to understand what is happening at the nanoscale and permits us to design and create semiconductor nanomaterials with novel properties that cannot be obtained by other means.

Probably the first realization of a structure specifically designed and engineered to have dimensionality on the nanoscale was the development of the quantum well in the 1970s by Bell Labs and other groups [115, 116]. A quantum well consists of a thin slice of semiconductor material with a low energy bandgap, sandwiched between two barriers made of another semiconductor with a larger bandgap (Fig. 72.14). So long as the thickness of the low bandgap material is on the order of the de Broglie electron wavelength, the energy of an electron confined to this region will become quantized. It should be noted that quantum mechanics is at play at all length scales; however, it is only below the de Broglie wavelength that the splitting between the quantized states becomes large enough that the energy levels no longer appear to be continuous.

What this means is that although the electron can have any arbitrary energy in the plane of the quantum well, in the nanoscaled out-of-plane direction the electron must occupy one of several discrete, fixed energy levels. The position of the energy levels is specific to a given quantum well nanostructure and is affected by the nanoscale width of the well and the bandgaps of the constituent semiconductor materials. The electron can move between these levels by absorbing or emitting packets (known as quanta) of energy. These quanta of energy are often absorbed or emitted in the form of a photon, that is, a particle of light with exactly the same energy. Since a photon's color is a function of its energy, this means that by simply adjusting the width of the quantum well nanostructure, we can control the color of light that it either absorbs or emits. This property of these nanomaterials led to one of the first applications of the quantum well: wavelength tunable semiconductor lasers [117].

A quantum well has one nanoscaled dimension, that is, its width. Following the development of quantum wells, semiconductor nanostructures that offer quantum confinement in two, and three dimensions have also since been demonstrated. These are, respectively, the quantum wire, which has two nanoscale dimensions (i.e., width and

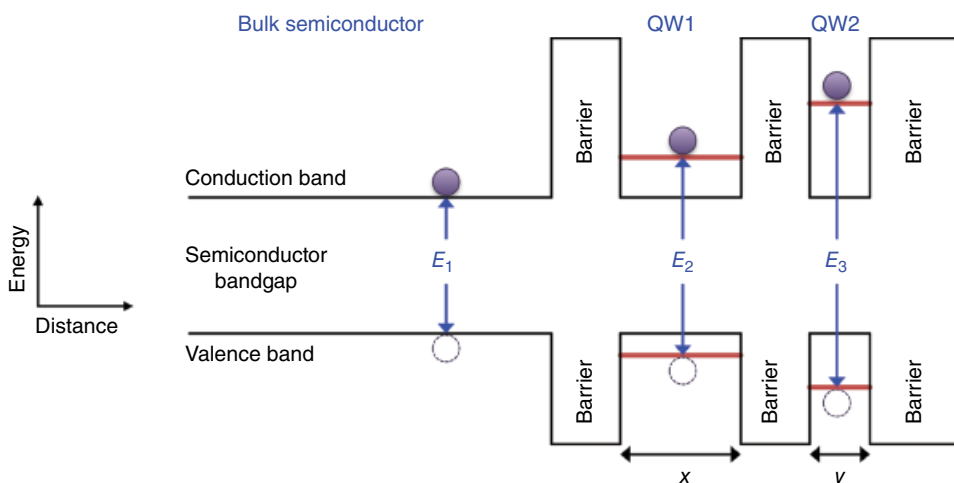


FIGURE 72.14 Schematic diagram showing the effect of quantization on electron energy levels. In a bulk semiconductor the electron transition is between the top of the valence band and the bottom of the conduction band. A photon emitted as a result of this transition will have energy E_1 . In quantum well, QW1, of width x (where x is on the order of the de Broglie electron wavelength), the electron energy is now quantized perpendicular to the plane of the well. The quantized energy levels exist below (above) the valence (conduction) band edges. A photon emitted due to an electron transition in QW1 will thus have energy E_2 , where $E_2 > E_1$. In quantum well, QW2, of width y (where $y < x$), the quantized electron energy levels are “squeezed” further below (above) the valence (conduction) band edges. A photon emitted due to an electron transition in QW2 will have energy E_3 , where $E_3 > E_2 > E_1$.

height), and the quantum dot, which is nanoscaled in all three dimensions (i.e., width, height, and length) (Fig. 72.15).

In a quantum wire, the electron energy is continuous along the length of the wire and quantized in the two nanoscaled directions. Quantum dots are nanoscaled in all three spatial dimensions so that the electron’s energy becomes fully quantized. In this respect, quantum dots are sometimes referred to as artificial atoms. Just as in an atom, the electrons in a quantum dot exist in well-defined energy levels, with rules about how multiple electrons must arrange themselves. The key difference between an atom and a quantum dot is that nature has fixed the atomic electron orbital energies, whereas we can readily tune the electron energy levels of a quantum dot simply by adjusting its size (Fig. 72.14).

As the number of nanoscaled dimensions increases from zero (bulk material), to one (quantum well), to two (quantum wire), to three (quantum dot), there is a big change in the density of states (DoS) (Fig. 72.16). The DoS is defined as the number of available “locations” that can be occupied by electrons at a given energy.

Together with the tunability of the electron energy levels, the capacity to change the DoS simply by adjusting the number of nanoscaled dimensions is perhaps the most

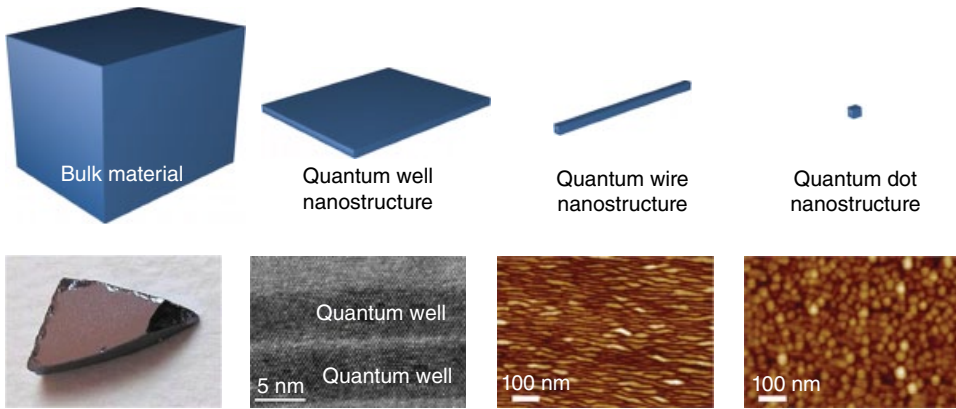


FIGURE 72.15 As we sequentially reduce each of the sides of a macroscopic bulk semiconductor crystal to the nanoscale, we create a specific family of quantum nanomaterials. (Top row): A quantum well offers quantum confinement in one dimension, a quantum wire in two, and a quantum dot has confinement in all three spatial dimensions. (Bottom row): Images of each of these quantum systems realized experimentally. Source: (bottom left) https://upload.wikimedia.org/wikipedia/commons/6/6f/Gallium_arsenide_crystal.jpg. CC-BY-SA 3.0; (bottom middle left) Paul [118]. Reproduced with permission of John Wiley & Sons, Inc. and (bottom middle right and bottom right) Li *et al.* [119]. Reproduced with permission of AIP Publishing LLC.

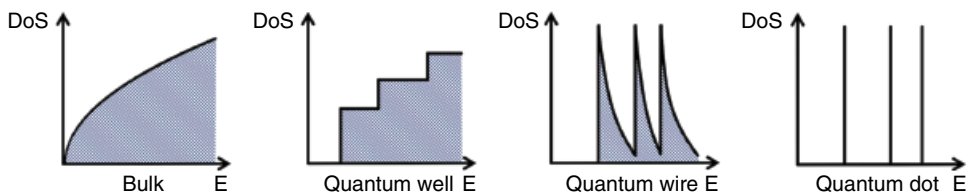


FIGURE 72.16 The change in density of states (DoS) in semiconductor nanomaterials as a function of the number of quantized dimensions. After [120].

attractive feature of semiconductor quantum nanostructures. This exquisite control over both the electron energy and DoS opens up new opportunities for designing materials for specific applications and for the discovery of new physics.

72.4.1 Quantum Well Nanostructures

Take the example of the semiconductor laser mentioned previously. Compared to a laser built from bulk semiconductor material, quantum well lasers have many superior characteristics [121, 122]. In a bulk semiconductor laser, the material chosen for the gain region dictates the emission wavelength. However, in a quantum well, quantization pushes the electron energy levels higher than the bulk semiconductor bandgap (Fig. 72.14).

This means that a photon emitted from a quantum well will have higher energy than a photon emitted from the same bulk semiconductor (in Fig. 72.14, $E_2 > E_1$). A photon with higher energy has a shorter wavelength. This property means that we can take a bulk semiconductor that emits light at a long wavelength and, by creating a quantum well out of it, make it emit light at a shorter, potentially more useful wavelength. For example, a bulk semiconductor that emits in the infrared region of the electromagnetic spectrum can be made into a quantum well laser that operates at visible wavelengths.

The wavelength of a bulk laser is fixed by the choice of semiconductor and can only be altered by changing the composition of the material. In contrast, in a quantum well laser, the emission wavelength is also a function of the width of the quantum well (Fig. 72.14). Higher energy (shorter wavelength) photons are emitted from thinner quantum wells, while wider quantum wells emit lower energy (longer wavelength) photons. Therefore simply by controlling the quantum well width, the lasing wavelength can be readily tuned ($E_3 > E_2$ in Fig. 72.14).

The ability to modify the electron DoS is the key to the extremely low threshold currents possible in quantum well lasers [123, 124]. The threshold current is the current above which stimulated emission exceeds spontaneous emission (i.e., the device begins to behave as a laser). Reducing the threshold current is highly desirable as it means that the laser uses much less power during its operation. In bulk material, the electron DoS increases with the square root of energy. In contrast, for a given quantized energy level in a quantum well, the electron DoS is independent of energy (i.e., constant). Compared with bulk material, a quantum well can therefore contain significantly more electrons with the same energy, which reduces the threshold current required to turn on a quantum well laser [125].

In addition, the threshold current in quantum well lasers is relatively insensitive to changes in temperature because of their step-like DoS (see Fig. 72.16) [122, 126]. Despite various cooling techniques, the temperature in a typical electronics rack can quickly rise, even under normal operating conditions. It is important therefore that a laser can cope with increasing temperatures and provide stable performance. The threshold current density, J_{th} , is related to the temperature of the laser, T , by the expression

$$J_{th}(T) = J_{th}(0) \exp(T / T_0)$$

The larger the value of the constant T_0 , the less J_{th} varies as T is raised or lowered. Even early in their development, $T_0 = 437^\circ\text{C}$ was reported for quantum well lasers [126], which compares extremely favorably with $T_0 = 185^\circ\text{C}$ for bulk lasers developed around the same time [127].

A more recent development in lasers based on quantum well nanostructures is the quantum cascade laser [128–131]. In a single quantum cascade laser, hundreds of quantum wells are placed side by side. The quantum wells are carefully designed so that their quantized energy levels create a “staircase” for electrons to flow down (Fig. 72.17) and, in doing so, emit very long wavelength photons.

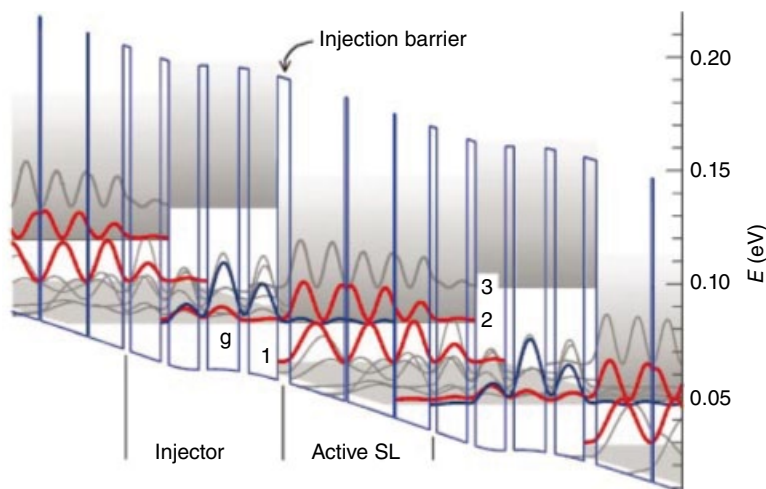


FIGURE 72.17 Calculated conduction band structure of a small part of a quantum cascade laser showing the multiple quantum wells with different widths. To achieve large gains, quantum cascade lasers will typically contain hundreds of repeats of these injector/active region sections. In this device, photons with energy 18 meV (4.35 THz) are emitted by electron transitions from state 2 to state 1 in the Active SL section. After traversing the injector ground state (g), the electron finds itself in the excited state 2 of the Active SL section and the photon is emitted. The electron is then injected into the next Active SL section and the process is repeated. Source: Köhler *et al.* [130]. Reproduced with permission of Nature Publishing Group.

Clearly, realizing a quantum cascade laser depends on the ability to create quantum wells with precise widths. This requires a growth technique that is stable over the many hours needed to grow these complex structures, most commonly, molecular beam epitaxy. These quantum cascade structures are currently the most effective ways of obtaining THz laser emission, the importance of which we discussed in Section 72.3.8.

In addition to enabling new devices, the ability to create quantum wells from ultrahigh electron mobility materials allowed researchers to directly test certain predictions made by quantum theory and led to the discovery of new physics. Perhaps most notable of these new findings was the discovery in 1980 of the integer quantum Hall effect, for which the Nobel prize in Physics was awarded in 1985 [132]. Discovery of the fractional quantum Hall effect followed in 1982 and led to the 1998 Nobel prize in Physics [133].

When a current is passed through a bulk conductive material in the presence of an orthogonally aligned magnetic field, a voltage is developed perpendicular to both. This is the classical Hall effect. The magnetic field curves the path of the electrons as they travel. As a result, a voltage is created by the accumulation of a net negative charge on one side of the conductor and positive charge on the other. The Hall coefficient, R_H , has an inverse linear dependence on the magnetic field strength.

To observe the quantum Hall effect, a quantum well is created in a high electron mobility semiconductor and cooled down, typically to 1.5 K or below, in the presence

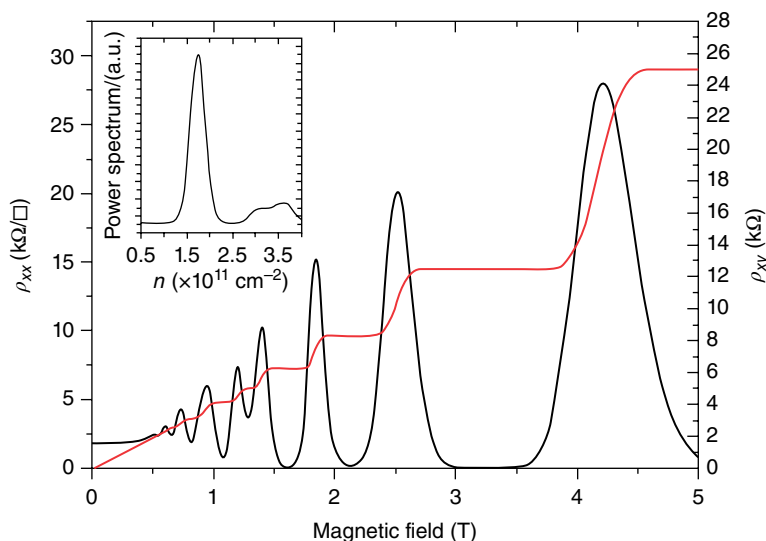


FIGURE 72.18 Measurement of longitudinal (ρ_{xx}) and Hall (ρ_{xy}) resistivities at 1.5 K as a function of magnetic field for an $\text{In}_{0.75}\text{Ga}_{0.25}\text{As}$ quantum well. The inset shows the fast Fourier transform of the longitudinal data, which gives an electron sheet density of $1.6 \times 10^{11} \text{ cm}^{-2}$. The electron mobility in this sample was approximately $200,000 \text{ cm}^2/\text{Vs}$. Source: Simmonds *et al.* [134]. Reproduced with permission of AIP Publishing LLC.

of a suitably large magnetic field. In contrast with the bulk case, when the magnetic field strength is swept, plateaus are observed in the linear Hall resistance (ρ_{xy} curve passing through the origin in Fig. 72.18). This is the signature of the quantum Hall effect [135, 136]. The longitudinal resistance in the direction of the current flow fluctuates between zero and finite values in a phenomenon known as Shubnikov-de Haas oscillations (curve oscillating between zero and non-zero values of ρ_{xx} in Fig. 72.18).

The magnetic field causes electrons in the plane of a quantum well to move in cyclotron orbits due to the Lorentz force. The orbits in which the confined electrons move have quantized radii, each of which is known as a Landau level [137]. As the strength of the magnetic field is increased, the degeneracy of these Landau levels (i.e., the number of electrons they can hold) rises. Electrons in higher Landau levels are therefore able to “fall” into lower Landau levels until the higher level is empty. At this point the Fermi energy (the highest energy level occupied by an electron) “jumps” down to the lower Landau level. This process is repeated as the magnetic field continues to rise, until all the electrons are in the lowest Landau level.

When the Fermi level is within a Landau level, the Hall resistance increases. However, when the Fermi level is between two Landau levels, the Hall resistance remains constant, and we see the characteristic plateaus of the quantum Hall effect (Fig. 72.18). The plateaus take values

$$\rho_{xy} = h / \nu e^2 \quad (\nu = 1, 2, 3, \dots)$$

ν is referred to as the filling factor and is equivalent to the number of filled Landau levels in a spin-split system (where the magnetic field is strong enough to give electrons with opposite spins different energies). Note that at magnetic fields that aren't strong enough to lift the electron spin degeneracy, only the plateaus corresponding to even values of ν are seen. h/e^2 is the resistance quantum. As a result of quantum Hall effect measurements, its value (25.813... k Ω) is known with extreme precision and is now used as the international standard for resistance. The quantum Hall effect is used as a method for measuring the electron density in a quantum well and the electron mobility of the semiconductor material [134].

In extremely high mobility materials at very low temperatures in large magnetic fields, additional quantum Hall plateaus have been found at noninteger values of ν . This is the fractional quantum Hall effect [133]. The origins of the fractional quantum Hall effect are not yet fully understood, but experimental support is rising for a theory that invokes a new state of matter based on fractionally charged quasiparticles [138–141].

72.4.1.1 2D Materials Since the discovery of graphene in 2004 [142], research into atomically thin materials has exploded. Their 2D nature means that these materials are in some ways similar to the quantum wells discussed previously. However, the difference is that it is not simply the electrons that form a 2D sheet due to quantum confinement; here the atoms making up the material itself are arranged in a 2D sheet. This has a huge impact on the behavior of 2D materials, which are often drastically different from samples of the same materials consisting of multiple atomic layers. For example, graphite consists of many layers of graphene stacked on top of each other, but graphene has several unique properties that are not shared by graphite. Perhaps most important is graphene's highly unusual electronic structure. Electrons in graphene behave as if they have no mass and hence move through the 2D material at 1/300 the speed of light [143]. It is this property that explains graphene's excellent electrical and thermal conductivities [144]. Its hexagonal structure is incredibly strong, making it an ideal nanoscaffolding material. When rolled up, these sheets are the basis for carbon nanotubes (see Section 72.4.2.2). The quantum Hall effect was measured in graphene at low temperature very shortly after its discovery [143]. Successful measurement of the quantum Hall effect was then reported at room temperature [145]. This result is especially striking when one considers that previously the highest temperature at which the quantum Hall effect had been seen in any material was less than 50 K [146]. Given its highly attractive properties, a diverse range of applications are proposed for graphene from transparent, flexible electronics to hydrogen storage in fuel cells.

More recently, additional 2D materials have been developed, including hexagonal boron nitride, transition metal dichalcogenides, silicene, and germanene [147–152]. By stacking layers of different 2D materials (Fig. 72.19), it is hoped that their various properties can be combined in hybrid structures [152]. The properties of some of these

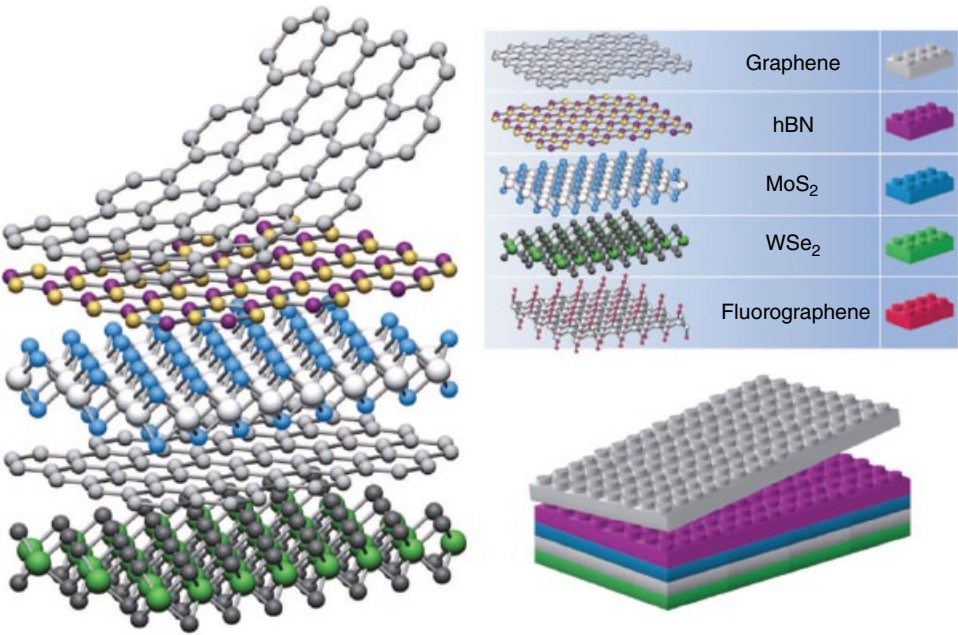


FIGURE 72.19 Hybrid structures composed from individual layers of 2D materials have been proposed. Loosely interconnected by van der Waals forces, the 2D layers will stack on top of one another like a child’s building blocks. Source: Geim and Grigorieva [152]. Reproduced with permission of Nature Publishing Group.

materials are yet to be fully elucidated, but it is expected that they will complement, or in certain areas surpass, graphene. For example, silicene and germanene are expected to possess a small semiconductor bandgap (unlike graphene) [153]. This bandgap could enable these materials to be used as the channel material in a transistor, enhancing computer chip performance while reducing size.

72.4.2 Quantum Wire Nanostructures

Continuing from left to right in Figure 72.15, we come to quantum wire nanostructures. Quantum wires are 1D nanostructures that offer quantum confinement in two dimensions. From Figure 72.16 we see that there is another big change in the DoS when moving from quantum wells to quantum wires. To create a quantum wire, you can imagine taking a quantum well and “squeezing” down one of its long edges. If the length of this edge is reduced sufficiently, then the electron energy will also be quantized in this direction.

72.4.2.1 Gate-Defined Quantum Wires In fact, literally squeezing a quantum well was precisely the approach used to create the first quantum wires [154–156]. To do this, tiny metal pads (known as gates) are placed next to a quantum well nanostructure

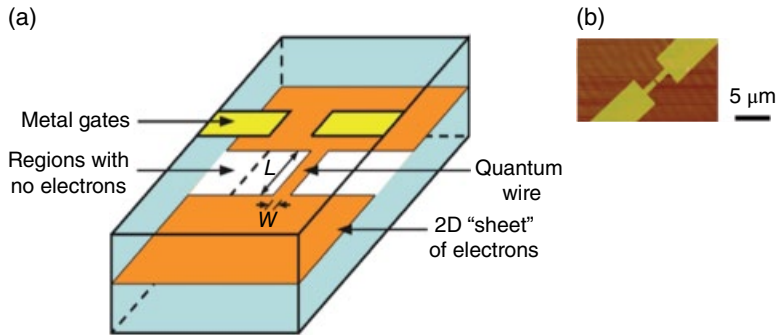


FIGURE 72.20 (a) Using electrostatic gates to create a quantum wire in a quantum well. A standard quantum well is created containing a 2D sheet of electrons. Metal gates are placed close to the quantum well. Applied gate voltage depletes electrons from underlying regions of the electron sheet. A quantum wire is formed by increasing the gate voltage to reduce the effective channel width (W) below the electron de Broglie wavelength. (b) A microscope image of a pair of gold gates on a sample of InGaAs containing a quantum well. Source: Simmonds *et al.* [157]. Reproduced with permission of the American Vacuum Society.

(Fig. 72.20a). The gates are electrically isolated from the quantum well by a thin insulating layer. Applying a negative voltage to the gates creates an electric field around each of them, and this field “depletes” the electrons from the immediately adjacent region of the quantum well. Depending on the shape and position of the metal gates, patterns can be created in the quantum well so that there are insulating areas where all the electrons have been removed and conductive areas where the electrons continue to exist as before.

Two parallel gates can be used to create a thin channel in the quantum well (Fig. 72.20a). Figure 72.20b shows a microscope image of a pair of these gates that were fabricated above a quantum well using electron beam lithography. In this example, the thin gap between the two gold finger-shaped gates is 500 nm. It is in this gap that the quantum wire is formed.

Increasing the negative gate bias raises both the strength and lateral spread of the electric field around the gates. By increasing the field we increase the size of the depleted regions of the quantum well where there are no electrons (Fig. 72.20a), which has the effect of reducing the channel width, W . Eventually, W will become small enough that it is on the order of the electron de Broglie wavelength. At this point the electron energy levels in the channel are quantized in two directions: firstly in the direction normal to the plane of the original quantum well and secondly in the direction between the two gates. Electron energy is however still continuous along the length of the channel making this a quantum wire nanostructure.

Quantum wires created in this way are typically a few hundred nanometers long. If the quantum well/wire is made from high-quality semiconductor material with a very low background of impurities, then the electron mean free path (the average distance electrons travel between collisions) can be several microns. In this situation,

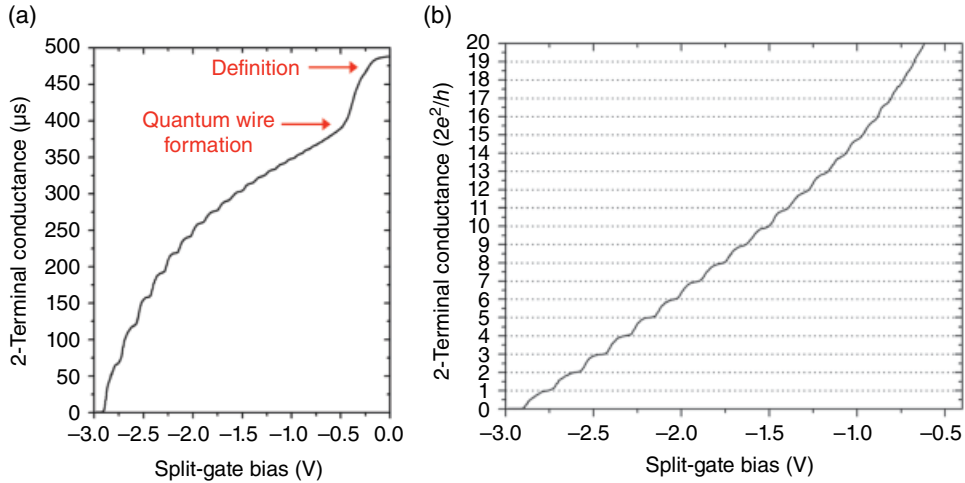


FIGURE 72.21 (a) The conductance as a function of gate bias for ballistic electron transport through a 1D channel defined in a GaAs quantum well. Below -0.5 V the conductance exhibits quantized steps. (b) The data in (a) normalized to the conductance quantum $2e^2/h$ [137].

electrons can traverse the entire length of the quantum wire without scattering, and this is known as ballistic transport. It was by studying these ultrapure 1D nanostructures that researchers in 1988 discovered new physics related to their unique quantum mechanical properties [155, 156].

As the negative gate bias is increased, the conductance of the channel initially undergoes a sharp but continuous decrease as W starts to reduce (Fig. 72.21a). This behavior is called “definition.” However, at some particular negative gate bias (around -0.5 V in Fig. 72.21a), the channel conductance abruptly changes slope signifying the formation of the quantum wire. At increasingly negative gate bias, the channel conductance is no longer continuous but becomes quantized into steps (Fig. 72.21a). When researchers first observed these plateaus in the conductance through the quantum wire it came as a surprise.

After subtracting any contact and series resistances from the measurement, researchers discovered that the height of each step was identical (Fig. 72.21b). The increase in conductance at each step is the conductance quantum $2e^2/h$ (where e is the charge of the electron and h is Planck’s constant) [155, 156]. This conductance quantum is the exact contribution to the conductance offered by electrons moving through a single 1D edge state. As the negative gate bias increases and W gets smaller, these 1D edge states are sequentially depopulated or “squeezed out” of the quantum wire and so the conductance goes down by $2e^2/h$. Using magnetic fields to separate the electrons into two populations depending on their spin, it is possible to observe additional conductance plateaus at e^2/h values [155, 157–159].

This is not the end of the story however. Below the last $2e^2/h$ plateau where the last 1D conductance channel has been depopulated and the conductance should go to zero,

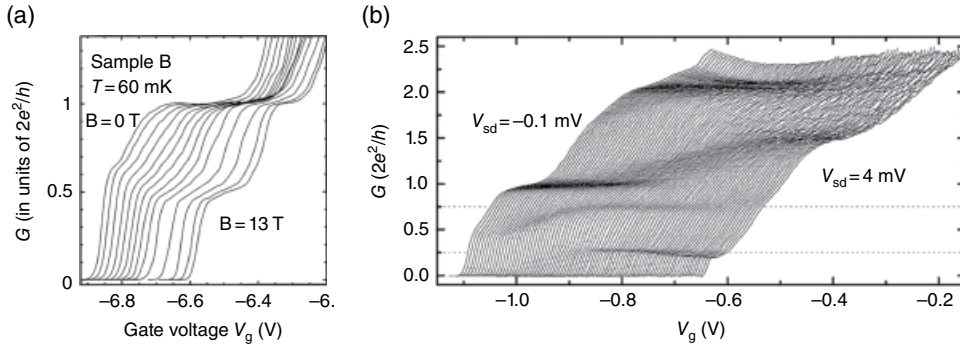


FIGURE 72.22 (a) Plateau in the conductance of a GaAs/AlGaAs quantum wire at $0.7(2e^2/h)$ showing how in the presence of an increasingly strong magnetic field it develops into a plateau at $0.5(2e^2/h)$. Source: Thomas *et al.* [160]. Reproduced with permission of the American Physical Society. (b) The evolution of sub- $(2e^2/h)$ conductance plateaus as an increasingly large voltage, V_{sd} , is applied along the length of an $\text{In}_{0.75}\text{Ga}_{0.25}\text{As}$ quantum wire (successive V_{sd} traces are offset laterally for clarity). Particularly striking is the emergence of plateau features at exactly $0.25(2e^2/h)$ and $0.75(2e^2/h)$. Source: Simmonds *et al.* [158]. Reproduced with permission of AIP Publishing LLC.

some strange behavior has been observed. Additional plateaus appear under certain measurement conditions. The first and most prominent is the so-called “0.7 structure,” which appears at $0.7(2e^2/h)$ (Fig. 72.22a) [160]. Other well-known sub- $2e^2/h$ features are seen at $0.20\text{--}0.25(2e^2/h)$ and $0.75\text{--}0.85(2e^2/h)$ (Fig. 72.22b) [158, 161, 162]. These features are believed to arise due to spontaneous spin polarization in the quantum wires as a result of electron–electron interaction. Their precise origin remains the subject both of considerable debate and a great deal of research [158, 160–165].

72.4.2.2 Carbon Nanotubes There are other ways of creating quantum wires in addition to forming a gated channel in a quantum well. The first is to use carbon nanotubes. Discovered in 1991 by Iijima, carbon nanotubes consist of one or more sheets of graphene rolled up to form a cylinder [166]. As discussed in Section 72.4.1.1, graphene is a single atomic layer of carbon atoms arranged in a hexagonal sheet with purely in-plane sp^2 bonding [142]. The precise way in which the graphene sheet is rolled can have a big effect on both the shape (Fig. 72.23) and the properties of the carbon nanotube, for example, causing it to be metallic or semiconducting in nature [167, 168]. The walls of a carbon nanotube can be as thin as just one atom, and their diameter is typically in the range $0.4\text{--}5\text{ nm}$ (depending on the number of graphene sheets rolled up), but their length can be several microns. Carbon nanotubes therefore represent almost perfect 1D nanomaterials.

To measure their 1D electronic properties, carbon nanotubes must be contacted with metal electrodes at both ends. One way to do this is to physically place a carbon nanotube on an insulating surface so that it touches two gold electrodes. By sweeping the

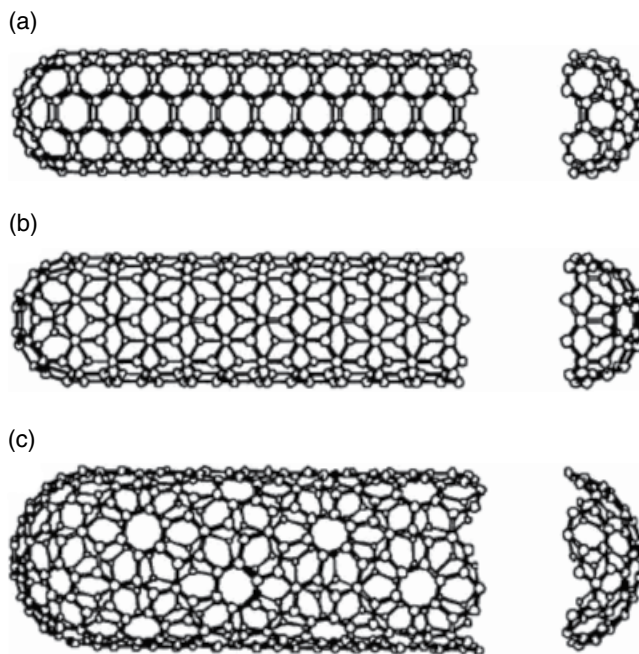


FIGURE 72.23 Three examples of the crystallographic arrangements that can be adopted by carbon nanotubes. The differences between them come from the angle at which the graphene sheet is cut before it is rolled up. The three examples previously of crystallographic arrangements adopted by carbon nanotubes are often referred to as (a) “armchair” nanotubes—graphene cut at 30° ; (b) “zigzag” nanotubes—graphene cut at 0° ; and (c) “helical” nanotubes—graphene cut at some intermediate angle. Source: Dresselhaus *et al.* [167]. Reproduced with permission of Springer.

voltage between the two electrodes and measuring the current that flows in the carbon nanotube, it is possible to map out the “ladder” of quantized 1D electron energy levels (Fig. 72.24).

As the voltage is raised, electrons in the left electrode become aligned in energy with a quantized 1D electron state in the carbon nanotube and a step in the current is observed. Each step in the current indicates that an additional 1D state in the carbon nanotube can now be accessed.

A group at the Georgia Institute of Technology used an ingenious method to measure quantized conductance in carbon nanotubes [170]. A fiber consisting of an enormous number of carbon nanotubes was created. At the tip of the fiber, a few individual carbon nanotubes were sticking out from the fiber further than their neighbors. The nanotube fiber was slowly brought down until it came into contact with a liquid metal (either mercury or gallium), which served as the second electrode. The conductance through the fiber was monitored throughout the experiment. Initially, when there is no contact, the conductance is zero. However, once the first carbon nanotube contacts the liquid metal electrode, there is a sudden jump in the conductance by G_0 , where

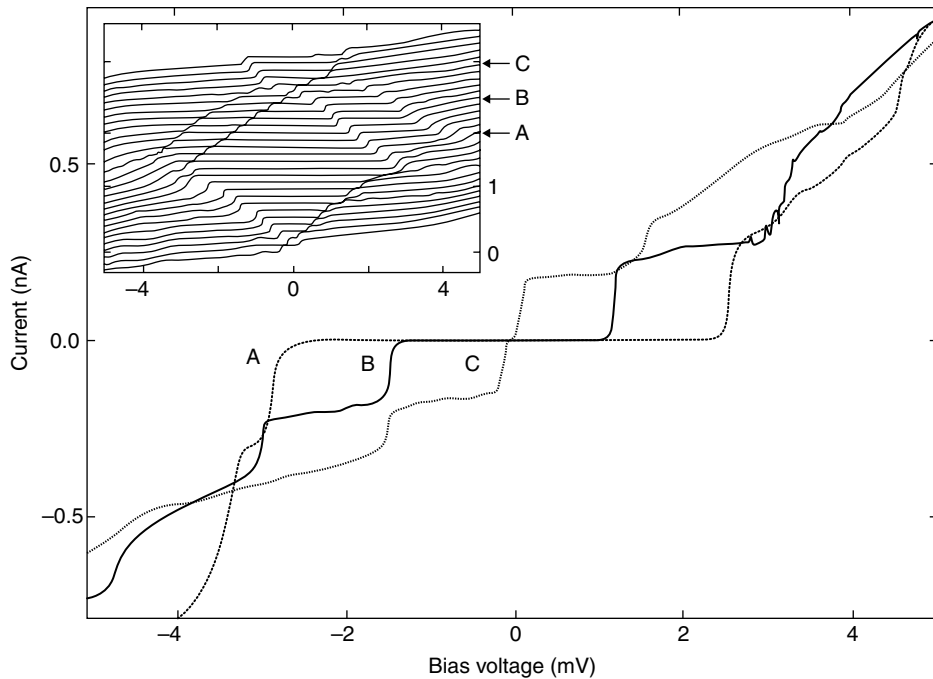


FIGURE 72.24 Current–voltage measurements of a carbon nanotube with electrical contacts at each end reveal its quantized electron levels. A third electrode is used to apply different gate bias (curves A, B, and C). Source: Tans *et al.* [169]. Reproduced with permission of Nature Publishing Group.

$G_0 = 2e^2/h$. As the fiber is pushed down further, additional nanotubes touch the metal, and, each time, a step of G_0 in the conductance is seen. $2e^2/h$ is the same conductance quantum that was measured for each 1D edge state in the gate-defined quantum wires earlier. This experiment proves that carbon nanotubes behave as perfect 1D quantum wires [167, 168, 170].

In addition to their 1D electrical properties, carbon nanotubes are incredibly useful for a very wide range of applications [171]. Due to their sp^2 bonding, they couple enormous mechanical strength under tension with very low weight, making them an excellent additive for strengthening anything from construction materials to clothing. They can be grown in macroscopic lengths of more than 10 cm, which, given their nanoscale diameter, corresponds to extraordinary aspect ratios in the hundreds of millions [172]. Given their excellent electrical properties, carbon nanotubes with such enormous lengths could be used to directly build circuits and devices. They can be used as tips in various scanning probe microscopy techniques and are even a highly promising candidate for hydrogen storage, which is a key challenge in the development of practical fuel cell technology. Their diverse properties mean that current applications for carbon nanotubes range from biomedicine to clean energy [173, 174].

72.4.2.3 Quantum Wires in Self-Assembled Nanopillars In an offshoot from the work on self-assembled nanopillars discussed in Section 72.3.12, researchers have shown that individual nanopillars can be turned into quantum wires.

Individual semiconductor nanopillars are grown, snapped off and placed on an insulating substrate, and finally contacted with electrodes similar to the carbon nanotube in Figure 72.24a. Since self-assembled nanopillars typically have minimum diameters in the region of 50 nm, gate electrodes are used to electrostatically reduce the effective pillar width until it falls below the electron de Broglie wavelength and quantization occurs [175]. Evidence exists that exotic new quasiparticles, called Majorana fermions, are observed when a quantum wire formed in an InSb nanopillar is brought into contact with a superconductor due to the strong spin–orbit coupling in the semiconductor [176].

1D conductance quantization has also been demonstrated in atomically thin metal wires. Quantum wires are created by mechanically pulling apart nanoscale gold contacts while continually measuring the electrical conductance. Chains of at least four gold atoms form in which the highest measured conductance is $2e^2/h$, proving that they are 1D in nature. Electrons traverse these gold atomic chains without scattering meaning that they could be very useful for nanoscale electronics [177].

72.4.2.4 Epitaxial Self-Assembled Quantum Wires Finally, arrays of self-assembled quantum wires can be created during the growth of certain compressively strained semiconductor heterostructures. When InAs is deposited on InP(001) (or several of its lattice-matched alloys: InAlAs, InGaAs, AlAsSb, and GaAsSb), it spontaneously forms thin “dash-like” nanostructures parallel to the $[-110]$ direction (Fig. 72.25) [178–180]. These nanostructures are typically less than 5 nm high and 15–20 nm wide but can be 500 nm or more in length fulfilling the requirements for a quantum wire.

The InAs quantum wires self-assemble via the Stranski–Krastanov mode where growth initially proceeds by the formation of a flat 2D wetting layer. However, after

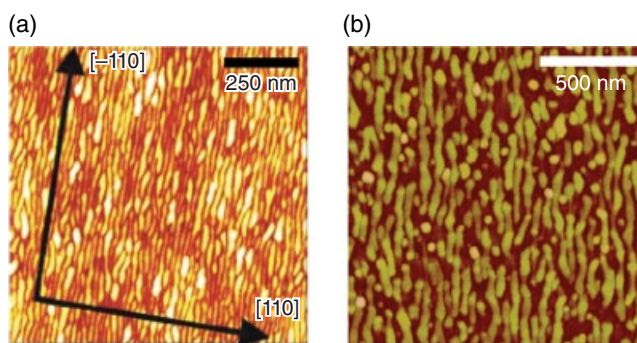


FIGURE 72.25 Self-assembled InAs quantum wires formed on (a) InAlAs (Source: Simmonds *et al.* [178]. Reproduced with permission of the American Vacuum Society) and (b) GaAsSb surfaces (Source: Simmonds *et al.* [179]. Reproduced with permission of AIP Publishing LLC). Regardless of the buffer material, the InAs wires form parallel to the $[-110]$ direction.

the deposition of some critical thickness, there is a sudden transition to 3D growth due to the accumulation of compressive strain in the InAs layer and the quantum wires form. The anisotropic self-assembly of the wires is attributed to the fact that growth is energetically more favorable in the $[-110]$ direction [178, 180].

This common InAs wetting layer and the fact that the InAs nanostructures are closely packed into arrays mean that measuring the electrical properties of a single quantum wire is not possible. Quantized conductance cannot therefore be demonstrated in these quantum wires, but their 1D nature has been confirmed by mapping out their density of states (Fig. 72.26). A qualitative comparison of the 10 K curves in Figure 72.26 with

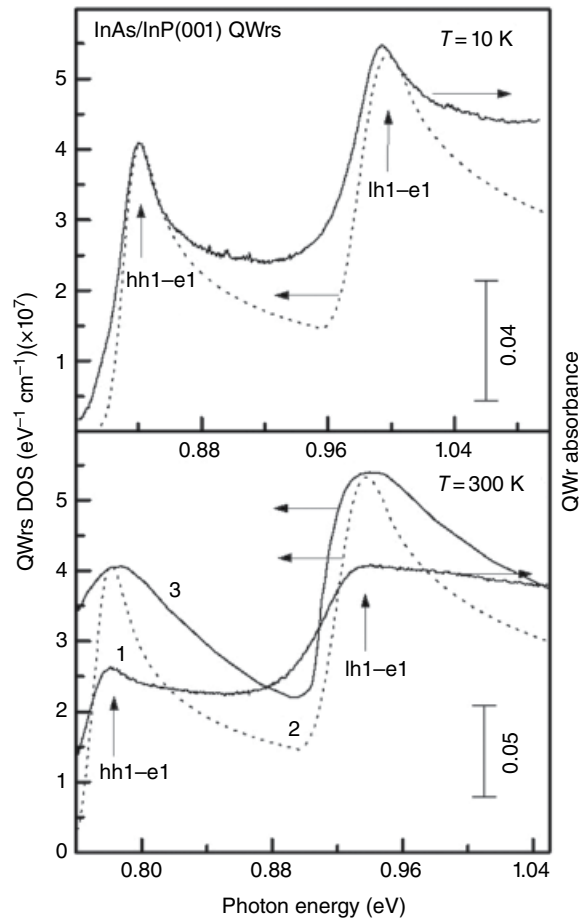


FIGURE 72.26 Comparison between simulated density of states for self-assembled InAs quantum wires on InP(001) and their experimental optical absorption spectra at 10 K (top) and 300 K (bottom). In the top panel the simulated (experimental) data are the dotted (solid) lines. In the bottom panel the experimental data are line 1, while two density of states curves simulated to take into account different broadening mechanisms appear as lines 2 and 3. Source: Mazur *et al.* [181]. Reproduced with permission of IOP Publishing.

the 1D DoS in Figure 72.16 confirms the characteristic shape of a 1D quantum system and indicates that these are indeed quantum wires.

Self-assembled InAs quantum wires have found multiple applications. Their high density of states at each quantized electron energy level is helpful for reducing the threshold current density in lasers [181–183]. What is more, the wavelength of light emitted by InAs quantum wires on InP and its alloys is highly tunable. Changing the size of the wires by changing the amount of InAs deposited changes the position of the quantized energy levels (in the same way as we saw for quantum well nanostructures). The emission wavelength can readily be tuned over a very wide range from 1.1 to 1.8 μm (Fig. 72.27). This range covers the 1.3 and 1.55 μm wavelengths that are critically important to fiber optic-based telecommunications [181].

Until very recently, the only way to create self-assembled quantum wires was to use compressive strain. However, a method for creating self-assembled quantum wires from tensile-strained semiconductors has recently been discovered [184]. The underlying growth processes by which these tensile-strained nanostructures self-assemble are very similar to those in traditional compressively strained self-assembly [185–187]. However, the tensile strain has some interesting and potentially very useful effects on the properties of the quantum wires. Perhaps the most striking result is that unlike compressive strain, which increases the semiconductor bandgap of the quantum wire, tensile strain *reduces* the bandgap. As a result, tensile-strained quantum wires emit light at longer wavelengths than can be reached in either bulk semiconductor material or in

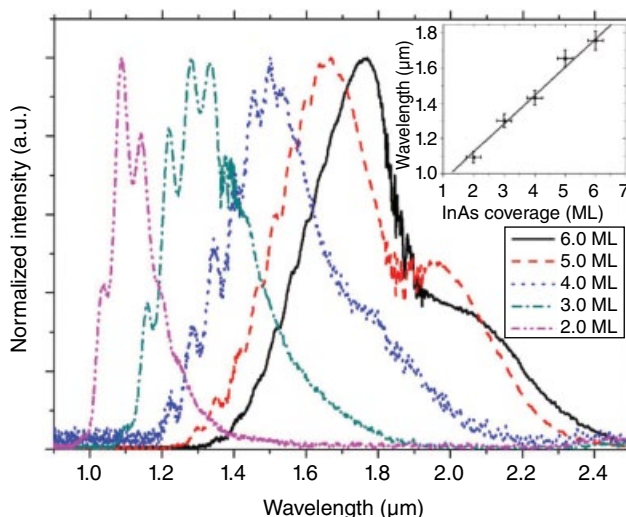


FIGURE 72.27 Photoluminescence from InAs quantum wires measured at 4.2 K showing straightforward control of the emission wavelength with InAs deposition thickness in monolayers (ML). (Inset) Peak emission wavelength from the 5 PL spectra plotted against InAs coverage revealing a linear dependence of 170 nm/ML. Source: Simmonds *et al.* [178]. Reproduced with permission of the American Vacuum Society.

compressive-strained quantum wires. This push toward the infrared end of the electromagnetic spectrum could have important implications for technologies including night vision, astronomy, and trace gas sensing [184].

72.4.3 Quantum Dot Nanostructures

Taking the progression from left to right in Figure 72.15 to its logical conclusion, we reach quantum dots, in which electrons experience quantum confinement in all three spatial dimensions [188]. The theoretical 0D DoS consists of a series of delta functions at discrete energies, although in reality each energy level is somewhat broadened due to inhomogeneity and thermal effects. This 0D DoS has many practical benefits for device applications. Using the example of the laser again, quantum dot lasers exhibit low threshold currents, high gains, and large T_0 values (i.e., excellent thermal stability of the threshold current) [189]. In addition, the discrete electron energy levels in a quantum dot should result in highly monochromatic light emission. By changing the quantum dot size, the emission wavelength can be tuned, as for both quantum wells and wires. The unique electronic structure of quantum dots places them at the forefront of research into a wide variety of cutting-edge technologies, including quantum computation and quantum cryptography [190–192].

72.4.3.1 Gate-Defined Quantum Dots A 2D sheet of electrons in a quantum well can be patterned into quantum dots using electrostatic gates (in the same way as we saw for quantum wires) [193, 194]. In this case, the gates are shaped to create quasicircular regions in the electron sheet. Controlling the gate voltages allows the size of the circular region to be shrunk until its size is less than the de Broglie wavelength in both in-plane directions and the out-of-plane direction of the quantum well. Once inside the quantum dot, the energy of an electron is quantized in all three directions.

Electrons can even be pushed in and out of the quantum dots using additional gates as “plungers.” Indeed, due to the ease with which additional gates can be added, arrays of multiple quantum dots can be created in almost any geometry. Such versatile systems serve as a rich toolbox for the exploration of phenomena such as coupling between adjacent quantum dots, quantum computation, or even as part of a quantum refrigerator (Fig. 72.28) [195–199].

72.4.3.2 Colloidal Quantum Dots In Section 72.3.13 we discussed this family of quantum dot nanoparticles in relation to their applications for nanomedicine [95]. However, as well as their size and the ease with which they can be functionalized, the quantum mechanical properties of colloidal dots are also of great interest. These solution-based quantum dots are created using chemical reactions between various metal oxide and metal organic precursors [200]. This approach is extremely versatile. Colloidal quantum dot nanoparticles can be created from a range of different materials (most commonly the II–VI family of semiconductors, such as CdSe and CdTe) [201].

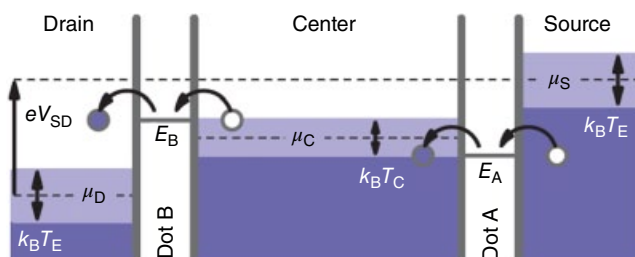


FIGURE 72.28 A quantum dot refrigerator created by connecting three 2D sheets of electrons in quantum wells with two quantum dots. An electron flows from the Source into the Center via the quantized energy level E_A in dot A. The quantized energy of the Center 2D electron sheet is thermally broadened. The quantized energy level E_B of dot B is made to be coincident with that of the hottest electrons in the Center so that these are selected to flow out to the Drain. In summary, electrons are injected into the Center with low energy and removed with higher energy such that the overall amount of thermal energy removed is $E_B - E_A$ and this cools the Center region. Source: Prance *et al.* [195]. Reproduced with permission of the American Physical Society.



FIGURE 72.29 Depending on their size, colloidal quantum dots dispersed in hexane emit light across the entire visible spectrum (wavelengths 430–620 nm) when viewed under UV light. The inset image shows the QDs illuminated by the tungsten room light without UV irradiation. Source: Kwak *et al.* [203]. Reproduced with permission of the American Chemical Society.

Coating a low bandgap nanoparticle with a larger bandgap shell creates the conditions necessary for quantum confinement of electrons [202]. Light emission wavelength can be readily tuned by controlling the core diameter of these core-shell quantum dots (Fig. 72.29) [203, 204].

As we have seen, colloidal quantum dots are ideally suited to functionalization for use in biomedical applications [95, 97–100]. However, these colloidal quantum dots

can additionally be used as building blocks and packed together into 3D superlattices [205]. Their solution-based nature means that thin films of these quantum dots can be easily prepared using standard spin-coating or inkjet printing techniques [206]. These simple production techniques are significantly cheaper than other synthesis techniques such as molecular beam epitaxy or metal organic chemical vapor deposition. The attractive properties and cost-effectiveness of colloidal quantum dots mean that they are being developed for use in a wide range of electronic and optoelectronic devices. For a comprehensive review of colloidal quantum dots, their properties, and their applications, see Ref. 207.

72.4.3.3 Epitaxial Self-Assembled Quantum Dots Similar to the self-assembled nanowires discussed in Section 72.4.2.4, the strain in semiconductor materials can be harnessed to drive the creation of quantum dot nanostructures. Traditionally, *compressive* strain has been used as the driving force for the growth of these nanomaterials. The two most commonly explored self-assembled systems are In(Ga)As quantum dots in a GaAs matrix [208] and Ge quantum dots in a Si matrix [209]. Since their discovery in the early 1990s, these and other quantum dot systems have been explored extensively, and their properties employed in a wide range of device applications, including lasers, solar cells, and quantum computers [119, 179, 192, 210–215].

Self-assembled quantum dots usually nucleate and grow at random locations across the substrate surface (Fig. 72.30 (left)). However, by patterning the substrate before growth, quantum dots can be made to nucleate at only at predetermined positions (Fig. 72.30 (right)) [216–218]. This technique can be used to generate quantum dot molecules or to assist with locating the quantum dots during subsequent device fabrication. For example, one area of current interest is to create photonic crystal cavities

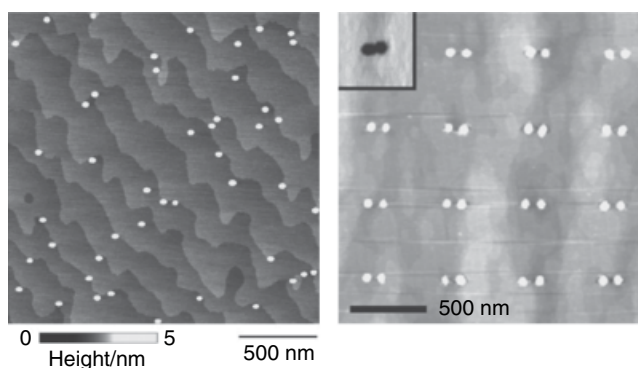


FIGURE 72.30 (Left) InAs self-assembled quantum dots typically form at random locations across the surface of a GaAs substrate. (Right) If small pits are patterned into the GaAs surface prior to InAs deposition, the quantum dots can be made to form only in the location of the pits, leading to an ordered array. In this case the growth conditions have been controlled such that the quantum dots assemble in pairs. Source: Atkinson *et al.* [216]. Reproduced with permission of Elsevier.

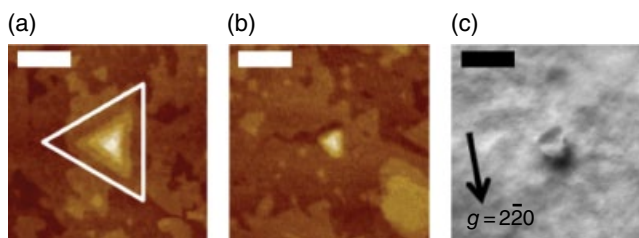


FIGURE 72.31 (a) Atomic force micrograph of a large tensile-strained quantum dot showing its symmetric triangular shape (the white equilateral triangle is drawn for comparison). (b) A tensile quantum dot of smaller, more typical size. (c) Plan-view transmission electron micrograph of a capped quantum dot showing that it is under tensile strain but free from strain-related defects. All scale bars are 100 nm. Source: Yerino *et al.* [221]. Reproduced with permission of AIP Publishing LLC.

that contain a single quantum dot. Interesting optical effects occur when the emission from the quantum dot coincides with the resonance of the photonic crystal cavity. These hybrid structures can be created from randomly dispersed quantum dot layers by creating many photonic crystal cavities and then looking for one that contains a single dot [219]. However, a more reliable, efficient, and high-yield approach is first to know precisely where the quantum dots are and then to create the photonic crystals around them [216].

More recently a method that uses *tensile* strain to drive quantum dot self-assembly has also been discovered [220, 221]. Like the tensile-strained quantum wires discussed at the end of Section 72.4.2.4, the residual tensile strain in these quantum dots causes them to emit light at longer wavelengths than is possible in either the bulk material or compressively strained quantum dots. This new way of creating self-assembled quantum dots opens up the possibility of creating families of quantum nanomaterials with small bandgaps, a feat that previously has been very difficult. What is more, these tensile-strained quantum dots can be grown with extremely high symmetry (Fig. 72.31), which gives them a very low fine structure splitting. This property makes them ideally suited to quantum information applications, such as quantum computing and quantum cryptography [221].

It is also possible to grow quantum dot nanostructures without using strain. The most common approach is known as droplet epitaxy. The first step is to deposit a tiny amount of metal onto the surface, which organizes itself into discrete liquid nanodroplets dispersed across the surface. The second step is to crystalize the semiconductor material out of these metal droplets by annealing. For example, to create an unstrained InAs quantum dot on GaAs, one begins by depositing pure indium metal to form the nanodroplets. Then, to turn the metal into InAs, the nanodroplets are annealed under an arsenic atmosphere, and the III–V semiconductor crystalizes. This approach allows quantum dots to be created on traditionally incompatible substrates where the strain-driven approach fails [222, 223]. In addition, this approach can also be used to create quantum dot molecules that may find application in quantum computation [224].

72.5 CONCLUSIONS

Given the enormous scope of current research into the development of nanomaterials with new and exciting properties, in this chapter it has been impossible to do more than give a flavor of some of the major areas of exploration. As a result there are many omissions and highly simplified or abridged descriptions of complex topics that I would have loved to address more completely given more space. Nevertheless, I trust that the interested reader will be able to use the references provided as a starting point from which to explore these areas in more detail.

Despite these caveats, it is my hope that the foregoing sections have provided a useful introduction to nanomaterials and their properties. In particular, I have focused on the influence of the high surface-to-volume ratio and the role of quantum confinement on nanomaterial properties. In doing so, I have attempted to explain why material properties at the nanoscale can be, at times, very different from those at the macroscale. From color and magnetic behavior to electronic structure, nanomaterials provide us with an opportunity to engineer and fine-tune the fundamental properties of a wide range of materials. As we have seen, numerous exciting scientific discoveries and hugely influential technological advances owe everything to the unique properties of nanomaterials. Ongoing research into nanomaterials will drive future innovation.

REFERENCES

1. A. B. Kesel, A. Martin, T. Seidl, Getting a grip on spider attachment: an AFM approach to microstructure adhesion in arthropods, *Smart Mater. Struct.* 13, 512–518 (2004).
2. Y. T. Cheng, D. E. Rodak, C. A. Wong, C. A. Hayden, Effects of micro- and nano-structures on the self-cleaning behaviour of lotus leaves, *Nanotechnology* 17, 1359–1362 (2006).
3. V. Saranathan *et al.*, Structure, function, and self-assembly of single network gyroid (I4132) photonic crystals in butterfly wing scales, *Proc. Natl. Acad. Sci.* 107, 11676–11681 (2010).
4. R. A. Shelby, D. R. Smith, S. Schultz, Experimental verification of a negative index of refraction, *Science* 292, 77–80 (2001).
5. J. D. Joannopoulos, P. R. Villeneuve, S. Fan, Photonic crystals: putting a new twist on light, *Nature* 386, 143–149 (1997).
6. E. Yablonovitch, Inhibited spontaneous emission in solid-state physics and electronics, *Phys. Rev. Lett.* 58, 2059–2062 (1987).
7. E. Yablonovitch, Photonic band-gap structures, *J. Opt. Soc. Am. B* 10, 283–295 (1993).
8. D. R. Smith, W. J. Padilla, D. C. Vier, S. C. Nemat-Nasser, S. Schultz, Composite medium with simultaneously negative permeability and permittivity, *Phys. Rev. Lett.* 84, 4184–4187 (2000).
9. D. Schurig *et al.*, Metamaterial electromagnetic cloak at microwave frequencies, *Science* 314, 977–980 (2006).
10. E. Roduner, Size matters: why nanomaterials are different, *Chem. Soc. Rev.* 35, 583–592 (2006).
11. A. Y. Cho, J. R. Arthur, Molecular beam epitaxy, *Prog. Solid State Chem.* 10, 157–191 (1975).

12. H. M. Manasevit, Single-crystal gallium arsenide on insulating substrates, *Appl. Phys. Lett.* 12, 156–159 (1968).
13. H. M. Manasevit, W. I. Simpson, The use of metal-organics in the preparation of semiconductor materials I. Epitaxial gallium-V compounds, *J. Electrochem. Soc.* 116, 1725–1732 (1969).
14. W.-D. Li, W. Wu, R. S. Williams, *Single-Digit Nanometer Nanoimprint Templates*, SPIE Newsroom, Bellingham, WC, 2013.
15. X. Wen, L. M. Traverso, P. Srisungsitthisunti, X. Xu, E. E. Moon, Optical nanolithography with $\lambda/15$ resolution using bowtie aperture array, *Appl. Phys. A* 117, 307–311 (2014).
16. V. Auzelyte *et al.*, Extreme ultraviolet interference lithography at the Paul Scherrer Institut, *J. Micro/Nanolithography, MEMS, MOEMS* 8, 021204 (2009).
17. L. Pan *et al.*, Maskless plasmonic lithography at 22 nm resolution, *Sci. Rep.* 1, 175 (2011).
18. Y. Kyoung Ryu, P. Aitor Postigo, F. Garcia, R. Garcia, Fabrication of sub-12 nm thick silicon nanowires by processing scanning probe lithography masks, *Appl. Phys. Lett.* 104, 223112 (2014).
19. D. Mijatovic, J. Eijkel, A. van den Berg, Technologies for nanofluidic systems: top-down vs. bottom-up—a review, *Lab Chip* 5, 492–500 (2005).
20. B. A. Joyce, D. D. Vvedensky, Self-organized growth on GaAs surfaces, *Mater. Sci. Eng. R* 46, 127–176 (2004).
21. J. W. van Honschoten, N. Brunets, N. R. Tas, Capillarity at the nanoscale, *Chem. Soc. Rev.* 39, 1096–1114 (2010).
22. P. Serp, K. Philippot, Eds., *Nanomaterials in Catalysis*, Wiley-VCH, Weinheim, Germany, 2013.
23. X. D. Wang, C. J. Summers, Z. L. Wang, Mesoporous single-crystal ZnO nanowires epitaxially sheathed with Zn₂SiO₄, *Adv. Mater.* 16, 1215–1218 (2004).
24. D. M. Eigler, E. K. Schweizer, Positioning single atoms with a scanning tunnelling microscope, *Nature* 344, 524–526 (1990).
25. U. Kreibig, M. Vollmer, *Optical Properties of Metal Clusters*, Springer, Berlin, 1995.
26. P. Buffat, J.-P. Borel, Size effect on the melting temperature of gold particles, *Phys. Rev. A* 13, 2287–2298 (1976).
27. K. M. Unruh, T. E. Huber, C. A. Huber, Melting and freezing behavior of indium metal in porous glasses, *Phys. Rev. B* 48, 9021–9027 (1993).
28. K. Koga, T. Ikeshoji, K. Sugawara, Size- and temperature-dependent structural transitions in gold nanoparticles, *Phys. Rev. Lett.* 92, 115507 (2004).
29. Y. Nakae *et al.*, Anomalous spin polarization in Pd and Au nano-particles, *Phys. B* 284–288, 1758–1759 (2000).
30. Y. Yamamoto *et al.*, Magnetic properties of the noble metal nanoparticles protected by polymer, *Phys. B* 329–333, 1183–1184 (2003).
31. N. Watari, S. Ohnishi, Atomic and electronic structures of Pd₁₃ and Pt₁₃ clusters, *Phys. Rev. B* 58, 1665–1677 (1998).
32. Y. Sakamoto *et al.*, Ferromagnetism of Pt nanoparticles induced by surface chemisorption, *Phys. Rev. B* 83, 104420 (2011).

33. C. Q. Sun, Dominance of broken bonds and nonbonding electrons at the nanoscale, *Nanoscale* 2, 1930–1961 (2010).
34. G. Polizos *et al.*, Scalable superhydrophobic coatings based on fluorinated diatomaceous earth: abrasion resistance versus particle geometry, *Appl. Surf. Sci.* 292, 563–569 (2014).
35. B. D’Urso, J. T. Simpson, M. Kalyanaraman, Nanocone array glass, *J. Micromech. Microeng.* 17, 717–721 (2007).
36. T. Aytug *et al.*, Optically transparent, mechanically durable, nanostructured superhydrophobic surfaces enabled by spinodally phase-separated glass thin films, *Nanotechnology* 24, 315602 (2013).
37. W. Sparreboom, A. van den Berg, J. C. T. Eijkel, Transport in nanofluidic systems: a review of theory and applications, *New J. Phys.* 12, 015004 (2010).
38. J. M. Oh, T. Faez, S. de Beer, F. Mugele, Capillarity-driven dynamics of water–alcohol mixtures in nanofluidic channels, *Microfluid. Nanofluid.* 9, 123–129 (2009).
39. R. B. Schoch, J. Han, P. Renaud, Transport phenomena in nanofluidics, *Rev. Mod. Phys.* 80, 839–883 (2008).
40. J. C. T. Eijkel, A. van den Berg, Nanofluidics: what is it and what can we expect from it? *Microfluid. Nanofluid.* 1, 249–267 (2005).
41. A. H. J. Yang *et al.*, Optical manipulation of nanoparticles and biomolecules in sub-wavelength slot waveguides, *Nature* 457, 71–75 (2009).
42. J. O. Tegenfeldt *et al.*, Micro- and nanofluidics for DNA analysis, *Anal. Bioanal. Chem.* 378, 1678–1692 (2004).
43. J. Goldberger, R. Fan, P. Yang, Inorganic nanotubes: a novel platform for nanofluidics, *Acc. Chem. Res.* 39, 239–248 (2006).
44. S. S.-Y. Chui, S. M.-F. Lo, J. P. H. Charmant, A. G. Orpen, I. D. Williams, A chemically functionalizable nanoporous material $[\text{Cu}_3(\text{TMA})_2(\text{H}_2\text{O})_3]_n$, *Science* 283, 1148–1150 (1999).
45. D. Bradshaw, J. B. Claridge, E. J. Cussen, T. J. Prior, M. J. Rosseinsky, Design, chirality, and flexibility in nanoporous molecule-based materials, *Acc. Chem. Res.* 38, 273–282 (2005).
46. F. Ç. Cebeci, Z. Wu, L. Zhai, R. E. Cohen, M. F. Rubner, Nanoporosity-driven superhydrophilicity: a means to create multifunctional antifogging coatings, *Langmuir* 22, 2856–2862 (2006).
47. C. Sanchez, C. Boissière, D. Grosso, C. Laberty, L. Nicole, Design, synthesis, and properties of inorganic and hybrid thin films having periodically organized nanoporosity, *Chem. Mater.* 20, 682–737 (2008).
48. A. Krotkus, Semiconductors for terahertz photonics applications, *J. Phys. D: Appl. Phys.* 43, 273001 (2010).
49. D. V. Seletskiy *et al.*, Efficient terahertz emission from InAs nanowires, *Phys. Rev. B* 84, 115421 (2011).
50. V. N. Trukhin *et al.*, Terahertz generation by GaAs nanowires, *Appl. Phys. Lett.* 103, 072108 (2013).
51. A. Atrashchenko *et al.*, Giant enhancement of terahertz emission from nanoporous GaP, *Appl. Phys. Lett.* 105, 191905 (2014).

52. M. Reid, I. V. Cravetchi, R. Fedosejevs, I. M. Tiginyanu, L. Sirbu, Enhanced terahertz emission from porous InP (111) membranes, *Appl. Phys. Lett.* 86, 021904 (2005).
53. R. Adomavičius *et al.*, Terahertz pulse emission from nanostructured (311) surfaces of GaAs, *J. Infrared Millimeter Terahertz Waves* 33, 599–604 (2012).
54. A.-H. Lu, F. Schüth, Nanocasting: a versatile strategy for creating nanostructured porous materials, *Adv. Mater.* 18, 1793–1805 (2006).
55. J. Erlebacher, M. J. Aziz, A. Karma, N. Dimitrov, K. Sieradzki, Evolution of nanoporosity in dealloying, *Nature* 410, 450–453 (2001).
56. X. Zhao *et al.*, Influence of surface properties on the electrical conductivity of silicon nano-membranes, *Nanoscale Res. Lett.* 6, 402 (2011).
57. C. Jiang, V. V. Tsukruk, Freestanding nanostructures via layer-by-layer assembly, *Adv. Mater.* 18, 829–840 (2006).
58. C. Jiang, S. Markutsya, Y. Pikus, V. V. Tsukruk, Freely suspended nanocomposite membranes as highly sensitive sensors, *Nat. Mater.* 3, 721–728 (2004).
59. R. Narayanan, M. A. El-Sayed, Catalysis with transition metal nanoparticles in colloidal solution: nanoparticle shape dependence and stability, *J. Phys. Chem. B* 109, 12663–12676 (2005).
60. Z. Peng, H. Yang, Designer platinum nanoparticles: control of shape, composition in alloy, nanostructure and electrocatalytic property, *Nano Today* 4, 143–164 (2009).
61. C.-J. Zhong *et al.*, Fuel cell technology: nano-engineered multimetallic catalysts, *Energy Environ. Sci.* 1, 454–466 (2008).
62. B. Seger, P. V. Kamat, Electrocatalytically active graphene-platinum nanocomposites. Role of 2-D carbon support in PEM fuel cells, *J. Phys. Chem. C* 113, 7990–7995 (2009).
63. P. H. Matter, L. Zhang, U. S. Ozkan, The role of nanostructure in nitrogen-containing carbon catalysts for the oxygen reduction reaction, *J. Catal.* 239, 83–96 (2006).
64. M.-L. Seol, J.-H. Ahn, J.-M. Choi, S.-J. Choi, Y.-K. Choi, Self-aligned nanoforest in silicon nanowire for sensitive conductance modulation., *Nano Lett.* 12, 5603–5608 (2012).
65. K. A. Bertness, N. A. Sanford, A. V. Davydov, GaN nanowires grown by molecular beam epitaxy, *IEEE J. Sel. Top. Quantum Electron.* 17, 847–858 (2011).
66. A. Lin *et al.*, Extracting transport parameters in GaAs nanopillars grown by selective-area epitaxy, *Nanotechnology* 23, 105701 (2012).
67. A. C. Scofield *et al.*, Bottom-up photonic crystal cavities formed by patterned III-V nanopillars, *Nano Lett.* 11, 2242 (2011).
68. T. Mårtensson *et al.*, Epitaxial III–V nanowires on silicon, *Nano Lett.* 4, 1987–1990 (2004).
69. G. Mariani, A. Scofield, C.-H. Hung, D. L. Huffaker, GaAs nanopillar-array solar cells employing in situ surface passivation, *Nat. Commun.* 4, 1497 (2013).
70. R. Yan, D. Gargas, P. Yang, Nanowire photonics, *Nat. Photonics* 3, 569–576 (2009).
71. F. Qian, S. Gradečák, Y. Li, C.-Y. Wen, C. M. Lieber, Core/multishell nanowire heterostructures as multicolor, high-efficiency light-emitting diodes, *Nano Lett.* 5, 2287–2291 (2005).
72. M. Borgström, K. Deppert, L. Samuelson, W. Seifert, Size- and shape-controlled GaAs nano-whiskers grown by MOVPE: a growth study, *J. Cryst. Growth* 260, 18–22 (2004).
73. B. Mandl *et al.*, Growth mechanism of self-catalyzed group III–V nanowires, *Nano Lett.* 10, 4443–4449 (2010).

74. H. Sekiguchi, K. Kishino, A. Kikuchi, Emission color control from blue to red with nanocolumn diameter of InGaN/GaN nanocolumn arrays grown on same substrate, *Appl. Phys. Lett.* 96, 231104 (2010).
75. Z. Fan *et al.*, Ordered arrays of dual-diameter nanopillars for maximized optical absorption, *Nano Lett.* 10, 3823–3827 (2010).
76. A. C. Scofield *et al.*, Bottom-up photonic crystal lasers, *Nano Lett.* 11, 5387–5390 (2011).
77. J. C. Johnson *et al.*, Single gallium nitride nanowire lasers, *Nat. Mater.* 1, 106–110 (2002).
78. P. Senanayake *et al.*, Photoconductive gain in patterned nanopillar photodetector arrays, *Appl. Phys. Lett.* 97, 203108 (2010).
79. G. Mariani *et al.*, Hybrid conjugated polymer solar cells using patterned GaAs nanopillars, *Appl. Phys. Lett.* 97, 013107 (2010).
80. K. Tomioka, M. Yoshimura, T. Fukui, A III–V nanowire channel on silicon for high-performance vertical transistors, *Nature* 488, 189–192 (2012).
81. T. Tanaka *et al.*, Vertical surrounding gate transistors using single InAs nanowires grown on Si substrates, *Appl. Phys. Express* 3, 025003 (2010).
82. Z. L. Wang, Piezopotential gated nanowire devices: piezotronics and piezo-phototronics, *Nanotoday* 5, 540–552 (2010).
83. C. J. Barrelet, A. B. Greytak, C. M. Lieber, Nanowire photonic circuit elements, *Nano Lett.* 4, 1981–1985 (2004).
84. F. Caruso, Nanoengineering of particle surfaces, *Adv. Mater.* 13, 11–22 (2001).
85. M. D. Stoller, S. Park, Y. Zhu, J. An, R. S. Ruoff, Graphene-based ultracapacitors, *Nano Lett.* 8, 3498–3502 (2008).
86. T. Ramanathan *et al.*, Functionalized graphene sheets for polymer nanocomposites, *Nat. Nanotechnol.* 3, 327–331 (2008).
87. Z. Liu, J. T. Robinson, X. Sun, H. Dai, PEGylated nanographene oxide for delivery of water-insoluble cancer drugs, *J. Am. Chem. Soc.* 130, 10876–10877 (2008).
88. D. Tasis, N. Tagmatarchis, A. Bianco, M. Prato, Chemistry of carbon nanotubes, *Chem. Rev.* 106, 1105–1136 (2006).
89. J. Chen *et al.*, Solution properties of single-walled carbon nanotubes, *Science* 282, 95–98 (1998).
90. R. J. Chen, Y. Zhang, D. Wang, D. Hongjie, Noncovalent sidewall functionalization of single-walled carbon nanotubes for protein immobilization, *J. Am. Chem. Soc.* 123, 3838–3839 (2001).
91. Y. Cui, Q. Wei, H. Park, C. M. Lieber, Nanowire nanosensors for highly sensitive and selective detection of biological and chemical species, *Science* 293, 1289–1292 (2001).
92. A. Kolmakov, D. O. Klenov, Y. Lilach, S. Stemmer, M. Moskovits, Enhanced gas sensing by individual SnO₂ nanowires and nanobelts functionalized with Pd catalyst particles, *Nano Lett.* 5, 667–673 (2005).
93. Y. L. Bunimovich *et al.*, Quantitative real-time measurements of DNA hybridization with alkylated nonoxidized silicon nanowires in electrolyte solution, *J. Am. Chem. Soc.* 128, 16323–16331 (2006).
94. X. Michalet *et al.*, Quantum dots for live cells, in vivo imaging, and diagnostics, *Science* 307, 538–544 (2005).

95. X. Gao, Y. Cui, R. M. Levenson, L. W. K. Chung, S. Nie, In vivo cancer targeting and imaging with semiconductor quantum dots, *Nat. Biotechnol.* 22, 969–976 (2004).
96. I. L. Medintz, H. T. Uyeda, E. R. Goldman, H. Mattoussi, Quantum dot bioconjugates for imaging, labelling and sensing, *Nat. Mater.* 4, 435–446 (2005).
97. J. Gao, H. Gu, B. Xu, Multifunctional magnetic nanoparticles: design, synthesis, and biomedical applications, *Acc. Chem. Res.* 42, 1097–1107 (2009).
98. V. Bagalkot *et al.*, Quantum dot-aptamer conjugates for synchronous cancer imaging, therapy, and sensing of drug delivery based on bi-fluorescence resonance energy transfer, *Nano Lett.* 7, 3065–3070 (2007).
99. C. Minelli, S. B. Lowe, M. M. Stevens, Engineering nanocomposite materials for cancer therapy, *Small* 6, 2336–2357 (2010).
100. D. Kim, Y. Y. Jeong, S. Jon, A drug-loaded aptamer-gold nanoparticle bioconjugate for combined CT imaging and therapy of prostate cancer, *ACS Nano* 4, 3689–3696 (2010).
101. R. Teki *et al.*, Nanostructured silicon anodes for lithium ion rechargeable batteries, *Small* 5, 2236–2242 (2009).
102. P. Simon, Y. Gogotsi, Materials for electrochemical capacitors, *Nat. Mater.* 7, 845–854 (2008).
103. A. S. Aricò, P. Bruce, B. Scrosati, J.-M. Tarascon, W. van Schalkwijk, Nanostructured materials for advanced energy conversion and storage devices, *Nat. Mater.* 4, 366–377 (2005).
104. C. K. Chan *et al.*, High-performance lithium battery anodes using silicon nanowires, *Nat. Nanotechnol.* 3, 31–35 (2008).
105. B. Kang, G. Ceder, Battery materials for ultrafast charging and discharging, *Nature* 458, 190–193 (2009).
106. X. W. (David) Lou, L. A. Archer, Z. Yang, Hollow micro-/nanostructures: synthesis and applications, *Adv. Mater.* 20, 3987–4019 (2008).
107. P. V. Kamat, Meeting the clean energy demand: nanostructure architectures for solar energy conversion, *J. Phys. Chem. C* 111, 2834–2860 (2007).
108. M.-L. Kuo *et al.*, Realization of a near-perfect antireflection coating for silicon solar energy utilization, *Opt. Lett.* 33, 2527–2529 (2008).
109. E. Buitrago *et al.*, The top-down fabrication of a 3D-integrated, fully CMOS-compatible FET biosensor based on vertically stacked SiNWs and FinFETs, *Sensors Actuators B Chem.* 193, 400–412 (2014).
110. W. K. Choi *et al.*, Synthesis of silicon nanowires and nanofin arrays using interference lithography and catalytic etching, *Nano Lett.* 8, 3799–3802 (2008).
111. B. Alemán *et al.*, Transfer-free batch fabrication of large-area suspended graphene membranes, *ACS Nano* 4, 4762–4768 (2010).
112. R. R. Nair *et al.*, Graphene as a transparent conductive support for studying biological molecules by transmission electron microscopy, *Appl. Phys. Lett.* 97, 153102 (2010).
113. J. R. Morones *et al.*, The bactericidal effect of silver nanoparticles, *Nanotechnology* 16, 2346–2353 (2005).
114. IOFFE, Basic properties of GaAs at 300 K, Ioffe Physico-Technical Institute (available at <http://www.ioffe.rssi.ru/SVA/NSM/Semicond/GaAs/basic.html>, accessed on October 30, 2015).

115. R. Dingle, A. C. Gossard, W. Wiegmann, Direct observation of superlattice formation in a semiconductor heterostructure, *Phys. Rev. Lett.* 34, 1327–1330 (1975).
116. N. Holonyak Jr., R. M. Kolbas, R. D. Dupuis, P. D. Dapkus, Quantum-well heterostructure lasers, *IEEE J. Quantum Electron.* 16, 170–186 (1980).
117. J. P. van der Ziel, R. Dingle, R. C. Miller, W. Wiegmann, W. A. Nordland Jr., Laser oscillation from quantum states in very thin GaAs-Al_{0.2}Ga_{0.8}As multilayer structures, *Appl. Phys. Lett.* 26, 463–465 (1975).
118. D. J. Paul, The progress towards terahertz quantum cascade lasers on silicon substrates, *Laser Photon. Rev.* 4, 610–632 (2010).
119. H. W. Li *et al.*, Quantum dot resonant tunneling diode for telecommunication wavelength single photon detection, *Appl. Phys. Lett.* 91, 073516 (2007).
120. M. Asada, Y. Miyamoto, Y. Suematsu, Gain and the threshold of three-dimensional quantum-box lasers, *IEEE J. Quantum Electron.* 22, 1915–1921 (1986).
121. Y. Arakawa, A. Yariv, Quantum well lasers—gain, spectra, dynamics, *IEEE J. Quantum Electron.* 22, 1887–1899 (1986).
122. Y. Arakawa, H. Sakaki, Multidimensional quantum well laser and temperature dependence of its threshold current, *Appl. Phys. Lett.* 40, 939 (1982).
123. T. Fujii, S. Yamakoshi, K. Nanbu, O. Wada, S. Hiyamizu, Very low threshold current GaAs-AlGaAs GRIN-SCH lasers grown by MBE for OEIC applications, *J. Vac. Sci. Technol. B* 2, 259–261 (1984).
124. W. T. Tsang, Extremely low threshold (AlGa) As modified multiquantum well heterostructure lasers grown by molecular-beam epitaxy, *Appl. Phys. Lett.* 39, 786–788 (1981).
125. D. J. Klotzkin, *Introduction to Semiconductor Lasers for Optical Communications: An Applied Approach*, Springer, New York, 1st ed., 2013.
126. R. Chin *et al.*, Temperature dependence of threshold current for quantum-well Al_xGa_{1-x}As-GaAs heterostructure laser diodes, *Appl. Phys. Lett.* 36, 19–21 (1980).
127. M. Ettenberg, C. J. Nuese, H. Kressel, The temperature dependence of threshold current for double-heterojunction lasers, *J. Appl. Phys.* 50, 2949–2950 (1979).
128. J. Faist *et al.*, Quantum cascade laser, *Science* 264, 553–555 (1994).
129. C. Gmachl, F. Capasso, D. L. Sivco, A. Y. Cho, Recent progress in quantum cascade lasers and applications, *Reports Prog. Phys.* 64, 1533–1601 (2001).
130. R. Köhler *et al.*, Terahertz semiconductor-heterostructure laser, *Nature* 417, 156–159 (2002).
131. B. S. Williams, Terahertz quantum-cascade lasers, *Nat. Photonics* 1, 517–525 (2007).
132. K. V. Klitzing, G. Dorda, M. Pepper, New method for high-accuracy determination of the fine-structure constant based on quantized Hall resistance, *Phys. Rev. Lett.* 45, 494–497 (1980).
133. D. C. Tsui, H. L. Stormer, A. C. Gossard, Two-dimensional magnetotransport in the extreme quantum limit, *Phys. Rev. Lett.* 48, 1559–1562 (1982).
134. P. J. Simmonds, H. E. Beere, D. A. Ritchie, S. N. Holmes, Growth-temperature optimization for low-carrier-density In_{0.75}Ga_{0.25}As-based high electron mobility transistors on InP, *J. Appl. Phys.* 102, 083518 (2007).

135. R. E. Prange, S. M. Girvin, Eds., *The Quantum Hall Effect*, Springer-Verlag, New York, 2nd ed., 1989.
136. H. L. Stormer, D. C. Tsui, The quantized hall effect, *Science* 220, 1241–1246 (1983).
137. P. J. Simmonds, Molecular beam epitaxy of InGaAs and InAlAs for low-dimensional electrical transport, Ph.D. thesis, University of Cambridge, 2007.
138. R. B. Laughlin, Anomalous quantum hall effect: an incompressible quantum fluid with fractionally charged excitations, *Phys. Rev. Lett.* 50, 1395–1398 (1983).
139. F. D. M. Haldane, Fractional quantization of the hall effect: a hierarchy of incompressible quantum fluid states, *Phys. Rev. Lett.* 51, 605–608 (1983).
140. J. K. Jain, Composite-Fermion approach for the fractional quantum hall effect, *Phys. Rev. Lett.* 63, 199–202 (1989).
141. J. Martin *et al.*, Localization of fractionally charged quasi-particles, *Science* 305, 980–984 (2004).
142. K. S. Novoselov *et al.*, Electric field effect in atomically thin carbon films, *Science* 306, 666 (2004).
143. K. S. Novoselov *et al.*, Two-dimensional gas of massless Dirac fermions in graphene, *Nature* 438, 197–200 (2005).
144. S. V. Morozov *et al.*, Giant intrinsic carrier mobilities in graphene and its bilayer, *Phys. Rev. Lett.* 100, 016602 (2008).
145. K. S. Novoselov *et al.*, Room-temperature quantum hall effect in graphene, *Science* 315, 1379 (2007).
146. G. Landwehr *et al.*, Quantum transport in n-type and p-type modulation-doped mercury telluride quantum wells, *Phys. E* 6, 713–717 (2000).
147. K. S. Novoselov *et al.*, Two-dimensional atomic crystals, *Proc. Natl. Acad. Sci.* 102, 10451–10453 (2005).
148. L. Song *et al.*, Large scale growth and characterization of atomic hexagonal boron nitride layers, *Nano Lett.* 10, 3209–3215 (2010).
149. S. Z. Butler *et al.*, Progress, challenges, and opportunities in two-dimensional materials beyond graphene, *ACS Nano* 7, 2898–2926 (2013).
150. P. Vogt *et al.*, Silicene: compelling experimental evidence for graphene like two-dimensional silicon, *Phys. Rev. Lett.* 108, 155501 (2012).
151. M. E. Dávila, L. Xian, S. Cahangirov, A. Rubio, G. Le Lay, Germanene: a novel two-dimensional germanium allotrope akin to graphene and silicene, *New J. Phys.* 16, 095002 (2014).
152. A. K. Geim, I. V. Grigorieva, Van der Waals heterostructures, *Nature* 499, 419–425 (2013).
153. Z. Ni *et al.*, Tunable bandgap in silicene and germanene, *Nano Lett.* 12, 113–118 (2012).
154. T. J. Thornton, M. Pepper, H. Ahmed, D. Andrews, G. J. Davies, One-dimensional conduction in the 2D electron gas of a GaAs-AlGaAs heterojunction, *Phys. Rev. Lett.* 56, 1198–1201 (1986).
155. D. A. Wharam *et al.*, One-dimensional transport and the quantisation of the ballistic resistance, *J. Phys. C Solid State Phys.* 21, L209–L214 (1988).
156. B. J. van Wees *et al.*, Quantized conductance of point contacts in a two-dimensional electron gas, *Phys. Rev. Lett.* 60, 848–850 (1988).

157. P. J. Simmonds *et al.*, Molecular beam epitaxy of high mobility In_{0.75}Ga_{0.25}As for electron spin transport applications, *J. Vac. Sci. Technol. B* 27, 2066–2070 (2009).
158. P. J. Simmonds *et al.*, Quantum transport in In_{0.75}Ga_{0.25}As quantum wires, *Appl. Phys. Lett.* 92, 152108 (2008).
159. N. K. Patel *et al.*, Properties of a ballistic quasi-one-dimensional constriction in a parallel high magnetic field, *Phys. Rev. B* 44, 10973–10975 (1991).
160. K. Thomas *et al.*, Possible spin polarization in a one-dimensional electron gas, *Phys. Rev. Lett.* 77, 135–138 (1996).
161. N. K. Patel *et al.*, Evolution of half plateaus as a function of electric field in a ballistic quasi-one-dimensional constriction, *Phys. Rev. B* 44, 13549–13555 (1991).
162. A. Kristensen *et al.*, Bias and temperature dependence of the 0.7 conductance anomaly in quantum point contacts, *Phys. Rev. B* 62, 10950–10957 (2000).
163. D. J. Reilly, Y. Zhang, L. DiCarlo, Phenomenology of the 0.7 conductance feature, *Phys. E* 34, 27–30 (2006).
164. A. C. Graham, M. Pepper, M. Y. Simmons, D. A. Ritchie, New interaction effects in quantum point contacts at high magnetic fields, *Phys. E* 34, 588–591 (2006).
165. S. M. Cronenwett *et al.*, Low-temperature fate of the 0.7 structure in a point contact: a kondo-like correlated state in an open system, *Phys. Rev. Lett.* 88, 226805 (2002).
166. S. Iijima, Helical microtubules of graphitic carbon, *Nature* 354, 56–58 (1991).
167. M. S. Dresselhaus, G. Dresselhaus, P. C. Eklund, A. M. Rao, Carbon nanotubes, in *The Physics of Fullerene-Based and Fullerene-Related Materials*, W. Andreoni, Ed. Springer, Dordrecht, the Netherlands, 2000, pp. 331–379.
168. C. Dekker, Carbon nanotubes as molecular quantum wires, *Phys. Today* May, 22–28 (1999).
169. S. J. Tans *et al.*, Individual single-wall carbon nanotubes as quantum wires, *Nature* 386, 474–478 (1997).
170. S. Frank, P. Poncharal, Z. L. Wang, W. A. de Heer, Carbon nanotube quantum resistors, *Science* 280, 1744–1746 (1998).
171. S. Iijima, Carbon nanotubes: past, present, and future, *Phys. B* 323, 1–5 (2002).
172. X. Wang *et al.*, Fabrication of ultralong and electrically uniform single-walled carbon nanotubes on clean substrates, *Nano Lett.* 9, 3137–3141 (2009).
173. X. Shi *et al.*, Fabrication of porous ultra-short single-walled carbon nanotube nanocomposite scaffolds for bone tissue engineering, *Biomaterials* 28, 4078–4090 (2007).
174. Y. Jung, X. Li, N. K. Rajan, A. D. Taylor, M. A. Reed, Record high efficiency single-walled carbon nanotube/silicon p–n junction solar cells, *Nano Lett.* 13, 95–99 (2013).
175. I. van Weperen, S. R. Plissard, E. P. A. M. Bakkers, S. M. Frolov, L. P. Kouwenhoven, Quantized conductance in an InSb nanowire, *Nano Lett.* 13, 387–391 (2013).
176. V. Mourik *et al.*, Signatures of majorana fermions in hybrid superconductor-semiconductor nanowire devices, *Science* 336, 1003–1007 (2011).
177. A. I. Yanson, G. Rubio Bollinger, H. E. van den Brom, N. Agrait, J. M. van Ruitenbeek, Formation and manipulation of a metallic wire of single gold atoms, *Nature* 395, 783–785 (1998).

178. P. J. Simmonds *et al.*, Growth by molecular beam epitaxy of self-assembled InAs quantum dots on InAlAs and InGaAs lattice-matched to InP, *J. Vac. Sci. Technol. B* 25, 1044–1048 (2007).
179. P. J. Simmonds *et al.*, Structural and optical properties of InAs/AlAsSb quantum dots with GaAs(Sb) cladding layers, *Appl. Phys. Lett.* 100, 243108 (2012).
180. O. Bierwagen, W. T. Masselink, Self-organized growth of InAs quantum wires and dots on InP(001): the role of vicinal substrates, *Appl. Phys. Lett.* 86, 113110 (2005).
181. Y. I. Mazur *et al.*, Spectroscopy of shallow InAs/InP quantum wire nanostructures, *Nanotechnology* 20, 065401 (2009).
182. F. Lelarge *et al.*, Recent advances on InAs/InP quantum dash based semiconductor lasers and optical amplifiers operating at 1.55 μm , *IEEE J. Sel. Top. Quantum Electron.* 13, 111–124 (2007).
183. R. H. Wang *et al.*, Room-temperature operation of InAs quantum-dash lasers on InP (001), *IEEE Photonics Technol. Lett.* 13, 767–769 (2001).
184. C. D. Yerino *et al.*, Tensile GaAs(111) quantum dashes with tunable luminescence below the bulk bandgap, *Appl. Phys. Lett.* 105, 071912 (2014).
185. P. J. Simmonds, M. L. Lee, Tensile strained island growth at step-edges on GaAs(110), *Appl. Phys. Lett.* 97, 153101 (2010).
186. P. J. Simmonds, M. L. Lee, Self-assembly on (111)-oriented III-V surfaces, *Appl. Phys. Lett.* 99, 123111 (2011).
187. P. J. Simmonds, M. L. Lee, Tensile-strained growth on low-index GaAs, *J. Appl. Phys.* 112, 054313 (2012).
188. A. P. Alivisatos, Semiconductor clusters, nanocrystals, and quantum dots, *Science* 271, 933–937 (1996).
189. M. V. Maximov, N. N. Ledentsov, Quantum dot lasers, in *Dekker Encyclopedia of Nanoscience and Nanotechnology*, Vol. 4, J. A. Schwarz, C. I. Contescu, K. Putyera, Eds. Marcel Dekker, Inc., New York, 2004, pp. 3109–3126.
190. A. J. Shields, Semiconductor quantum light sources, *Nat. Photonics* 1, 215–223 (2007).
191. A. Shields, Quantum logic with light, glass, and mirrors, *Science* 297, 1821 (2002).
192. D. Kim, S. G. Carter, A. Greilich, A. S. Bracker, D. Gammon, Ultrafast optical control of entanglement between two quantum-dot spins, *Nat. Phys.* 7, 223–229 (2011).
193. M. A. Reed *et al.*, Observation of discrete electronic state in a zero-dimensional semiconductor nanostructure, *Phys. Rev. Lett.* 60, 535–537 (1988).
194. T. P. Smith III, K. Y. Lee, C. M. Knoedler, J. M. Hong, D. P. Kern, Electronic spectroscopy of zero-dimensional systems, *Phys. Rev. B* 38, 2172–2176 (1988).
195. J. R. Prance *et al.*, Electronic refrigeration of a two-dimensional electron gas, *Phys. Rev. Lett.* 102, 146602 (2009).
196. F. R. Waugh *et al.*, Single-electron charging in double and triple quantum dots with tunable coupling, *Phys. Rev. Lett.* 75, 705–708 (1995).
197. L. P. Kouwenhoven *et al.*, Transport through a finite one-dimensional crystal, *Phys. Rev. Lett.* 65, 361–364 (1990).
198. D. Loss, D. P. DiVincenzo, Quantum computation with quantum dots, *Phys. Rev. A* 57, 120 (1998).

199. J. R. Petta *et al.*, Coherent manipulation of coupled electron spins in semiconductor quantum dots, *Science* 309, 2180–2184 (2005).
200. J. M. Pietryga *et al.*, Utilizing the lability of lead selenide to produce heterostructured nanocrystals with bright, stable infrared emission, *J. Am. Chem. Soc.* 130, 4879–4885 (2008).
201. C. B. Murray, D. J. Norris, M. G. Bawendi, Synthesis and characterization of nearly monodisperse CdE (E = S, Se, Te) semiconductor nanocrystallites, *J. Am. Chem. Soc.* 115, 8706–8715 (1993).
202. J. J. Li *et al.*, Large-scale synthesis of nearly monodisperse CdSe/CdS core/shell nanocrystals using air-stable reagents via successive ion layer adsorption and reaction, *J. Am. Chem. Soc.* 125, 12567–12575 (2003).
203. J. Kwak *et al.*, High-power genuine ultraviolet light-emitting diodes based on colloidal nanocrystal quantum dots, *Nano Lett.* 15, 3793–3799 (2015).
204. B. O. Dabbousi *et al.*, (CdSe)ZnS core-shell quantum dots: synthesis and characterization of a size series of highly luminescent nanocrystallites, *J. Phys. Chem. B* 101, 9463–9475 (1997).
205. C. B. Murray, C. R. Kagan, M. G. Bawendi, Self-organization of CdSe nanocrystallites into three-dimensional quantum dot superlattices, *Science* 270, 1335–1338 (1995).
206. J. Zhao *et al.*, Efficient CdSe/CdS quantum dot light-emitting diodes using a thermally polymerized hole transport layer, *Nano Lett.* 6, 463–467 (2006).
207. D. V. Talapin, J.-S. Lee, M. V. Kovalenko, E. V. Shevchenko, Prospects of colloidal nanocrystals for electronic and optoelectronic applications, *Chem. Rev.* 110, 389–458 (2010).
208. D. Leonard, M. Krishnamurthy, C. M. Reaves, S. P. Denbaars, P. M. Petroff, Direct formation of quantum-sized dots from uniform coherent islands of InGaAs on GaAs surfaces, *Appl. Phys. Lett.* 63, 3203–3205 (1993).
209. D. J. Eaglesham, M. Cerullo, Dislocation-free Stranski-Krastanow growth of Ge on Si(100), *Phys. Rev. Lett.* 64, 1943–1946 (1990).
210. D. L. Huffaker, G. Park, Z. Zou, O. B. Shchekin, D. G. Deppe, 1.3 μm room-temperature GaAs-based quantum-dot laser, *Appl. Phys. Lett.* 73, 2564 (1998).
211. C. G. Bailey, D. V. Forbes, R. P. Raffaele, S. M. Hubbard, Near 1 V open circuit voltage InAs/GaAs quantum dot solar cells, *Appl. Phys. Lett.* 98, 163105 (2011).
212. Y. Song, P. J. Simmonds, M. L. Lee, Self-assembled In_{0.5}Ga_{0.5}As quantum dots on GaP, *Appl. Phys. Lett.* 97, 223110 (2010).
213. K. Jacobi, Atomic structure of InAs quantum dots on GaAs, *Prog. Surf. Sci.* 71, 185–215 (2003).
214. B. Damilano, N. Grandjean, F. Semond, J. Massies, M. Leroux, From visible to white light emission by GaN quantum dots on Si(111) substrate, *Appl. Phys. Lett.* 75, 962 (1999).
215. D. Bimberg *et al.*, InGaAs-GaAs quantum-dot lasers, *IEEE J. Sel. Top. Quantum Electron.* 3, 196–205 (1997).
216. P. Atkinson, O. G. Schmidt, S. P. Bremner, D. A. Ritchie, Formation and ordering of epitaxial quantum dots, *Comptes Rendus Phys.* 9, 788–803 (2008).
217. Y. Nakamura *et al.*, Regular array of InGaAs quantum dots with 100-nm-periodicity formed on patterned GaAs substrates, *Phys. E* 21, 551–554 (2004).

218. S. Kiravittaya, H. Heidemeyer, O. Schmidt, Growth of three-dimensional quantum dot crystals on patterned GaAs (001) substrates, *Phys. E* 23, 253–259 (2004).
219. K. Rivoire, S. Buckley, Y. Song, M. L. Lee, J. Vučković, Photoluminescence from In_{0.5}Ga_{0.5}As/GaP quantum dots coupled to photonic crystal cavities, *Phys. Rev. B* 85, 045319 (2012).
220. P. J. Simmonds *et al.*, Tuning quantum dot luminescence below the bulk band gap using tensile strain, *ACS Nano* 7, 5017–5023 (2013).
221. C. D. Yerino *et al.*, Strain-driven growth of GaAs(111) quantum dots with low fine structure splitting, *Appl. Phys. Lett.* 105, 251901 (2014).
222. E. Stock *et al.*, Single-photon emission from InGaAs quantum dots grown on (111) GaAs, *Appl. Phys. Lett.* 96, 093112 (2010).
223. T. Kondo, K. Saitoh, Y. Yamamoto, T. Maruyama, S. Naritsuka, Fabrication of GaN dot structures on Si substrates by droplet epitaxy, *Phys. Status Solidi A* 203, 1700–1703 (2006).
224. B.-L. Liang *et al.*, Energy transfer within ultralow density twin InAs quantum dots grown by droplet epitaxy, *ACS Nano* 2, 2219–2224 (2008).

CHEMICAL SENSING

W. RUDOLF SEITZ

Department of Chemistry, University of New Hampshire, Durham, NH, USA

73.1 INTRODUCTION

Chemical sensing involves measuring the concentration of an analyte in a sample. It has the following two essential features: (i) the measurement is made in the sample or at the site of the sample rather than in a separate laboratory and (ii) the concentration of interest is measured either continuously or repeatedly such that changes in concentration with time can be followed. This chapter will review analytical measurement technologies that meet these requirements with emphasis on those methods that are not covered elsewhere in this book.

Most chemical sensors are used in one of two contexts. They can be part of a system for some type of process control. For example, a pH electrode can be used to sense the pH of a fermentation process. If the pH starts to deviate from the desired value, it sends a signal that causes acid or base to be dispensed into the fermentation to restore the pH to the desired value. Another example involves glucose measurements by diabetics, which are used to determine how much insulin should be dispensed. The ideal system for this would involve an implantable glucose sensor coupled to an insulin pump with the appropriate control algorithm. In practice, it hasn't yet proved feasible to develop a glucose sensor that functions long enough to warrant implantation. Instead, glucose is measured intermittently and the resulting value is used to establish the need for insulin.

The other context where sensors are frequently used involves alarm monitoring. The sensor is designed to produce a signal if a concentration of some analyte becomes high enough to cause a problem. Carbon monoxide sensors in the home fall into this

category. Environmental sensing is often undertaken to make sure concentrations don't exceed a predetermined value where the analyte may prove problematic.

The ideal sensor would be perfectly selective for the analyte of interest and would respond continuously without maintenance. In practice, these goals are difficult to realize. Selectivity is often difficult to achieve or to combine with continuous response. Biosensors, described in another chapter, have been developed because biological reagents like enzymes and antibodies offer excellent selectivity. However, few biosensors offer continuous response and those that do have a limited lifetime because biological reagents are not indefinitely stable. The development of stable receptors that offer selectivity approaching that of biological reagents with improved stability is an active area of research that is likely to lead to improved sensors in the future.

Sensor arrays offer an alternative approach to selectivity. The idea is to use multiple sensors that have different selectivity patterns. While none is completely selective for a particular analyte, pattern recognition techniques can be applied to the responses from the array to distinguish different analytes. Gas sensor arrays have become known as "electronic noses" and have proven themselves to be well suited for certain types of problems. Sensor arrays designed for measurements in solution are known as "electronic tongues." These remain a subject of research and have yet to have the same practical impact as electronic noses.

Longevity and calibration are issues for any sensor. Reagent stability can limit sensor lifetime. For example, this is an issue with continuous biosensors, for example, the glucose electrode based on the enzyme glucose oxidase. Another problem involves sensor fouling, that is, accumulation of unwanted material on the surface of a sensor that affects its response. When applying a chemical sensor to a new type of sample, a series of control experiments needs to be undertaken to establish the length of time that a sensor will provide accurate response. Based on these experiments, a maintenance protocol that involves cleaning and recalibration can be established to assure that the sensor is responding with the desired accuracy at all times. Most physical sensors need little if anything in the way of maintenance. However, this is not the case for many chemical sensing technologies.

Ideally chemical sensors respond continuously to changes in analyte concentration. Measurement technologies such as in situ Raman spectroscopy that directly measure analyte concentration intrinsically provide continuous response. However, most chemical sensors involve an interaction between that analyte and the sensor that modifies the properties of a sensing element. If this interaction is reversible, then response is continuous. Other "continuous" sensors involve steady-state measurements that involve analyte consumption at such a slow rate that it does not significantly change the analyte concentration.

As we go through various sensor technologies, we will consider the issues that affect that particular technology. Other sources of information will be referenced. The goal will be to provide a sense of the state of the art for each sensing technology plus an indication of the main thrusts of current research and how they may be expected lead to improved sensor technology in the future.

The field of chemical sensing encompasses a variety of analytical technologies, many of them quite different. Electrical engineers, chemists, and physicists are all active in chemical sensor research. There is an excellent book on chemical sensing that considers general sensing issues more in depth than this chapter and covers many of the same measurement technologies more in depth than this chapter [1]. There is also a recent comprehensive book that covers both chemical and biological sensors [2]. However, for the most part, it's easier to find more detailed information on a particular type of sensor than it is to find general coverage of all chemical sensing approaches.

The chemical sensing technologies that we will consider in this chapter are categorized as electrical, optical, and mass, depending on the type of signal that is measured. Several of the sensing technologies covered can be coupled to biological recognition agents to make biosensors.

73.2 ELECTRICAL METHODS

73.2.1 Potentiometry

73.2.1.1 Solution Phase Measurements Potentiometry is an established technique for measuring the activity of ions in solution. The measurement of pH with a glass electrode is by far the most widely applied and best known example of potentiometry. However, electrodes are available for sensing the activities of a wide variety of other cations and anions.

The basic arrangement for potentiometry is shown in Figure 73.1. What is actually measured is the difference in potential between two electrodes. The reference electrode is designed so that its potential is independent of solution composition. The saturated calomel electrode is most commonly used as a reference. The other electrode is known as the indicator electrode. Its potential varies with the activity of the ion of interest.

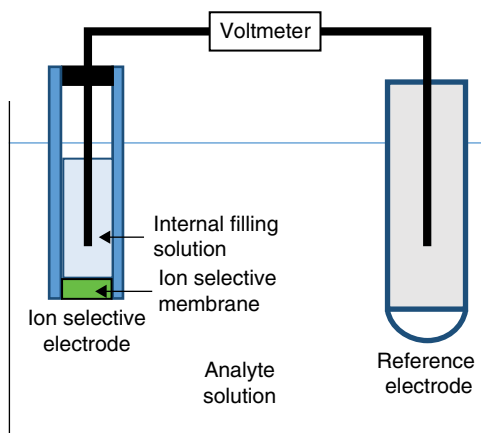


FIGURE 73.1 Instrumental arrangement for potentiometry.

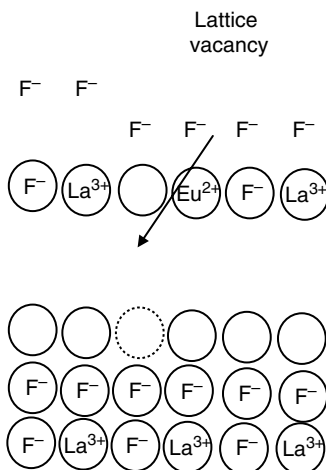


FIGURE 73.2 Lanthanum fluoride membrane used in ion selective electrode that responds selectively to fluoride. The membrane is doped with a +2 ion Eu(II) creating lattice vacancies because Eu(II) only needs to be surrounded by 2 rather than 3 fluoride ions. Fluoride ions can move into the vacancy and thus can be transported through the membrane. Transport is selective because other ions don't fit into the lanthanum fluoride crystal lattice.

In the case of pH measurements, the two electrodes are often combined onto a single housing known as a combination electrode.

Most indicator electrodes are membrane electrodes. The potential that is measured arises because there is a difference in the concentration of the ion of interest on either side of a membrane. To make an electrode that is selective for a particular ion, you need a membrane that is selectively permeable to that ion. We illustrate this with the ion selective electrode for fluoride, the second most widely applied ion selective electrode after the glass pH electrode. The membrane for this electrode, lanthanum fluoride doped with Eu(II), is shown in Figure 73.2. The presence of a +2 ion in LaF_3 leaves a vacancy in the crystal lattice. This results in a solid-state material that is permeable to fluoride because fluoride can hop from site to site. Selective fluoride transport arises because other anions are too large to conveniently fit into the lanthanum fluoride lattice. When the fluoride ion concentration in the internal filling solution of the ion selective electrode is higher than the fluoride concentration in the external solution, there is net transport of fluoride from the internal filling solution. The transport rate is so slow that this does not produce a large change in fluoride concentration in the external solution. However, it does result in an electrode potential that depends on the ratio of fluoride concentrations on either side of the lanthanum fluoride membrane.

Lanthanum fluoride is an example of a solid-state membrane. Another common type of ion selective membrane involves a compound known as an ionophore that selectively binds a particular ion dissolved in plasticized polyvinyl chloride.

The equation for ion selective electrode response is shown as follows:

$$E = k + \beta \frac{0.0592}{z} \log([A] + K_{AI} [I])$$

E is the measured electrochemical cell potential in volts. “ k ” and “ β ” are constants that must be determined by calibration. “ z ” is the charge on the ion that is being measured. For example, if a sulfide (S^{2-}) ion selective electrode is being used, the value of “ z ” is -2 . $[A]$ is the concentration of the analyte ion. $[I]$ is the concentration of an interfering ion. K_{AI} is the selectivity coefficient for that particular interfering ion. The smaller the value of the selectivity coefficient, the more selective the electrode for analyte A compared to interference I . When the product $K_{AI} [I]$ is significant compared to $[A]$, the interfering ion starts to contribute to the measured potential.

The two constants, k and β , are determined by measuring electrode potential at two or more analyte concentrations in the absence of interfering ions. The theoretical value of β is 1.00 at 25°C. In practice it is usually slightly smaller. The value of β is determined from the slope of the response. Once β is known, the value of k can be calculated.

When you acquire an ion selective electrode, it will come with a series of selectivity coefficients. These can be used to determine whether a particular ion can be sensed without interference in a given context.

One important characteristic of ion selective electrodes is that they measure ion activities rather than total concentrations. Activity is a physical chemistry concept that may not be familiar to all readers. Analyte activity equals the analyte concentration times an activity coefficient. The value of the activity coefficient is 1.00 at infinite dilution but decreases as the ionic strength of the solution increases. To measure concentration, the usual practice is to adjust the ionic strength of the standards used to calibrate the electrode so that it equals the ionic strength of the sample. In this case, the activity coefficient is the same in both standards and sample, and the measured parameter is the concentration.

Another way of viewing this is that ion selective electrodes only measure the concentration of a particular ion. Thus, a Cu(II) electrode only measures the concentration of Cu(II) in solution. It does not respond to total Cu. Cu in the form of a complex ion, for example, $Cu(NH_3)_4^{2+}$, will not be sensed by an electrode. The glass pH electrode only measures the concentration of H^+ ion. It does not respond to total acid.

The instrumentation for potentiometry is nothing more than a voltmeter with a high input impedance. The ubiquitous pH meter is nothing more than such a voltmeter designed to directly read out in pH if it is appropriately calibrated using standard pH buffers.

Electrodes are available that respond selectively to a large number of ions including F^- , Br^- , Cl^- , I^- , CN^- , SCN^- , Ag^+ , Cu^{2+} , Cd^{2+} , Pb^{2+} , Na^+ , Li^+ , K^+ , NH_4^+ , and S^{2-} .

Potentiometry is also widely used to measure the concentrations of acidic or basic dissolved gases, particularly carbon dioxide and ammonia. To make a carbon

dioxide-sensitive electrode, a pH electrode is covered by a CO_2 permeable membrane. A thin layer of a bicarbonate solution is contained between the membrane and the pH sensor. The dissolved CO_2 diffuses through the gas permeable membrane into the bicarbonate solution. The CO_2 reacts with water-forming carbonic acid, H_2CO_3 , which changes the pH of the solution. The measured potential due to pH changes in a systematic known manner with changing carbon dioxide partial pressures. Ammonia is sensed in a similar manner using an ammonia permeable membrane and a thin layer of an ammonium ion solution over a pH electrode.

Potentiometry is a mature technology. While work continues on new membrane formulations that may offer improved selectivity for certain applications, only incremental improvements can be expected. More in-depth information on potentiometry and its applications may be found in common analytical chemistry texts such as *Quantitative Chemical Analysis* by Daniel Harris.

73.2.1.2 Gas Phase Measurements There are also potentiometric gas sensors. The basic configuration is exemplified by the widely applied oxygen sensor based on zirconia shown in Figure 73.3. As with potentiometric ion selective electrodes, the measured parameter is a potential across a membrane resulting from a concentration gradient. One side contains a known reference concentration of the analyte, and the other side contains the sample of interest. In the zirconia sensor the reference side is air, which contains a constant percentage of oxygen.

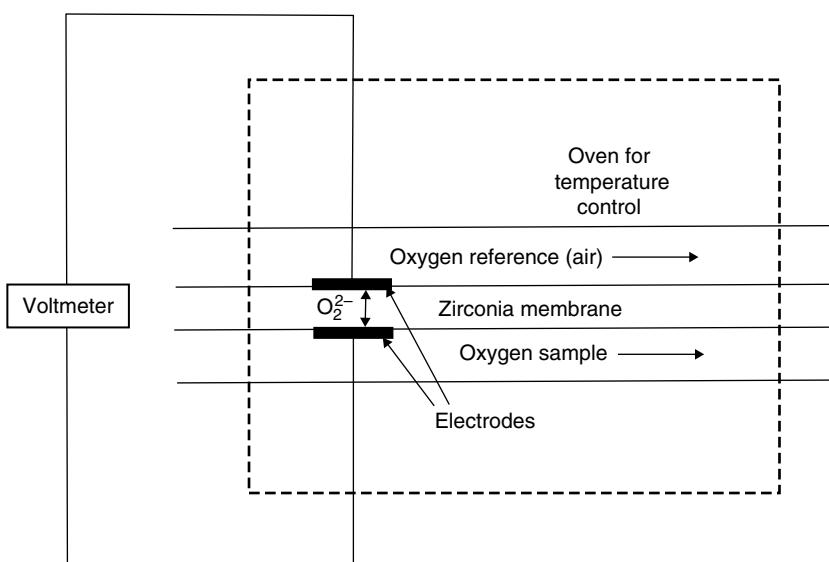


FIGURE 73.3 Schematic of high-temperature zirconia oxygen sensor. The reference gas is normally air. Oxygen is transported across the membrane as O^{2-} . The potential across the membrane is measured directly using electrodes applied to either side of the membrane.

The potential across the zirconia membrane is measured directly using two platinum electrodes applied directly to the membrane surface. While potentials develop at the interfaces between the platinum and the zirconia, the values of these two potentials are equal in magnitude and opposite in sign, so they cancel out leaving only the potential due to the difference in analyte concentration on either side of the metal oxide. This potential is due to selective oxygen transport across the zirconia. The oxygen reacts with zirconia to form O^{2-} ions that can hop from site to site. The transport mechanism is analogous to that observed with lanthanum fluoride membranes used for aqueous phase fluoride sensing. As with the other type of membrane electrode, the potential across the electrode varies with the logarithm of the ratio of oxygen concentrations on either side of the membrane.

The zirconia-based oxygen sensor is widely used to sense the oxygen concentrations in fuel–air mixtures in internal combustion engines. They are incorporated into feedback control systems to maintain oxygen concentrations at the level required for efficient combustion reactions. Current research on this type of oxygen sensor has been recently reviewed [3].

This approach is applicable to other oxidizing gases such as chlorine.

73.2.2 Voltammetry

Voltammetry refers to methods where the potential of a working electrode is controlled relative to a reference electrode, and the resulting current is measured. This current is due to oxidation or reduction of a solute. The magnitude of the current is related to solute concentration. Voltammetry can be implemented directly in aqueous samples that and, therefore, can be used for continuous sensing. In practice, such applications are few. The one requirement is that the aqueous solution contains enough dissolved salt to be electrically conducting.

In practice voltammetry is not widely used for continuous chemical sensing. One problem is that most of the complex samples that we deal with in practical contexts contain components that deposit on the electrode surface. This blocks the electrode surface, reducing its effective area. Since current is proportional to electrode area, this causes an interference. The result is the voltammetry requires frequent electrode maintenance to the point where it is usually deemed impractical. A second problem is that it is often difficult to find a potential at which the analyte of interest can be oxidized or reduced without also oxidizing or reducing other solutes as well.

There are two strategies that have led to successful voltammetric sensing. One is to determine gases that can be oxidized or reduced using a membrane to protect the electrodes. By far the most important application involves the measurement of dissolved oxygen. The voltammetric oxygen electrode is shown in Figure 73.4. The voltammetric electrode is separated from the analytical sample by a membrane that is permeable to gases but not to solution. This eliminates interferences from all nonvolatile solutes. The most common membrane is Teflon, a substance that does not adhere strongly to

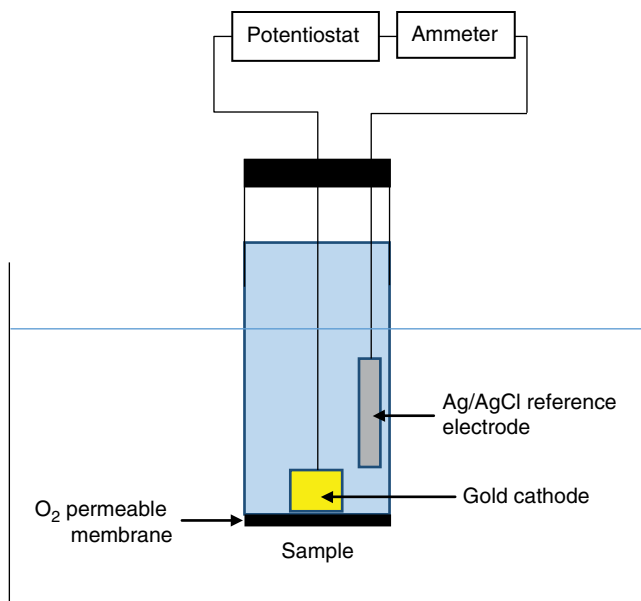


FIGURE 73.4 Schematic of voltammetric oxygen electrode. The measured parameter is the current due to the oxygen reduction at the gold cathode. The potential of this electrode is controlled by the potentiostat at a value sufficiently cathodic relative to the silver/silver chloride reference electrode to reduce oxygen.

anything. As a result, the dissolved oxygen electrode is not subject to fouling by solutes in the sample of interest. The dissolved oxygen electrode is sometimes known as a Clark electrode, after its inventor, Leland Clark.

There is a thin layer of a salt solution between the membrane and the electrode. This is required so that the potential of the working electrode can be controlled relative to the potential of a reference electrode. This voltage is held at a sufficiently negative value to reduce dissolved oxygen to water, $\text{O}_2(\text{g}) + 4\text{H}^+(\text{aq}) + 4\text{e}^- \rightarrow 2\text{H}_2\text{O}(\text{l})$. The resulting current depends on the concentration of dissolved oxygen.

The oxygen electrode is operated in a steady-state mode. The electrode potential is held a sufficiently negative value to reduce all oxygen at the surface of the electrode. This creates a concentration gradient. Oxygen from the sample diffuses across the membrane and the thin solution layer to reach the working electrode surface where it is reduced. If the sample is stirred, the sample will be homogeneous, and the oxygen concentration at the membrane surface will stay essentially constant. While the oxygen that is reduced at the working electrode surface reduces the solution concentration, this amount of oxygen that is reduced is a tiny percentage of the total oxygen amount so that concentration changes are too small to have a significant effect on oxygen concentration.

This approach is applicable to gases that can be oxidized or reduced. Examples include H_2 , CO , NO_2 , NO , H_2S , and SO_2 . More information on voltammetric gas sensing is available from an excellent review article [4].

A second strategy for voltammetric sensing involves disposable electrodes prepared by screen printing. The most important example involves glucose sensing using a strip consisting of disposable electrodes covered by a layer of the immobilized enzyme glucose oxidase that converts glucose to a product that can be determined by voltammetry. This is a biosensor, covered in another chapter of this handbook.

73.2.3 Chemiresistors

The resistance of many substances depends on their environment. Changes in resistance can be used to sense this environment. Figure 73.5 shows two ways of configuring chemiresistors.

The first widely applied chemiresistor was developed by Taguchi for propane sensing. The sensing element is SnO_2 . Propane adsorbing on the surface of the tin oxide donates electrons to the tin oxide, reducing its electrical resistance. This particular device is widely applied as an alarm sensor to detect gas leaks.

Metal oxides are one type of chemiresistive material. Response involves adsorption of gases on a surface. The change in resistance per mole of adsorbed gas will depend on the surface to volume ratio of the resistive material. Hence, to maximize sensitivity, it is desirable to use porous materials with high surface to volume ratios. By providing routes to the preparation of highly porous materials, nanotechnology is leading to improvements in the sensitivity of gas sensing chemiresistors.

The mechanism of metal oxide chemiresistor response involves changes in the number of charge carriers in a semiconductor due to adsorption of a gas at the semiconductor surface. Oxidizing gases like oxygen tend to attract electrons. Adsorption of oxygen on the surface of an n-type semiconductor attracts the negatively charged electrons, leading to an increase in electrical resistance. Adsorption of a reducing gas will have the opposite effect, increasing the number of charge carriers. If the resistive element is a p-type semiconductor, the opposite effects will be observed. By attracting

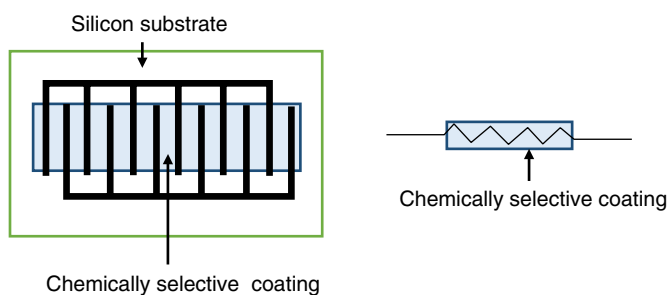


FIGURE 73.5 Two configurations for chemiresistors. In the configuration on the left, the chemically sensitive layer is deposited over interdigitated electrodes on the surface of a silicon semiconductor substrate. On the right, a chemically sensitive coating is coated directly onto a resistor.

electrons, adsorption of an oxidizing gas will increase the number of holes available to carry charge causing a decrease in resistance.

Organic conductors such as polyaniline also function as chemiresistors. In this case, gases partition into the organic material causing a change in resistance. Equilibration times depend on the thickness of the organic material.

Chemiresistors are true sensors in that they involve an equilibrium between the concentration of a gas and number of adsorbed gas molecules or atoms on the surface of the resistive elements. Temperature will affect this equilibrium and needs to be controlled. Selectivity depends on both the affinities of the surface for different gases and the abilities of different gases to affect resistance.

A different way of preparing a chemically sensitive resistor is to deposit small particles of a conducting material such as graphite into a polymer. When volatile organic compounds interact with the polymer, they induce swelling, which increases the distance between conducting particles, causing resistance to increase. In this case, selectivity depends on the coefficient for gas partitioning into the polymer.

Further information on chemiresistors may be obtained from a book on this subject [5]. The properties of metal oxides that are sensitive to gas concentrations are also described in a separate book [6].

The selectivities of chemiresistors are limited. However, their cost is very low. To increase selectivity, arrays of chemiresistors with different response characteristics have been constructed. As discussed later in the chapter, the combination of sensor arrays with the appropriate mathematics has proven to be a useful method for identifying complex mixtures of gases. Such devices are commonly known as electronic noses. A recent review article summarizes the state of the art with respect to chemiresistors in electronic noses [7].

73.2.4 Field Effect Transistors

Field effect transistors can be used to sense either gases or ions in solution. The basic arrangement is shown in Figure 73.6. In a field effect transistor, charge is applied to a gate that is separated from doped semiconductor by a thin layer of silicon dioxide. A positive charge applied to the gate attracts negative charge carriers to the surface of p-type semiconductor material creating a channel where current can flow from the source to the drain. The magnitude of the current flow depends on the size of the charge. If the charge is related to a chemical concentration, then the field effect transistor will function as a chemical sensor.

One class of applications involves ion sensing. The gate is coated with a membrane that selectively binds the ion of interest. This produces an ion selective field effect transistor (ISFET). For stable response, a reference electrode is required. It's also necessary to encapsulate the device in some sort of polymer that protects all components of the field effect transistor except the membrane-coated gate from exposure to the external aqueous samples.

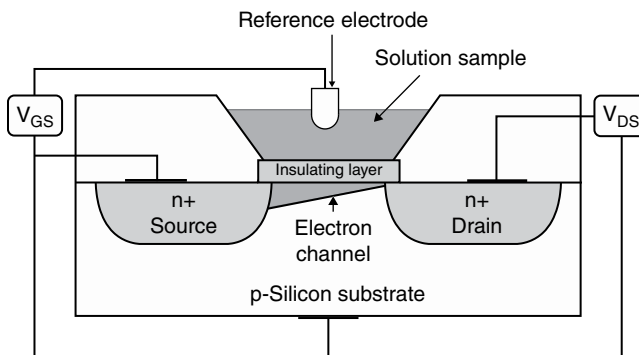


FIGURE 73.6 A solution phase chemically sensitive field effect transistor. The buildup of positive charge on the top surface of the insulating layer attracts electrons to the bottom layer. This affects the magnitude of the current between the source and the drain.

The membranes used for ISFETs involve the same polymers and ionophores that are used for ion selective electrode membranes. ISFETs are alternatives to ion selective electrodes with similar selectivities. Given that ISFETs are amenable to mass production with extremely low cost per device, it is expected that they will eventually replace ion selective electrodes. However, at present, this has not happened, and most potentiometric measurements are carried out with conventional electrodes.

Gas sensing is also possible using field effect transistors [8]. In this case, gas adsorption on the surface of the insulating layer affects the number of charge carriers in the channel that conducts current from the source to the drain.

73.3 OPTICAL METHODS

73.3.1 In situ Optical Measurements

Optical fibers consist of a high refractive index core surrounded by a lower refractive index cladding so that total internal reflection occurs when light strikes the core cladding interface at an angle greater than the critical angle. They provide a convenient method for transporting light to and from a sample. Both plastic and glass fibers are available, transmitting light in the ultraviolet, visible, and near-infrared regions of the electromagnetic spectrum. Both transmission and emission measurements are readily realized. One common configuration is shown in Figure 73.7. A central fiber carries light from a source into a sample. It is combined with six emission fibers that surround the excitation fiber and carry light back to a detection system. If a reflector is placed at the common end of the fiber bundle, this arrangement serves to measure sample absorbance. Without a reflector this arrangement measures scattered or emitted light either at the same wavelength or at a different wavelength.

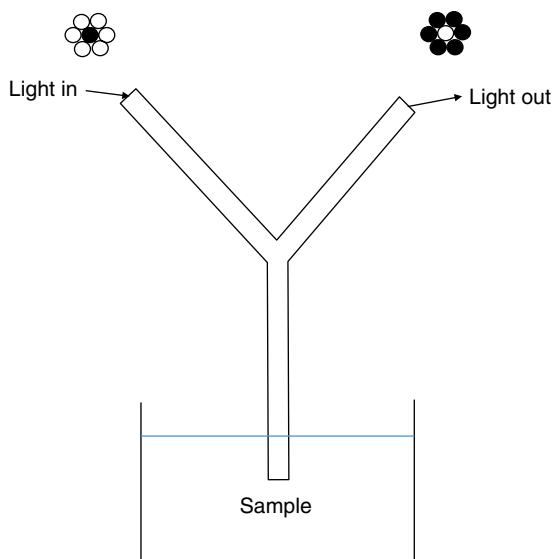


FIGURE 73.7 Common fiber optic arrangement for in situ spectroscopy. A single central fiber conducts light to a sample. A set of six fibers around the central fiber collect light from the sample and transport it back to a detection system.

Another common approach is to remove the cladding from an optical fiber and to directly expose the core to a sample. If the sample refractive index is lower than the fiber core refractive index, total internal reflection occurs at the core/sample interface. However, some of the light enters into the sample where it can be absorbed. This is an alternate method for measuring sample absorbance, essentially a form of internal reflection spectroscopy. If the absorbed light excites fluorescence, this can be captured within the fiber and transmitted to a detection system.

Several types of spectroscopy can be carried out within liquid samples. Absorption measurements in the ultraviolet and visible are readily performed using the fiber optic arrangement of Figure 73.7 with a reflector. Fiber optic attachments are readily available to perform this type of measurement.

Near-infrared spectroscopy is very conveniently performed through optical fibers. This form of spectroscopy is described in another chapter of this handbook and will not be considered further here.

Fluorescence and Raman spectra are also readily measured in situ. Fluorescence is limited to analytes that fluoresce. Raman spectroscopy is more generally applicable to a variety of sensing applications and will be considered further in the next section of this chapter.

A recent review provides more detailed descriptions of the various types of spectroscopy that can be implemented through optical fibers [9].

73.3.2 Raman Spectroscopy

Scattering refers to light that changes direction when it interacts with matter. Rayleigh scattering refers to light that changes direction without changing wavelength. Raman scattering refers to light that changes wavelength as well as changing direction. As shown in Figure 73.8, scattering may be viewed as resulting from an interaction between matter and a photon of light that promotes matter to a higher-energy “virtual state.” This state does not involve a specific energy level of the excited matter and is extremely short lived. Normally, the virtual state returns to the original state, reemitting a photon of light with the same energy. However, it is observed that a small percentage of virtual states return to a different vibrational energy level, resulting in a shift of photon energy. Most often the initial state is a ground energy state, and the final state is a higher-energy vibrational state. This is known as Stokes scattering. The resulting Raman emission is at a longer wavelength than the incident wavelength. The difference in photon energies corresponds to the energy of a vibrational transition. Anti-Stokes scattering involves photons that originate in a higher-energy vibrational state and return to a ground state. Because the vast majority of molecules are initially in their ground vibrational state at room temperature, the Stokes lines are much more intense than the anti-Stokes line. Both are much weaker than Rayleigh scattering. The Raman signal is approximately 1 million times weaker than the Rayleigh signal.

The instrumental challenge of Raman spectroscopy is to measure weak emission at wavelengths close to the much stronger emission due to Rayleigh scattering. For many years, this required expensive instrumentation limiting practical applications. Recently, however, advances in optical technology have greatly reduced the cost of Raman spectroscopy. These include semiconductor diode lasers as high-intensity monochromatic sources that produce stronger Raman signals, notch filters for selectively removing the high-intensity Rayleigh line that produces less background, and charge coupled device

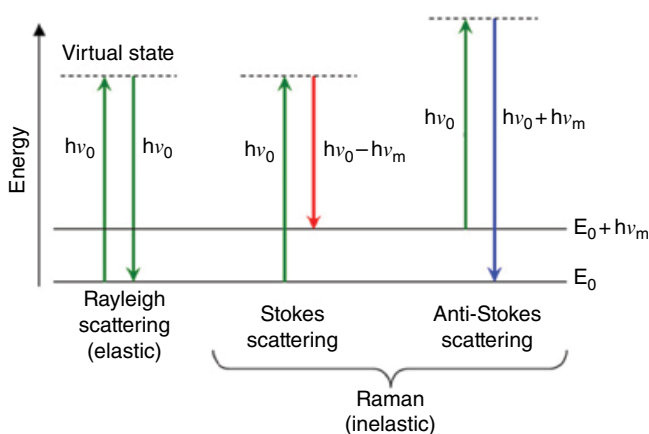


FIGURE 73.8 Processes involved in Raman scattering.

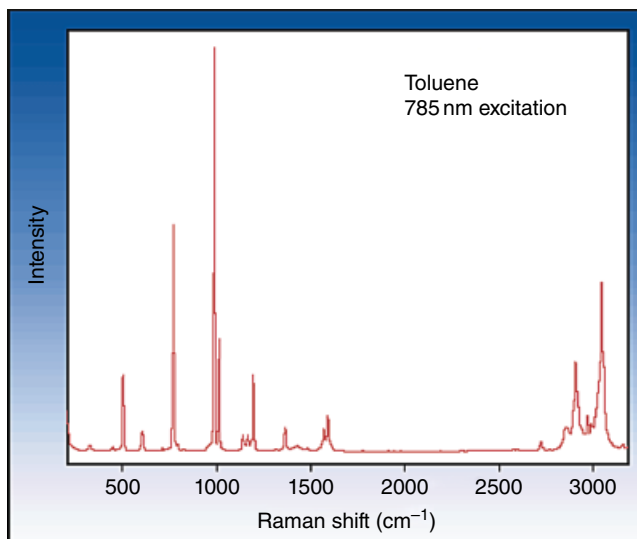


FIGURE 73.9 Raman spectrum of toluene.

detector arrays for simultaneously acquiring the complete Raman spectrum with high sensitivity.

When excited in the ultraviolet or visible region of the electromagnetic spectrum, many samples emit fluorescence. Fluorescence emission occurs at longer wavelengths than the excitation source and will overlap Raman spectra, often completely obscuring it. The probability of fluorescence can be greatly reduced by exciting at a sufficiently long wavelength in the visible or near-infrared region of the electromagnetic spectrum.

Figure 73.9 shows the Raman spectrum of toluene obtained using 785 nm semiconductor laser as the excitation source. The x -axis is plotted as the shift in energy compared to the Rayleigh band. This way Raman spectra acquired using excitation at different wavelengths will appear the same. The actual wavelengths depend on the wavelength of the excitation source and the magnitude of the Raman shift.

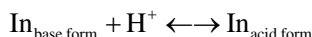
Raman spectra include bands due to both stretching and bending vibrations. The wavelength of the stretching vibrations depends on the particular type of bond involved in the stretch and may be used to identify functional groups present in the sample. The bending vibrations are sensitive to overall structure. When chemicals react, bonds in the reactants break forming new bonds in the products. Thus, all chemical reactions will give rise to changing spectra and can be followed by Raman spectroscopy. It is important to note that Raman spectra will include not just the sample of interest but also a large signal from the solvent used for a reaction, which may obscure otherwise useful spectral regions.

Raman spectroscopy is already establishing itself as a useful method for process control in the chemical and pharmaceutical industries, a trend that is likely to continue [10].

The importance of Raman spectroscopy for chemical sensing is likely to increase in the near future since several low-cost instruments are now available for this type of application.

73.3.3 Indicator-Based Optical Sensors

Indicators are reagents that change optical properties upon reversibly interacting with an analyte. The best known indicators are those that respond to pH. The reaction is shown in the following:



The range of pH that can be sensed by a particular indicator is given by the Henderson–Hasselbalch equation:

$$\text{pH} = \text{p}K_{\text{a}} + \log \frac{[\text{In}_{\text{base form}}]}{[\text{In}_{\text{acid form}}]}$$

The $\text{p}K_{\text{a}}$ is the negative log of the acid ionization constant, K_{a} , for the acid form of the indicator. The most common indicators are different colors in the acid and base form. The most familiar application context involves locating titration end points by observing the exact volume required to change indicator color.

Acid–base indicators can also be used to sense pH. The ratio of the concentration of the acid form of the indicator to the base form of the indicator can be determined instrumentally. One way of accomplishing this is to chemically bind the indicator to a solid substrate, which is then confined to the end of a fiber optic bundle similar to that shown in Figure 73.7. The ratio of the reflected intensity at a wavelength absorbed by the acid form of the indicator to the reflected intensity at a wavelength absorbed by the base form of the indicator will change with pH over the range $\text{pH} = \text{p}K_{\text{a}} \pm 1$. Outside this range of pH, the indicator will be essentially completely in either the acid or base form.

For continuous sensing purposes it's more advantageous to use indicators that fluoresce in both the acid and base forms. Because fluorescence measurements are highly sensitive, this makes it possible to work with very low indicator amounts. By relating pH to the ratio of two intensities, it's possible to make pH measurements in environments such as within a cell where calibration of a single intensity measurement would be difficult if not impossible.

An important requirement for sensing with an indicator is that the indicator concentration be low enough so that it does not significantly perturb the concentration of analyte. In the case of an acid–base indicator, this means that the number of protons that come from the indicator itself does not significantly perturb the pH of the solution.

The same concept can be applied to metal ion detection. In this case the indicator is a ligand that changes optical properties upon complexation with a metal ion. For example, there are commercially available Zn(II) indicators that shift fluorescence

emission wavelength upon complexation. Just as the response range of a pH indicator depends on indicator K_a , the response range of a metal ion indicator depends on formation constant for metal ion binding.

Like potentiometric electrodes, indicators measure metal ion activity, that is, the concentration of metal ion the form of the free metal ion, rather than measuring total metal. In environmental and biological samples, the metal ion activity rather than total metal concentration determines the effects of a given metal ions.

Metal ion indicators have been successfully developed for specialized applications. Ratiometric fluorescent indicators for Ca(II) and Zn(II) have revolutionized the study of the biology of these metal ions. Research to develop useful indicators for other metal ions is ongoing.

Another important type of optical indicator involves fluorescence quenching. One important example involves oxygen, a common quencher. The indicator in this case is an efficient fluorophore that is subject to oxygen quenching. The decrease in fluorescence intensity can be related to oxygen partial pressure. An even better approach is to measure the change in fluorescence lifetime that results from quenching. The advantage of lifetime measurements is that they do not depend on the amount of fluorophore and thus are not affected by slow fluorophore photodegradation. Nitro compounds, particularly explosives like 2,4,6-trinitrotoluene, are efficient fluorescent quenchers and may be sensed by low levels by fluorescent indicators.

A recent review on optical sensors includes work on indicator-based optical sensors [9].

73.4 MASS SENSORS

There are two widely used types of mass sensor, the quartz crystal microbalance and the surface acoustic wave device. Both are based on the piezoelectric effect in quartz. Application of an electric potential causes the quartz to deform slightly. Application of a varying electrical potential creates an acoustic wave in the quartz. In the case of the quartz crystal microbalance, electrodes are attached to opposite sides of a quartz crystal. An oscillating potential leads to a shear wave that propagates through the crystal as shown in Figure 73.10. In the case of the surface acoustic wave device, the wave propagates along the surface of the quartz as shown in Figure 73.11.

These devices are incorporated into electrical circuits that allow them to oscillate at a resonant frequency. This frequency depends on the mass of the quartz. Changes in mass cause a shift in the resonant frequency that is proportional to the change in mass. The proportionality constant will depend on the viscoelastic properties of the coating that is changing mass.

These waves propagate freely in air but are damped by liquids. The damping is more severe for surface waves. The surface acoustic wave is more sensitive to mass loading and is able to detect mass changes in the small picogram range. However, because it is more subject to damping by liquids, it is mainly used for gas sensing. The quartz

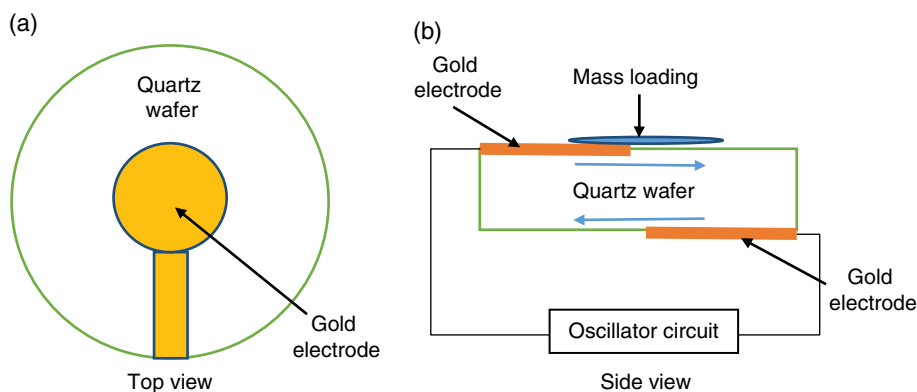


FIGURE 73.10 Schematic of quartz crystal microbalance. Electrical energy propagates as a shear wave through the quartz wafer. The oscillation circuit assumes the resonant frequency, which shifts in a linear fashion with mass loading. The arrows in the quartz wafer represent the mass displacement accompanying shear wave propagation through the quartz.

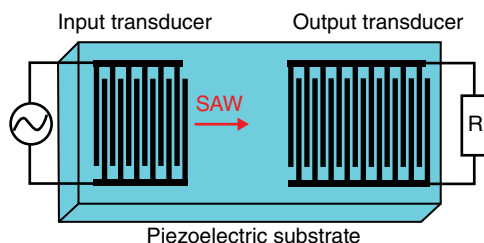


FIGURE 73.11 Arrangement for producing surface acoustic waves.

crystal balance is more easily operated in liquids with the ability to detect mass changes in the low nanogram range. It is more widely used for solution phase measurements.

Mass sensors can be coupled with recognition phases to make chemical sensors. What is detected is the shift in frequency accompanying analyte binding. One type of application involves gas sensors based on thin polymer coatings on surface acoustic wave devices. Sensitivity depends on the coefficient for gas partitioning into the polymer. Selectivity is low, depending on the relative affinity of the polymer for different gases. However, these sensors are good candidates for sensor arrays using several SAW devices each coated with a different polymer with a different selectivity pattern. This type of array is known as an electronic nose and will be covered in the last part of this chapter.

Quartz crystal microbalances are well suited for solution phase measurements. They are best suited for detecting large analytes such as proteins because of the larger mass change per analyte molecule. This type of device generally employs a biological recognition agent such as an antibody and is, therefore, classified as a biosensor.

An excellent book provides further details on acoustic sensors [11]. Recent activity in this area is summarized in a recent review article [12].

73.5 SENSOR ARRAYS (ELECTRONIC NOSE)

Many of the sensor technologies covered in this chapter are insufficiently selective to be useful for monitoring a specific analyte in a complex sample. However, they are sufficiently inexpensive that it is practical to prepare an array of sensors each of which responds with a somewhat different selectivity. This approach has proven to be practical for gas analysis. The chemically selective layers are polymers. They can be coated onto several of the base sensors described in this chapter including chemiresistors, surface acoustic wave devices, and field effect transistors. These sensor arrays are widely known as electronic noses because their response is similar to olfaction. Several electronic noses are available commercially.

The same concept can be applied to solution measurements. This type of array has been called an electronic tongue. However, while research is ongoing, it has yet to prove useful for practical problems and is unlikely to do so soon.

Electronic noses do not directly measure concentrations. Instead they are trained to recognize a particular type of sample. Determining the freshness of fish is an example of the kind of problem that can be addressed using an electronic nose. It would be initially exposed to a series a fish of known freshness. Then it would be used to assess the freshness of unknown fish using pattern recognition techniques to compare the vapor signature of the unknown fish to the signal from the known fish. Electronic noses are widely applicable to a variety of vapor detection problems [13–15].

REFERENCES

1. Janata, J.; *Principles of Chemical Sensors*, Second edition, Springer Verlag, 2009.
2. Banica, F.-G.; *Chemical Sensors and Biosensors: Fundamentals and Applications*, John Wiley & Sons Inc., 2012.
3. Moos, R.; Izu, N.; Rettig, F.; Reiss, S.; Shin, W.; Matsubara, I.; Resistive Oxygen Sensors for Harsh Environments, *Sensors* 11 (2011) 3439–3065.
4. Stetter, J.R.; Li, J.; Amperometric Gas Sensing – A Review, *Chem. Rev.* 108 (2008) 352–366.
5. Aswal, D.K.; Gupta, S.K.; *Science and Technology of Chemiresistor Gas Sensors*, Nova Publishers, 2007.
6. Eranna, G.; *Metal Oxide Nanostructures as Gas Sensing Devices*, CRC Press, 2011.
7. Chiu, S.-W.; Tang, K.-T.; Towards a Chemiresistive Sensor – Integrated Electronic Nose: A Review, *Sensors* 13 (2013) 14214–14247.
8. Zhang, C.; Chen, P.; Hu, W.; Organic Field Effect Transistor-Based Gas Sensors, *Chem. Soc. Rev.* 44 (2015) 2087–2107.
9. Qazi, H.H.; Mohammad, A.B.; Akram, M.; Recent Progress in Optical Sensors, *Sensors* 12 (2012) 16522–16556.

10. Gala, U.; Chauhan, H.; Principles and Applications of Raman Spectroscopy in Pharmaceutical Drug Discovery and Development, *Expert Opin. Drug Discov.* 2 (2015) 187–206.
11. Ballantine Jr., D.S.; Martin, S.J.; Ricco, A.J.; Frye, G.C.; Wohltjen, H.; White, R.M.; Zellers, E.T.; *Acoustic Wave Sensors, Theory, Design, and Physico-Chemical Applications*, Academic Press, 1996.
12. Vashist, S.K.; Vashist, P.; Recent Advances in Quartz Crystal-Based Sensors, *J. Sens.* 2011 (2011) 13 pages.
13. Gardner, J.; Bartlett, P.N.; Eds., *Sensors and Sensory Systems for an Electronic Nose*, Springer, 2013.
14. Bartlett, P.N.; *Electronic Noses: Principles and Applications*, Oxford University Press, 1999.
15. Patel, H.K.; *The Electronic Nose: Artificial Olfaction Technology*, Springer Verlag, 2014.

INDEX

- ABCD matrices, 1955
- absolute measurements, 2526
- absolute Seebeck coefficient, 2188–9
- absorbance, 2591, 2594
- absorbance spectrum, 2591, 2609
 - dark current spectrum, 2592
 - reference spectrum, 2592
 - sample spectrum, 2592
- absorption edges (K,L,M), 2503–5
- accuracy, 2246, 2251, 2253, 2254,
2259, 2262, 2265, 2271,
2272, 2313
- accuracy class, 2294, 2295
- acetylation, 2445
- ac-mode, 2002
- acoustic wave, 2722–4
- activation entropy, 2572
- activator, 2227, 2228
- active power, 2276, 2281, 2282, 2285–90,
2303, 2304, 2306
- activity, ion, 2711
- adsorption chromatography
 - applications, 2414–15
 - definition, 2414
 - mobile phases, 2414
 - stationary phases, 2414
- affinity chromatography
 - applications, 2422–3
 - definition, 2421
 - mobile phases, 2422
 - stationary phases, 2422
- affinity ligand, 2421–2
- AFM *see* atomic force microscope (AFM)
- aging-related diseases, 2449, 2450
- a-ions, 2438, 2439
- air-core current transducer, 2262–7
- alarm, 2707
- alcohol OH exchange, 2570–1
- alcohol type, NMR test, 2570–1
- algorithms
 - non-iterative partial alternating least
squares (PCA), 2350
 - PLS-1 algorithm (NIPALS), 2366
 - PLS-2 algorithm (NIPALS), 2367
 - singular value decomposition
(PCA), 2349
- aliasing
 - sample time, 2112

- Allan variance
 - definition, 2113
 - frequency drift, 2117
 - modified, 2117
 - pre-whitening, 2117
 - slope of, 2115, 2123, 2126
 - time, 2117, 2126
- alpha particles, 2076
- ambiguous classification, 2374
- amide bond rotation, 2571
- amino group, 2433
- aminotyrosine, 2444–6
- ammonia, 2711–12
- ampere hour meter, 2275
- Ampere's law, 2263, 2269
- amplitude–distance curve, 2007, 2009, 2010, 2014, 2015
- amplitude–frequency curve, 2003, 2005, 2012
- amplitude modulation, 2003
- analog-to-digital converter, 2288, 2289
- analysis
 - AT1 scale, 2133
 - Fourier, 2117
 - iterative, 2120
 - statistical, 2110
 - worst-case, 2110, 2111
- analysis of variance (ANOVA), 2311
- analytical affinity chromatography, 2423
- angstrom unit, 2500
- angular momentum flowmeter, 2212–13
- anharmonic oscillator, 2590
- anion-exchange chromatography, 2417–18
- antifogging coatings, 72.8, 2668
- APD *see* avalanche photodiode (APD)
- apparent power, 2282, 2285, 2289, 2290, 2306
- aromaticity, NMR test, 2555
- Aron connection, 2287
- aspect ratio, 2668, 2669, 2672, 2673, 2687
- ASTM E-1137, 2185
- at-line NIR analysis, 2609, 2634
- atomic force microscope (AFM), 72.15, 72.20, 72.25, 72.30, 72.31, 2006–8, 2016, 2026, 2660, 2669, 2687
 - contact mode, 2008
- autocollimator (optical), 2168–71, 2176
- Automatic Computer Time Service (ACTS)
 - see* service, time
- autoscaling, 2334
- avalanche photodiode (APD), 2513, 2520
- averaging, data
 - finite length, 2126
 - numerical example, 2131
 - time limits, 2130
- background rejection, 2523–5
- ballast, 2051
- ballast factor (BF), 2051
- ballistic transport, 2684
- bandgap, semiconductor, 72.14, 2675, 2677, 2682, 2690, 2692, 2694
- bandwidth, 2248, 2253, 2255, 2257, 2259, 2261, 2262, 2266, 2272, 2278, 2295, 2297, 2299, 2302–5
- batch modelling, 2392
- beam cleaner, 2518
- beamline, 2512
- Beer–Lambert law, 2593–4
- beetle, 2039
- bend magnet, 2508
- bent crystal Laue analyzer, 2525
- beryllium window, 2512
- beta particles, 2076
- bias, 2365
- bilinear model, 2323, 2355
- bimorph, 2002–4
- bimorph-driven system, 2002, 2012–16, 2018
- bioinformatics, 2434
- biointeraction affinity chromatography, 2423
- b-ions, 2438, 2439
- biosensors, 2708
 - glucose, 2708, 2715
- biospecific elution, 2422
- biotinylation, 2445, 2447
- BIPM (International Bureau of Weights and Measures), 2133
- birefringence, 2268, 2271–2
- bistable state, 2008, 2016

- Bitter pattern technique, 2001
- blackbody, 2045–6
- Bloch, Felix, 2530
- body fluid, 2434
- Boltzmann, 2229, 2231
- Boltzmann distribution, 2231
- bonds
 - breaking, 2661, 2663
 - dangling, 2661
- borohydride ion (BH_4^-), 2560–1
- bottom-up, 2660, 2672 *see also* top-down
- Bragg analyzer, 2525
- Bragg diffraction, 2513
- bremsstrahlung, 2507
- broadband source, 2060, 2061
- burden, 2249–52, 2256

- calibrated, 2227, 2235, 2237, 2242
- calibration, 2232, 2235–8, 2708
 - potentiometry, 2711
 - transfer between NIR instruments, 2603
 - variance, 2337
- calorimetric flowmeter, 2216–17
- capacitance pressure sensors, 2206
- capacitance thermometers, 2192–3
- capacitive coupling, 2003, 2009, 2010, 2020, 2021
- capacitive divider, 2254, 2256–7, 2272
- capacitive voltage transformer (CVT), 2254–5
- capillary action, 2664, 2666
- capsule PRT, 2183–4
- carbon dioxide, 2711–12
- carbon-glass thermometers, 2187
- carbon monoxide, 2707–8
- carbon nanotubes, 72.23, 72.24, 2661, 2673, 2681, 2685–7 *see also* quantum wires
- carbon resistors, 2187
- cardiovascular system and related diseases, 2452, 2456
- carrier frequency, 2537
- Carr–Purcell–Meiboom–Gill sequence *see* CPMG sequence
- Casimir force (measurements of), 2177–78
- cation-exchange chromatography, 2417–18
- CCT *see* correlated color temperature (CCT)
- cerebrospinal fluid, 2434
- Cernox thermometer, 2186–7
- chameleon mechanism, 2164
- characterization, 2247–8
- charge mobility, 2084, 2098
- chemical derivatization, 2444, 2463
- chemical-labeling method, 2445
- chemical shift
 - hydrogen bonding effects, 2556
 - isotope effects, 2549, 2561
 - principles, 2530, 2541, 2552–4
 - range, 2555–6
 - reference standards, 2552–3
 - temperature effects, 2573
 - water in solvents, 2547
- chemometrics, 2309
- chiral separations, 2423
- chroma, 2072
- chromaticity
 - coordinate, 2067–71, 2074
 - diagram, 2067–71
 - discrimination, 2069–70
- chromatogram, 2410–11
- chromatographic-based immunoassays, 2423
- chromatography, 2409
- chromium(III) acetylacetonate complex, 2568–9
- CID *see* collision-induced dissociation (CID)
- clock
 - generic, 2109
 - model, 2120
- cluster analysis, 2318
 - hierarchical cluster analysis (HCA), 2322
 - k-means clustering, 2319
- coalescence temperature, 2572
- coatings, mirror, 2516
- COFRADIC *see* combined fractional diagonal chromatography (COFRADIC)
- collision-induced dissociation (CID), 2438, 2440, 2442, 2463
- color, 2043, 2044, 2046, 2071–2, 2659, 2663, 2675 *see also* colorimetry
- color channel, 2063

- colorimetry, 2063–72
- color matching function, 2065–6
- color rendering index (CRI), 2046, 2074
- color vision, 2063
- column bleed, 2415
- column chromatography, 2412
- combination frequencies, 2589–90
- combined fractional diagonal
 - chromatography (COFRADIC), 2433, 2446, 2449, 2450, 2463
- combined precursor isotopic labeling and isobaric tagging (cPILOT), 2444
- combined uncertainty, 2284
- combustion, 2240, 2241, 2243
- Commission Internationale de l'Éclairage (CIE), 2065
- common-view *see* GPS satellites
- complementary wavelength, 2068
- composite error, 2252–3
- conductivity detector, 2424–5
- cone photoreceptor, 2056–7, 2063, 2064
- confocal imaging, 1985
- Connes advantage, 2602
- contact angle, 72.6, 2665, 2668
- continuous processing, 2392
- continuous scan mode, 2513
- correlated color temperature (CCT), 2045–6, 2074
- correlation, 2284
- correlation loadings, 2331
- correlation spectroscopy (COSY), 2574–7, 2579
- counter, time interval, 2127
- CPMG sequence, 2567
- critical energy, 2509–10
- crossed diodes, 2542
- cross peaks, 2575–9
- cross section, x-ray, 2503–6
- cross talk, 2265
- cryogenics, 2181
- cryogenic sensors, 2181
- crystalline materials, 72.15, 72.23, 2660, 2661, 2668, 2694
- crystal “reflection,” 2514, 2517
- CT *see* current transformer (CT)
- cumulative distribution function, 2078
- current transducer, 2288, 2289, 2291, 2302, 2304, 2305
- current transformer (CT), 2246, 2248–53, 2291, 2302, 2305
 - errors, 2292, 2293
 - nonlinearities, 2294, 2295
 - protective, 2252–3
 - saturation, 2294, 2295
- cycle time, 2567
- damping constant, 2016
- dark current (DC) component, 2254, 2257, 2259, 2262, 2264, 2271
- dark energy, 2164
- dark matter, 2164
- day
 - astronomical, 2109
 - and leap seconds, 2157
- daylight, 2046
- de Broglie wavelength, 72.14, 72.20, 2675, 2683, 2688, 2691
- decay, 2227, 2230–4, 2238
- decay time, 2232–4
- decoupler heating, 2573
- decoupling, 2561–3
- deflection parameter, 2511–12
- delay, transmitting
 - ACTS system estimate, 2150
 - common-view estimate, 2137, 2145
 - GPS satellites, 2144
 - internet methods, 2147
 - melting-pot method, 2138
 - models, 2136
 - two-color refractivity estimate, 2139
 - two-way estimator, 2139
- density of states, 2029
- density of states (DoS), 72.16, 72.26, 2676–8, 2682, 2689–890
- depth of penetration of NIR Radiation, 2612
- derivatives
 - Savitzky–Golay derivative, 2624
 - segment–gap derivative, 2625
- deshielding, 2552

- design of experiments (DoE), 2311
- detection period, 2574
- detectors, 2518–21
- detrending, 2627
- detuning, 2517
- deuterated solvents, 2546–7
- diabetes, 2453, 2456
- dI/dV measurements, 2031
- dielectric constant of PZT materials, 2034
- differential pressure flowmeter, 2213–16
- diffraction gratings, 2598
 - moving gratings, 2598
 - order sorters, 2599
- diffractive analyzers, 2525
- diffuse reflectance measurements, 2594, 2596, 2611
- digital x-ray processor (DXP), 2520
- dimensionless sensitivity, 2196
- 2,2-dimethyl-2-silapentane-5-sulfonate-d6
 - sodium salt *see* DSS
- DIN IEC 60751, 2185
- diode thermometers, 2187–8
- direct force modulation, 2003, 2015
- dispersive mode, 2513
- distortions, 2522
- domains/motifs, 2459
- dominant wavelength, 2068–9
- dopant, 2227, 2229
- doped, 2227, 2231, 2243
- doped oxides, 2243
- doped phosphor, 2227, 2231
- dosimetry, 2081
- double crystal monochromator, 2513–14
- double Kennard–Stone method, 2387
- downfield, definition, 2552
- drug delivery, 72.13, 2673
- DSS chemical structure, 2553
- dual-mixer method, 2128
- dual nature of data, 2324
- duplexer, 2543
- dynamic NMR, 2568–73
- dynamic range, 2670
- dynamic strain measurements, 2205
- dynamic temperature measurements, 2199–201
- ECD *see* electron-capture dissociation (ECD)
- EDI-MS, 2446
- effective counts, 2518–19
- [18]-annulene Chemical shifts, 2555–6
- electrical excitation, 2197
- electric field, 2276–8
- electric potential, 2276, 2278
- electrode
 - Clark, 2713–14
 - disposable, 2715
 - indicator, 2709
 - ion selective, 2709
 - membrane, 2710, 2716–17
 - reference, 2709, 2716
- electromagnetic spectrum, 2043–4, 2585
- electron bunch, 2508
- electron-capture dissociation (ECD), 2440–2, 2463
- electron energy loss spectroscopy, 2507
- electronically compensated current
 - transducer, 2262 *see also* zero-flux transducer
- electronic transformers, 2302, 2303, 2305
- electron-transfer dissociation (ETD), 2440, 2442, 2463
- electron volt (eV), 2500
- electron yield mode detection, 2507, 2523
- electrospray ionization (ESI), 2425–6, 2433, 2449, 2450, 2463
- electrostatic force modulation, 2002, 2003, 2005, 2009, 2010, 2012–21
- ^{11}B , 2560–1
- elutropic strength, 2414
- empirical modelling, 2310
- endogenous nitroproteins, 2432, 2449, 2450, 2463
- energy definition, 2276
- energy flow, 2275, 2279, 2303
- engines, 2240
- enthalpy of activation, 2572
- entrance angle, 2072
- Eöt-wash group, 2166–7, 2171–2, 2175–7
- Ernst angle, 2567
- Ernst, Richard, 2530

- error *see also* Allan variance
 maximum time interval, 2111
 outliers, 2112
 ESI *see* electrospray ionization (ESI)
 ESI-MS/MS spectrum, 2437, 2438
 ESI-MS spectrum, 2437, 2438
 ethyl trans-crotonate ^{13}C spectrum, 2562–3
 ethyl trans-crotonate ^1H spectrum, 2559–60
 europium, 2227–9, 2231
 evaporative light scattering detector, 2424–5
 evolution period, 2574
 exchange rate, 2572
 excitation purity, 2069
 exciting current, 225–2252, 2262
 excluded volume, 2419
 experiment-based biological function
 system, 2459
 explained variation, 2311, 2336
 exploratory data analysis (EDA), 2310, 2317
 exponential, 2230, 2231
- Fabry–Perot filter, 2604
 Fano factor, 2083, 2091
 fast exchange limit, 2569
 fast neutron detection, 2104–6
 fault current, 2252, 2266
 feedback
 polarity, 2007, 2009, 2010, 2015, 2021
 signal, 2002, 2005
 stability, 2018
 Fellgett advantage, 2602
 Fermi contact mechanism, 2557–8
 Ferraris meter, 2288
 fiber optic probes for NIR analysis, 2617
 reflectance probes, 2617
 transmission probes, 2618
 fiber optics, 2717
 field-frequency locking, 2545
 field programmable gate arrays, 2520
 filling factor, 2681
 filtering, 2666–7
 fine structure splitting, 2694
 fission fragments, 2076
 five-times rule, 2060
 fixed points, 2182
 fixed wavelength absorbance detector, 2424
 flight tube, 2526
 flow, 2211–18
 flow meter, 2211–18
 angular momentum, 2212–13
 calorimetric, 2216–17
 differential pressure, 2213–16
 hot-wire anemometer, 2217–18
 laminar flow, 2215–16
 orifice, 2214–15
 packed screen, 2216
 packed spheres, 2216
 positive displacement, 2212
 thermal, 2216–17
 turbine, 2213
 venturi, 2215
 fluorescence, 2225–30, 2232, 2235, 2239,
 2240, 2242, 2243, 2718, 2721–2
 detector, 2424–5
 quenching, 2722
 fluorescence mode detection, 2507, 2521–2
 fluorescent, 222–2229, 2232, 2237, 2238,
 2242, 2243
 fluorescent lamp, 2046, 2048–51, 2058, 2070
 fluorophores, 72.13, 2486, 2673
 focussing conditions, 2514–15
 forced harmonic oscillation, 2004, 2016
 Fourier addition theorem, 2542
 Fourier pair, 2537, 2541
 free energy of activation, 2572
 free induction decay (FID), 2539
 frequency *see also* masers, hydrogen
 aging, 2123, 2132
 cesium standard, 2118, 2121, 2129,
 2131, 2157
 fractional, 2111, 2113
 offset model, 2120
 quartz generator, 2118, 2123, 2126
 rubidium standard, 2118, 2121, 2129,
 2131, 2132
 stability, 2112, 2114
 transmitting, 2135
 frequency excitation range, 2537
 F-residuals, 2342
 full width at half max (FWHM), 1976, 1978

- functionalization, 72.13, 2665, 2673–4, 2691, 2692
- fundamental frequencies, 2588
- G (gravitational constant), 2164–5, 2167, 2176
- GaAlAs diode, 2187–8
- galvanic insulation, 2289, 2291, 2299
- gaseous radiation detectors, 2081–92
- Gaussian
- beams, 1952, 1955, 1964, 1976, 1978
 - line shape, 2541–2
 - statistics, 2079
- gaussmeters, 2219
- Geiger counter, 2520
- Geiger–Muller detectors, 2092
- gel filtration chromatography, 2420
- gel permeation chromatography, 2420
- general (group specific) ligand, 2422
- general relativity, 2163
- germanene, 2681–2682
- germanium resistance thermometer, 2187
- glucose, 2707
- gold nanoparticles
- color, 2663
 - magnetism, 2664
 - melting point, 72.4, 2663
 - 1D conductance, 2688
- GPS satellites
- BIPM tracking schedule, 2146
 - common-view, 2145
 - description of, 2141
 - ionospheric delay, 2139
 - multipath, 2146
 - pseudorange, 2143
 - time codes, 2142
 - t-ram algorithm, 2143
 - tropospheric delay, 2145
- graphene, 72.19, 72.23, 2662–3, 2673, 2674, 2681, 2682, 2685
- graphite, 2716
- gravity (gravitation), 2163–78
- gradients, 2170
 - multipole techniques, 2170
 - Yukawa addition to, 2165
- grazing incidence, 2523
- gyromagnetic ratio, 2530, 2532–3
- Hall effect, 2258–9
- Hall-effect sensors, 2296, 2297
- Hall-effect transducers, 2258, 2262
- closed-loop, 2259–61
 - current transducers, 2259–62
 - open-loop, 2259
 - voltage transducers, 2261
- Hall sensor, 2219–20
- HALO *see* harmonics, alignment, linearity, and offsets (HALO)
- harmonic component, 2289, 2290
- harmonic rejection mirror, 2518
- harmonics, 2246, 2253–4, 2516–17 *see also* nonsinusoidal condition
- harmonics, alignment, linearity, and offsets (HALO), 2522
- heterodyne detection, 1959
- heterodyne method, 2129 *see also* dual-mixer method
- heteronuclear correlation (HETCOR), 2579
- heteronuclear multiple-quantum (HMQC), 2579
- heteronuclear overhauser effect spectroscopy (HOESY), 2579
- heteronuclear single-quantum correlation (HSQC), 2579
- high-performance liquid chromatography (HPLC), 2412
- high-performance thin layer chromatography (HPTLC), 2413
- high pressure sodium (HPS), 2045, 2046
- high specificity ligand, 2422
- history, NMR, 2530–1
- HOD peak, 2546
- Hotelling's T^2 , 2338
- Hotelling's T^3
- MSPC T2 control chart, 2390
- hot-wire anemometer, 2217–18
- HPLC-ESI-MS/MS, 2451, 2453
- hue, 2072
- hydrophilic interaction liquid chromatography (HILIC), 2417

- hydrophilicity/superhydrophilicity, 72.6,
72.8, 2665, 2666, 2668
- hydrophobicity/superhydrophobicity, 72.1,
72.6, 72.7, 2664–6, 2668
- illuminance, 2052–5, 2059, 2063, 2072, 2073
- illuminance meter, 2072, 2073
- immobilized metal-ion affinity
chromatography (IMAC), 2423
- immunoaffinity chromatography (IAC), 2423
- immunoaffinity enrichment, 2463
- immunodepletion, 2423
- immunohistochemistry, 2451
- incandescent lamp, 2045, 2046, 2048, 2051,
2064, 2068
- index of refraction, 2515
- industrial-grade platinum resistance
thermometer (IPRT), 2184–5
- inelastic x-ray scattering, 2507
- inertance, 2216
- inertia motor, 2038
- inflammation-related diseases, 2449, 2450
- influence plot, 2342
MSPC, 2391
- infrared (IR) energy, 2043, 2044, 2059–60
- infrared (IR) region of the spectrum
far infrared (FIR) region, 2588
mid infrared (MIR) region, 2586, 2588
near infrared (NIR) region, 2583, 2588–9
- Ingenuity Pathway Analysis (IPA), 2459
- in-line NIR analysis, 2609, 2635
- instantaneous power, 2277, 2279–81, 2289,
2290, 2304, 2305
- instrumentation, 2542–50
analog acquisition, 2483
frequency-domain lifetime, 2484
light detectors, 2481
light source, 2480
monochromator, 2480
photon counting acquisition, 2483
time-domain lifetime, 2384
- instrument transformer, 2246–55
- intact pharmaceutical tablet analysis,
2616, 2637
- integrating sphere, 2062–3
- The integrating sphere, 2612
- integration time, 2601, 2606
- integrator, 2264–6
- interactance measurements, 2613
- internal reflection, 2717–18
- international atomic time (TAI), 2133
- International Earth and Reference System
Service (IERS), 2158
- International Illumination Commission
(CIE), 2065
- InterProScan, 2459
- inverse-square law (ISL), 2053, 2060,
2163–6, 2170–3, 2175–8
constraints on, 2166, 2171–2, 2175
- Inverse Synthetic Aperture Radar (ISAR),
1966, 1972
- inversion recovery sequence, 2567
- ion chromatography, 2419
- ion-exchange chromatography (IEC)
anion-exchange chromatography, 2417–8
applications, 2418–19
cation-exchange chromatography, 2417–18
definition, 2417
ion chromatography, 2419
mobile phases, 2417–19
stationary phases, 2417–18
- ionic strength, 2711
- ionization chambers, 2085–9, 2519–20
- ion selective field effect transistor
(ISFET), 2716
- IR-MALDI-FT-ICR-MS, 2436
- isobaric tags for relative and absolute
quantification (iTRAQ), 2433, 2434,
2444, 2445, 2449, 2463
- isocratic elution, 2413
- isoelectric focusing (IEF), 2450
- isolation amplifier, 2257–8
- ITS-90 temperature scale, 2182–6
- Jacquinot advantage, 2602
- J coupling *see* scalar coupling
- Kalman estimator, 2133
state equations, 2134
- Kelvin probe microscopy, 2020

- Kerr effect, 2001
- kidney disease, 2453, 2456
- Knife-Edge method, 1976
- k*-space, 2506
- laboratory reference frame, 2534–5
- laminar flow element, 2215–16
- Landau level, 72.18, 2680–1
- lanthanum fluoride, 2710
- La₂O₂S:Eu, 2227–30, 2232, 2233, 2239, 2243
- Larmor equation, 2533
- Larmor formula, 2509
- Larmor frequency, 2531–2
- laser, 72.9, 72.17, 2668, 2672, 2675, 2677–9, 2690, 2691, 2693
 - amplification, 1948
 - beam propagation, 1954
 - cavity, 1949
 - CO₂, 1956, 1972, 1982
 - FIR, 1956, 1972, 1982
 - fundamental mode, 1951
 - history, 1945
 - imaging, 1972, 1974, 1982, 1985, 1987, 1994
 - mode-locking, 1991
 - principles, 1946
 - Q-switching, 1991
 - quantum cascade lasers (QCL), 1962
 - Ruby, 1948
 - titanium–sapphire (Ti:Al₂O₃), 1991, 1995
- LC-ESI-MS/MS, 2450–5
- LC-MS/MS, 2433, 2444, 2445
- lead zirconate titanate (PZT), 2032
- leap-seconds, 2143
 - implementation, 2158
 - in time scale, 2157
- least squares fitting, 2353
- leverage, 2339
- lifetime, 2226–8, 2230–3, 2235, 2237, 2242, 2243
 - measurements, 2232
 - thermometry, 2243
- light, 2043–4
- light emitting diode (LED), 2045, 2048, 2051–2, 2058–61, 2063
- light polarization, 2267–72
- light trapping, 2672
- limit of detection, 2315
- limit of quantification (LOQ), 2315
- linear discriminant analysis (LDA), 2370
 - principal component analysis–linear discriminant analysis (PCA–LDA), 2370
- linearity, 2248, 2257, 2259, 2263, 2272
- linear stopping power, 2077
- line width, 2542, 2548–50, 2566–7
- liquid chromatography (LC), 2410–11
 - adsorption chromatography, 2414–15
 - affinity chromatography, 2421–3
 - columns, 2426
 - comparison with gas chromatography, 2411–12
 - definition, 2409
 - detectors, 2423–6
 - ion-exchange chromatography, 2417–19
 - normal-phase liquid chromatography, 2415–16
 - pumps, 2426
 - reversed-phase liquid chromatography, 2415–17
 - size-exclusion chromatography, 2419–20
 - support materials, 2412–13
- liquid chromatography/electrochemical detection (LC-EC), 2424–5
- liquid chromatography/mass spectrometry (LC/MS), 2424–6
- liquid level, 2218–19
- liquid level meters, 2218–19
 - capacitance, 2218
 - hydrostatic head, 2218
 - resistive, 2218
 - superconducting, 2219
- locking *see* field-frequency locking
- longitudinal relaxation time *see* relaxation time, spin-lattice
- long-wavelength (L) cone, 2063, 2064
- long wave NIR region, 2593, 2607
- Lorentzian, 2017
- Lorentzian line shape, 2540–1
- louse, 2037

- low abundance, 2432
- low-performance liquid chromatography, 2412
- low power current transformer (LPCT), 2246
- low power voltage transformer (LPVT), 2246
- LTQ velos, 2442, 2443
- luminaire, 2048–52, 2061
- luminaire efficiency, 2050
- luminance, 2054, 2055, 2059, 2072–3
- luminance meter, 2072–3
- luminous efficacy (electrical), 2052
- luminous efficacy (optical), 2052
- luminous efficiency function, 2056–7, 2060, 2067, 2074
- luminous flux, 2047, 2057–9, 2061–3
- luminous intensity, 2047, 2059
- luminous intensity distribution, 2047–9

- MacAdam ellipse, 2070, 2071
- macroscopic quantum tunneling (MQT), 2001
- MAD *see* metastable atom-activated dissociation (MAD)
- magic angle spinning, 2525
- magnetic core, 2249, 2254, 2259, 2262
- magnetic field, 2219–20
 - homogeneity, 2548
 - shimming, 2548
 - strength, 2533, 2554–5, 2560
- magnetic field strength, 2512, 2680
- magnetic flux density *see* magnetic field strength
- magnetic force, 2002–4, 2007, 2016, 2019
- magnetic force gradient, 2019
- magnetic force microscopy (MFM), 2001–6, 2009, 2010, 2012, 2013, 2015, 2016, 2018–21, 2026
- magnetic moment, 2531–2
- magnetic resonance imaging *see* MRI
- magnetism, 72.5, 2664
- magnetogyric ratio *see* gyromagnetic ratio
- magnetometers, 2219
 - Hall effect, 2219–20
 - superconducting quantum interference device (SQUID), 2219
- magnetoresistance, 72.18, 2660, 2679–80
- magnetoresistive sensors, 2219–20
- Mahalanobis distance, 2338–9
- Majorana fermions, 2688
- MALDI *see* matrix-assisted laser desorption ionization (MALDI)
- masers, hydrogen, 2119
- mass spectrometry, 2432
- matrix-assisted laser desorption ionization (MALDI), 2432, 2434, 2435, 2446, 2449–54, 2463
- maximum time interval error (MTIE) *see* error
- Maxwell equations, 2277
- mean, 2078
- mean free path, 2077
- mean ionization energy, 2080
- measurements
 - anisotropy and polarization, 2492
 - decay times of fluorescence, 2490
 - emission spectrum, 2488
 - excitation spectrum, 2487
 - methods, 2275, 2282
 - dual mixers, 2128
 - heterodyne method, 2129
 - time-interval counter, 2127
 - quantum yield, 2492
- mechanisms, 2228
- medium-wavelength (M) cone, 2063, 2064
- melting point, 72.4, 2663
- meridional focussing, 2515
- mesopic luminance, 2057
- MetaCore Pathway Analysis programs, 2459
- metal alloy strain gages, 2202–5
- metal ions, 2721–2
- metallic resistance thermometers, 2183–6
- metal organic chemical vapor deposition, 72.12, 2660, 2672, 2693
- metal oxides, 2715
- metamer, 2064, 2065
- metastable atom-activated dissociation (MAD), 2440–2, 2463
- method of images, 2087–9
- MFM *see* magnetic force microscopy (MFM)
- Michelson interferometer, 2602

- microbore column, 2426
- micro-cantilevers, 2177
- microelectromechanical spectrometers (MEMS)
 - linear variable filter (LVF) instruments, 2605
 - oscillator, 2118
- mixing period, 2574
- mobile phase
 - adsorption chromatography, 2414
 - affinity chromatography, 2422
 - definition, 2409
 - ion-exchange chromatography, 2417–19
 - normal-phase liquid chromatography, 2415
 - reversed-phase liquid chromatography, 2416
 - role in liquid chromatography, 2413
 - size-exclusion chromatography, 2420
- molar absorptivity coefficient, 2589, 2593
- molecular beam epitaxy, 2660, 2679, 2693
- molecular imprinting, 2422
- monochromator, 2074
- monolithic supports, 2412
- Moore–Penrose pseudoinverse, 2354
- Moses effect, 72.7
- Motif scan, 2459
- motion control, 2512
- MQT *see* macroscopic quantum tunneling (MQT)
- MRI, 2531
- multidimensional NMR, 2573–9
- multiple linear regression (MLR), 2354
- multiplicative scatter correction (MSC), 2629
 - extended multiplicative scatter correction, 2631
- multivariate analysis, 2317
- multivariate classification, 2310, 2369
- multivariate projection, 2389
- multivariate regression, 2310, 2352
- Multivariate Statistical Process Control (MSPC), 2389
 - practical implementation, 2394
- Munsell color system, 2071–2
- nanocatalysis, 2662, 2668, 2670–2
- nanofluidic filtering, 2666–7
- nanofluidics, 2666–7
- nanoforests, 72.11, 2671
- nanomaterials
 - applications, 2657
 - characterization, 2660
 - definition, 2657
 - in nature, 72.1, 72.2, 2657, 2659, 2665
 - synthesis, 2660, 2693
 - topics not covered, 2658–9
- nanomembranes, 72.10, 2669–70, 2674
- nanopillars, 72.12, 2668, 2672–3, 2688, 2693
 - see also* nanowires
- nanoporosity, 72.7–72.9, 2668–9
- nanoscale
 - change in properties at, 2658
 - shrinking objects to, 2658
- nanowires, 72.12, 2668, 2672–3, 2688, 2693
 - see also* nanopillars
- near-infrared (NIR) instrumentation
 - acousto optical tunable filters (AOTF), 2603
 - filter based instruments, 2597
 - Fourier transform (FT) instruments, 2601
 - generic instrument configuration, 2595
 - holographic grating based instruments, 2598
 - linear variable filter (LVF) instruments, 2605
 - microelectromechanical spectrometers (MEMS), 2604
 - postdispersive instruments, 2596, 2600
 - predispersive instruments, 2596, 2601
 - stationary spectrograph instruments, 2600
 - transmission configuration, 2596
- near-infrared (NIR) spectroscopy, 2718
 - agricultural applications, 2634
 - bioprocessing applications, 2643
 - detectors for
 - diode array detector, 2600–1
 - indium gallium arsenide (InGaAs) detectors, 2606, 2608
 - lead sulfide (PbS) detectors, 2606–7
 - silicon (Si) detectors, 2606

- near-infrared (NIR) spectroscopy (*cont'd*)
 - dipole moment induced by, 2590
 - food and beverage applications, 2646
 - petrochemical applications, 2644
 - pharmaceutical/biopharmaceutical applications, 2636
 - process analytical technology (PAT), 2606, 2639
 - for raw material identification, 2636
- negative temperature coefficient, 2186
- network time protocol (NTP)
 - delay asymmetry, 2149
 - description of, 2147
 - polling interval, 2152
 - white phase noise, 2153
- neurodegenerative diseases, 2452, 2456
- The neurovisual system, 2452, 2456
- neutron detection, 2100–6
- neutrons, 2076
- nicotinamide spectrum, 2571–2
- nitroproteomics, 2432, 2443, 2444
- nitro group, 2433
- nitropeptide, 2431
- nitroprotein, 2431
- nitroprotein–protein complexes, 2457, 2459
- nitroproteome, 2432
- nitrotyrosine, 2431
- nitrotyrosine affinity column (NTAC), 2447, 2448, 2453, 2462
- nitrotyrosine antibody-based immunoaffinity, 2433
- NMR time scale, 2557, 2569
- Nobel prize, 2530–1
- No classification, 2374
- NOESY *see* nuclear overhauser effect spectroscopy (NOESY)
- noise *see also* Allan variance
 - flicker, 2124
 - optimum averaging, 2125, 2126
 - white frequency, 2119, 2122
 - white phase, 2114, 2116, 2121
- noncontact region, 2007, 2008, 2014, 2015
- nonsinusoidal condition, 2246, 2253–4 *see also* harmonics
- non-specific elution, 2422
- normal phase liquid chromatography (NPLC), 2415–16
- noscopine spectra, 2576–8
- NTAC-vMALDI-MS/MS, 2451
- nuclear overhauser effect *see* nuclear overhauser enhancement (NOE)
- nuclear overhauser effect spectroscopy (NOESY), 2574, 2577–9
- nuclear overhauser enhancement (NOE), 2562
- nuclide, 2531, 2532
- number of theoretical plates, 2412
- nutation angle, 2536
- off resonance, 2540
- ω component, 2004, 2012, 2021
- 1D edge state, 2684, 2687
- one-dimensional electrophoresis (1DE), 2454
- 1-pulse sequence, 2550
- on-line NIR analysis, 2609
- on resonance, 2535, 2540
- open-loop curve, 2008
- optical activity, 2268
- optical coherence tomography (OCT), 1987
- optical detection mode detection, 2507
- optical fibers, 2267, 2269, 2270, 2717
- optical metamaterials, 2659
- optical radiation, 2044
- optical transducer, 2267–72
 - current transducer, 2268–72
 - voltage transducer, 2271–2
- orifice flowmeter, 2214–15
- oscillators
 - model equation, 2130
 - types of, 2118
- outliers
 - PCA, 2338
- over-absorption, 2522
- overtone frequencies, 2588–90
- oxidative damage, 2463
- oxidative-/nitrate-mediated modification, 2463
- oxidative/nitrative stress, 2431

- oxygen
 - fluorescence quenching, 2722
 - potentiometry, 2712–13
 - voltammetry, 2713–14
- paramagnetic ion effects, 2566
- Parkinson's disease, 2456
- parsimony of multivariate models, 2336
- partial least squares discriminant analysis (PLS-DA), 2381
- partial least squares regression (PLSR), 2356
 - loading weights, 2357
 - PLS factor, 2356
 - PLS scores, 2356
 - PLS X-loadings, 2357
 - PLS Y-loadings, 2357
 - X-residuals in PLS, 2361
 - Y-residuals in PLS, 2361
- partition chromatography
 - applications, 2416–7
 - definition, 2415
 - hydrophilic interaction liquid chromatography, 2417
 - mobile phases, 2415–17
 - normal-phase liquid chromatography, 2415–16
 - reversed-phase liquid chromatography, 2415–17
 - stationary phases, 2415–17
- Pascal's triangle, 2558–9
- pathlength, 2594
- pathway system networks, 2459
- pellicular supports, 2412
- Peptizer algorithm, 2446
- perfusion particles, 2412
- period
 - accuracy, 2110
 - reference, 2109
 - stability, 2110
- Pfam, 2459
- pH, 2707, 2721
- phase detector, 2543
- phase displacement (or phase error), 2248, 2252, 2254
- phase shift, 2275, 2280
- phasing, 2540, 2551
- phospholipid quantitation, 2568–9
- phosphor, 2225–8, 2230–2, 2235, 2239, 2240, 2242, 2244
- phosphorescence, 2243–4
- phosphor thermometry, 2227, 2238, 2240, 2244
- photochemical decomposition, 2435, 2436
- photodecomposition, 2433
- photodiode array detector, 2424
- photodiode diodes, 2520
- photolithography, 2660, 2671
- photoluminescence, 72.27, 2660
- photometry, 2056–63
- photomultiplier tubes, 2094–6
- photonic crystals, 72.2, 2659, 2672, 2693–4
- photons, properties, 2500
- photopic luminous efficiency function, 2056–7, 2060, 2061, 2066, 2067, 2074
- piezoelectric effect, 2722–3
- piezoelectricity, 2032
- piezoelectric pressure sensors, 2210–11
- piezoresistive pressure sensors, 2208–10
- PIPS detector, 2520
- planar chromatography, 2413
- plant diseases, 2453, 2456
- plasma source, 2507
- plate height, 2412
- platinum nanoclusters, magnetism, 72.5, 2664
- platinum resistance thermometer, 2182–6
- PLTS-2000, 2182
- Pockels effect, 2271–2
- Poisson statistics, 2078–9
- polarization vector, 2500
- polarized XAFS, 2525
- polarizer, 1981
- poling of PZT materials, 2033
- polling *see also* network time protocol (NTP)
 - cost/benefit analysis, 2156
 - error detection, 2155
 - optimum accuracy, 2154
- polymer swelling, 2716
- positive displacement flowmeter, 2212
- positive-intrinsic-negative (PIN) diodes, 2520

- potential transformer (PT), 2246, 2248–54
see also voltage transformer (VT)
- power
 amplifier, 2538, 2542–3
 definition, 2276, 2304
 factor, 2282, 2303
- power quality (PQ), 2245, 2266, 2289, 2302–6
- Poynting vector, 2277
- preamplifier, 2543
- precession
 frequency, 2533
- precision, 2315
 class, 2251, 2253
 intermediate, 2315
- predicted residual sum of squares
 (PRESS), 2364
- predicted *vs.* reference plot, 2360
- preferred orientation, 2525
- preparation period, 2574
- preprocessing
 discrete data, 2333
 NIR data for chemometric analysis, 2621
 additive effects, 2621
 multiplicative effects, 2627
 spectral data, 2333
- pressure, 2205–11
- pressure sensors, 2205–11
 capacitance, 2206
 piezoelectric, 2210–11
 piezoresistive, 2208–10
 variable reluctance, 2206–8
- primary light source, 2065
- primary standards, 2182–3
- primary thermometer, 2182–3
- principal component analysis (PCA), 2323
 loadings, 2328
 scores, 2328
- principal component regression (PCR), 2355
- probability mass function, 2078
- process control, 2707, 2719
- ProDom, 2459
- propane, 2715
- proportional counters, 2090–2, 2520
- protons, 2076
- pseudorange, 2143 *see also* GPS satellites
- pulse-and-collect sequence *see* 1-pulse
 sequence
- pump-probe, 2513
- Purcell, Edward, 2530
- purple boundary, 2068, 2069
- Q-residuals, 2341
 MSPC control chart, 2391
- QSTAR elite (QTOF), 2442, 2443
- quality factor, 2016
- quantitative NMR, 2565, 2567–8
- quantization, 72.14, 72.21, 72.28, 2675–8,
 2680, 2682–4, 2686–9, 2691
- quantum cascade laser, 72.17, 2678–9
- quantum confinement, 72.14, 72.15, 2675,
 2681, 2682, 2691, 2692
- quantum dots, 72.15, 2676, 2691–4
 colloidal, 72.13, 72.29, 2673–4, 2691–3
 droplet epitaxy, 2694
 epitaxial self-assembly, 72.30, 72.31,
 2693–4
 gate-defined, 72.28, 2691
- quantum Hall effect
 fractional, 2679, 2681
 integer, 72.18, 2679–81
- quantum mechanics, 2659
- quantum well lasers, 2677–9
- quantum wells, 72.14–72.16, 72.20, 72.21,
 72.28, 2675–83, 2691
- quantum wires, 72.15, 2675–6, 2682–5,
 2688–91
 epitaxial self-assembly, 72.25–72.27,
 2688–91
 gate-defined, 72.20–72.22, 2682–5
 in nanopillars, 2688
- quartz crystal microbalance, 2722–3
- quartz-crystal oscillators, 2118
- radiant power, 2056–9
- radioactive decay, 2076, 2078
- radioactivity, 2076, 2078
- radio detection and ranging (RADAR), 1965
- radiofrequency (RF)
 magnetic field, 2544
 pulse, 2537–8

- pulse width calibration, 2551–2
- Raman spectroscopy, 1996
- rat brain, 2569
- rated life (light source), 2051–2
- ratio error, 2247, 2251, 2252, 2254
- ratio of prediction error to standard deviation (RPD), 2365
- rat liver, 2539, 2569
- Rayleigh scattering, 2719
- reactive power, 2276, 2281, 2282, 2285, 2287, 2289, 2290, 2304, 2306
- reactive volt-ampere, 2282
- receptivity, 2532, 2540
- reciprocating pump, 2426
- redundant array of independent measurements (RAIM), 2143 *see also* GPS satellites
- reference deconvolution, 2550
- references, 2404, 2652–6
- reference source (chromaticity), 2068, 2069
- reflectance, 2054–6, 2063
- reflections, crystal, 2514, 2517
- refractive index detector, 2423–4
- regression coefficients, 2358
- relative Seebeck coefficient, 2188
- relative time modelling (RTM), 2393
- relaxation
 - delay, 2550
 - spin-lattice, 2563–5, 2567–8
 - spin-spin, 2565–7
- relaxation time
 - spin-lattice, 2564
 - spin-spin, 2565
- repeatability, 2315
- representative data, 2311
- residuals
 - PCA, 2337
- residual variance, 2336
- resistance quantum, 2681
- resolution
 - slit width and true resolution, 2599–600
- resonance frequency, 2004, 2007, 2009, 2014
- response time, 2199–201, 2205, 2209, 2211, 2217
- retention time, 2411
- retention volume, 2411
- reversed-phase liquid chromatography (RPLC), 2415–17
- RhFe thermometer, 2186
- ring current, 2555–6
- robustness, 2315
- rod photoreceptor, 2056–7
- Rogowski coil, 2262–7, 2297–9, 2303
 - see also* air-core current transducer
- root mean square errors
 - calibration (RMSEC), 2364
 - cross validation (RMSECV), 2364
 - prediction (RMSEP), 2364
- rotating frame overhauser enhancement (ROESY), 2579
- rotating reference frame, 2535
- sagittal focussing, 2515
- Sagnac effect, 2151
- sample geometry, 2523
- sample preparation, 2521
- sample preparation for NIR analysis
 - grains and seeds, 2614
 - pharmaceutical tablets, 2615
 - silage and sugarcane, 2615
 - solid powders, 2614
- sample spinning, 2548–9
- sampling frequency, 2289, 2290, 2304
- sampling theorem, 2290
- saturation, 2252, 2254, 2259, 2565
- saturation recovery sequence, 2567
- scalar coupling, 2558
- scales, time
 - definitions, 2133
 - national, 2157
- scanning electron microscopy, 72.1, 72.2, 72.7, 72.11, 72.12
- scanning Hall probe, 2026
- scanning near-field optic microscope (SNOM/NSOM), 2026
- scanning tunneling microscope (STM), 72.15, 2025, 2660
 - coarse approach, 2036
 - constant current mode, 2035
 - constant height mode, 2035
 - scanner, 2032

- scanning tunneling microscopy and spectroscopy (STMS), 2029
- ScanProsite, 2459
- scintillators, 2092–6
- scotopic luminous efficiency function, 2056–7
- screens, 2216
- SCX-LC-ESI-MS/MS, 2452
- secondary standard, 2183
- second-order scalar coupling, 2558–60
- Seebeck coefficient, 2188–92
- selectivity, 2708
- selenium, 2059, 2072
- self-absorption, 2522
- self-actuation, 2016
- self-assembly, compressively strained, 72.25–72.27, 72.30, 2688–90, 2693–4
- self-assembly, tensile strained, 72.31, 2690–1, 2694
- self-heating, 2197–8
- semiconducting radiation detectors, 2096–9
- semiconductor-like resistance thermometers, 2186–7
- sensitivity, 2193–7
- sensor(s), 2181–2, 2669–70
- separation, 2002, 2003, 2007, 2010, 2019–21
- service, time *see also* WWVB
 - NIST Automatic Computer Time Service, 2149
- servo force, 2005, 2016, 2019, 2021
- shielding, 2553
- short-wavelength (S) cone, 2063, 2064
- short wave NIR region, 2592
- shot noise, 2603
- Shubnikov–de Haas oscillations, 72.18, 2680
- shunt, 2255–6
- signal averaging, 2513, 2551
- signaling networks, 2460, 2461
- signal/noise ratio, 2551
- signal to noise ratio, 2513
- silicene, 2681–2
- silicon, 2059–60, 2072
 - diode, 2187–8
 - drift detector, 2520
 - monochromator crystal, 2514
- single-phase circuits, 2285
- single-tube scanner, 2034
- sinusoidal condition, 2278, 2279, 2281, 2287, 2303, 2304
- size-exclusion chromatography (SEC)
 - applications, 2420
 - definition, 2419
 - gel filtration chromatography, 2420
 - gel permeation chromatography, 2420
 - mobile phases, 2419–20
 - stationary phases, 2420
- slow exchange limit, 2570
- smart meter, 2304
- snap-to-contact, 2008
- Soft Independent Modelling of Class Analogy (SIMCA), 2372
 - classification table, 2375
 - Coomans plot, 2375
 - distance vs. leverage plot, 2376
 - model distance, 2377
 - modelling power, 2378
 - variable discriminant power, 2378
- solid state detector, 2520–1
- solvent programming, 2413
- spectral power distribution (SPD), 2044–5, 2057, 2064, 2074
- spectroradiometer, 2074
- specular reflectance, 2607
- spheres, 2216
- spinning side bands, 2548–9
- spin-spin coupling *see* scalar coupling
- square law dependence, 2017
- stability, 2003, 2008, 2010, 2012, 2015, 2016, 2018, 2019, 2021 *see also* Allan variance
 - frequency, 2114
- standard deviation, 2078
- standard error of laboratory (SEL), 2364
- standard error of prediction (SEP), 2365
- standard normal variate (SNV), 2627
- standard platinum resistance thermometer (SPRT), 2182–6
- standard thermometer, 2182–5

- standard uncertainty, 2284
- stationary phase
 - adsorption chromatography, 2414
 - affinity chromatography, 2422
 - definition, 2409
 - ion-exchange chromatography, 2417–18
 - normal-phase liquid chromatography, 2415–16
 - reversed-phase liquid chromatography, 2416
 - size-exclusion chromatography, 2420
- step-scan mode, 2512
- STM *see* scanning tunneling microscope (STM)
- Stokes scattering, 2719
- Stokes theorem, 2276
- strain, 2201–5
- strain gages, 2201–5
 - metal alloy, 2202–3
 - semiconductor, 2202
- strategies, measurement
 - averaging time limit, 2130
 - frequency aging, 2132
 - GPS satellites, 2141
 - numerical example, 2131
 - optimum domain, 2133
 - two-way methods, 2147, 2151
 - white phase noise, 2129
- String theory, 2163–4
- strong mobile phase
 - adsorption chromatography, 2414
 - affinity chromatography, 2422
 - definition, 2413
 - ion-exchange chromatography, 2418
 - normal-phase liquid chromatography, 2415
 - reversed-phase liquid chromatography, 2416
- structural biology, 2434
- sum of squared residuals, 2363
- sunlight, 2045
- superconducting quantum interference device (SQUID), 2192, 2219
- support vector machine classification (SVMC), 2383
- Kernel functions, 2384
 - SVM Type 1, 2385
 - SVM Type 2, 2385
- surface-area-to-volume (SAV) ratio, 72.3, 2659, 2661–4, 2666, 2668, 2669, 2671, 2672, 2674
- surfaces, 2661–2
 - energies, 2661, 2664
- synchrotron radiation, 2508
- synthetic multilayers, 2514
- syringe pump, 2426
- systems biology methods, 2434
- T^1 *see* relaxation time, spin-lattice
- T^2 *see* relaxation time, spin-spin
- $T2^*$, 2566
- Taguchi, 2715
- tandem mass tags (TMT), 2433, 2444, 2445, 2449, 2463
- tapping/lift mode, 2003
- tapping region, 2007, 2008, 2010, 2014, 2015
- tautomerization, 2556–7
- temperature, 2182–201, 2265, 2269
 - effects, 2571–3
 - measurements, 2572–3
 - resolution, 2193
- tempering, 2197–9
- ^{10}B , 2560–1
- terahertz
 - compact radar range, 1972
 - FIR lasers, 1956, 1972, 1982
 - imaging, 1981
 - quantum cascade lasers (QCL), 1963
- terahertz (THz) emission, 72.9, 72.17, 2668–9, 2679
- terminology, 2397, 2648–51
- tetramethylsilane *see* TMS
- thermal anchoring, 2197–9
- thermal conduction, 2198
- thermal flowmeter, 2216–17
- thermal neutron detection, 2100–4
- thermistors, 2187
- thermocouples, 2188–92

- thermographic, 2226, 2238, 2243, 2244
- thermographic phosphor, 2238, 2243
- thermometers, 2182–201
 - capacitance, 2192–3
 - carbon, 2187
 - carbon-glass, 2187
 - Cernox (zirconium oxynitride), 2186–7
 - diode, 2187–8
 - germanium, 2187
 - platinum, 2182–6
 - RhFe, 2186
 - ruthenium oxide (RuO₂), 2186
 - thermistor, 2187
 - thermocouples, 2188–92
- thermometer use and comparisons, 2193–9
- thermometry, 2227, 2238, 2240, 2243, 2244
- thermopile, 2192
- thickness effects, 2521
- thin-layer chromatography (TLC), 2413
- ¹³C satellite, 2549, 2558
- (3R, 4S)-1-(4-(aminomethyl) phenylsulfonyl) pyrrolidine-3,4-diol (APPD), 2449
- 3D spatial structure, 2462
- three-phase circuits, 2286
- time
 - dead time, measurement, 2113
 - difference, 2111
 - difference model, 2130
 - different devices, comparing, 2119, 2127
 - error, 2121
 - GPS satellite estimate, 2143
 - measurements, 2119, 2127
 - model, difference, 2120
 - origin, 2110
 - transmitting, 2135
 - variance, 2116
- time-independent Schrödinger equation, 2026
- time resolved, 2513
- time transfer *see also* GPS satellites; network
 - time protocol (NTP)
 - accuracy, 2149
 - ACTS protocol, 2149
 - asymmetric delay, 2149
 - WWVB, 2135
- tip angle *see* nutation angle
- TMS
 - chemical structure, 2553
 - spectrum, 2549
- tolerance grades, 2185
- toluene, 2719
- top-down, 2660, 2672 *see also* bottom-up
- torsion balance *see* torsion pendulum
- torsion pendulum(s), 2164, 2166–78
 - readout noise of, 2168–9
 - rotating, 2176
 - thermal noise in, 2168–9
 - tilt-twist effect in, 2171
- total correlation spectroscopy (TOCSY), 2579
- traceability, 2158
 - and GPS signals, 2159
- transformation ratio, 2247, 2249–51, 2256, 2260
- transmission coefficient, 2027
- transmission electron microscopy, 72.15, 72.31, 2660
- transmission measurements, 2594, 2596, 2599, 2609–10
- transmission mode detection, 2507, 2521
- transverse electromagnetic wave, 2277
- transverse relaxation time *see* relaxation
 - time, spin-spin
- tributyl phosphate, 2568–9
- 1-(trimethylsilyl)propionic-2,2,3,3-d4 acid
 - sodium salt *see* TSP
- tripod scanner, 2034
- tristimulus value, 2066–7
- TSP chemical structure, 2553
- Tswett, Mikhail, 2409
- tube
 - camber, 2548–9
 - concentricity, 2548–9
 - sample, 2548–9
- tumorigenesis, 2456
- tungsten, 2217–18
- tunnel current, 2028, 2030
- tunneling phenomenon, 2026
- turbine, 2225, 2236, 2239–41, 2244
- turbine flowmeter, 2213
- 2-butanol spectrum, 2570–80
- two-dimensional electrophoresis (2DE), 2453

- two-dimensional gel electrophoresis (2DGE), 2446, 2447
- two-dimensional (2D) NMR, 2574–9
- 2D materials, 72.19, 2661, 2681–2
- two-pass technique, 2003
- 2,4-pentanedione spectrum, 2556–7
- two-photon fluorescence microscopy, 1994
- 2 ω component, 2009, 2012, 2017, 2019, 2021
- tyrosine nitration, 2431, 2432
- tyrosine phosphorylation motif, 2432
- ultra-performance liquid chromatography (UPLC), 2412
- ultrathin-layer chromatography (UTLC), 2413
- ultraviolet (UV) energy, 2043, 2044, 2059–60
- uncertainty evaluation, 2284
- undulator, 2511
- unexplained variation, 2311
- uniform chromaticity diagram, 2071–2
- unique classification, 2374
- univariate analysis, 2317
- upfield, definition, 2552
- UTC (coordinated universal time)
 - definition, 2110, 2133
 - and GPS satellites, 2143
 - optimum measurement, 2133
 - and UT1, 2143
- UT1 time scale, 2143
- UV/Vis absorbance detector, 2424
- vacuum energy, 2164
- vacuum, ultrahigh, 2512
- validation of chemometric models
 - categorical cross validation, 2388
 - cross validation, 2313, 2388
 - full cross validation, 2388
 - segmented cross validation, 2388
 - systematic cross validation, 2388
 - test set validation, 2313, 2386
- validation variance, 2337
- value, 2072
- van der Waals forces, 72.19, 2662–3, 2682
- variable reluctance pressure sensors, 2206–8
- variable wavelength absorbance
 - detector, 2424
- variance time, 2116 *see also* Allan variance
- varmeter, 2285, 2287
- venture flow meter, 2215
- Verdet constant, 2269, 2270
- vibrational transitions, 2719
- virtual state, 2719
- visible light, 2044
- visible region, 2586
- vMALDI-LTQ, 2439
- void volume, 2419
- voltage-controlled oscillator (VCO), 2005
- voltage divider, 2256–7
- voltage transducer, 2288, 2289, 2291, 2299, 2302, 2304, 2305
- voltage transformer (VT), 2246, 2248–54, 2299, 2302, 2305 *see also* potential transformer (PT)
- volt-ampere, 2282
- VT *see* voltage transformer (VT)
- VT errors, 2300–2
- watt, 2282
- wattmeter, 2285, 2287, 2288, 2304
- wavelength band (color), 2044
- wavenumber, electron, 2506
- weak affinity chromatography, 2422
- weak equivalence principle (WEP), 2163–5, 2167, 2170, 2173, 2175–7
 - constraints on, 2167, 2176
- weak mobile phase
 - adsorption chromatography, 2414
 - affinity chromatography, 2422
 - definition, 2413
 - ion-exchange chromatography, 2418
 - normal-phase liquid chromatography, 2415
 - reversed-phase liquid chromatography, 2416
- Weidemann–Franz law, 2197–8
- Western blotting, 2446, 2447, 2450–4, 2462
- wiggler, 2510
- woods anomaly, 2603
- work function, 2029
- WWVB (NIST Radio Station), 2135

- X-ray, 2076
 - absorption, 2502–3
 - attenuator, 2524
 - crystallography, 2462
 - crystal structure, 2462
 - diffraction, 2501–2
 - filter, 2524
 - lenses, 2515
 - mirrors, 2515–18
 - properties, 2499–500
 - source spectrum, 2509–11
 - tube, 2507
- X-ray absorption fine structure (XAFS),
 - 2502, 2505–26
- X-ray diffraction, 2660
- X-residuals, 2340
- YAG, 2232, 2234–7, 2243, 2244
- y-ions, 2439
- zero-flux transducer, 2262 *see also*
 - electronically compensated current transducer
- zero-flux transformer, 2296, 2297
- 0.7 structure, 72.22, 2685
- z-1 filter, 2524
- zirconia, 2712–13
- zonal constant, 2061–2